Development and Applications of Quantitative Systems Pharmacology Methods and Tools for Drug Discovery

by

Fen Pei

BS, Peking University, 2013

MS, Peking University, 2015

Submitted to the Graduate Faculty of the

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Fen Pei

It was defended on

November 6, 2020

and approved by

Russell S. Schwartz, Professor, Department of Biological Sciences and Lane Center for Computational Biology, Carnegie Mellon University

J. Timothy Greenamyre, Professor, Pittsburgh Institute for Neurodegenerative Diseases

Timothy R. Lezon, Assistant Professor, Department of Computational and Systems Biology, University of Pittsburgh

Dissertation Co-Director: Ivet Bahar, Distinguished Professor and JK Vries Chair, Department of Computational and Systems Biology, Associate Director, University of Pittsburgh Drug Discovery Institute
Dissertation Co-Director: D. Lansing Taylor, Distinguished Professor and Allegheny Foundation Professor, Department of Computational and Systems Biology, Director, University of Pittsburgh Drug Discovery Institute

Development and Applications of Quantitative Systems Pharmacology Methods and Tools for Drug Discovery

Fen Pei, PhD

University of Pittsburgh, 2020

Quantitative Systems Pharmacology (QSP) is a relatively new field, which aims to determine the mechanisms of disease progression and mechanisms of action of drugs on multi-scale systems and to optimize the development of therapeutic strategies through iterative and integrated computational and experimental methods. Given the unclear mechanisms and unmet medical needs for complex diseases, there is a great need for integrated and efficient computational tools to facilitate the drug discovery process. This thesis focuses on the development and applications of computational methods for QSP-driven drug discovery, including (1) the development of an integrated and efficient chemical-protein-pathway mapping tool for polypharmacology and chemogenomics, implemented in the *QuartataWeb* server, (2) the development of machine learning methods for predicting protein-protein interactions (PPIs), and (3) the applications of the developed QSP methodology to Huntington's disease, drug abuse, and non-alcoholic fatty liver disease (NAFLD) toward better understanding of disease mechanisms and facilitating the design of therapeutic strategies. To build *QuartataWeb*, we adopted a probabilistic matrix factorization (PMF) method using as input two databases: DrugBank v5.0 and STITCH v5, so as to predict new chemical-target associations as well as detect similarities among drugs/chemicals based on their interaction patterns with targets, as well as similarities between targets based on their interaction patterns with drugs/chemicals. Furthermore, this new tool links

targets to KEGG pathways and Gene Ontology (GO) annotations, completing the bridge from drugs/chemicals to function via protein targets and cellular pathways. In the second study, we developed a methodology for automated and efficient identification PPIs using a symmetric logistic matrix factorization method. Finally, the applications have been conducted with experimental collaborators. We customized our QSP approaches based on specific disease-centric inputs and experimental resources, identified the cellular mechanisms underlying the investigated diseases or disorders, and proposed drugs to potentially serve as lead compounds for developing drugs against Huntington's disease, drug abuse and NAFLD. Taken together, the development and applications of the QSP methodology presented here demonstrate the power of QSP-guided hypotheses as a key step required for gaining a better understanding of systems-level events underlying complex diseases/disorders and for accelerating drug discovery.

Table of Contents

List of Abbreviations xvii

Background xxiii

1.0 Development of QuartataWeb: a New Interface for Integrated Chemical-Protein-Pathway Mapping for Polypharmacology and Chemogenomics 1

1.1 Introduction 1

1.2 Implementation and Pipeline of QuartataWeb 7

1.3 Applications of QuartataWeb 10

1.3.1 Chemogenomics Analysis for a List of Chemicals 10

1.3.2 Polypharmacological Evaluation of Drug Pairs 14

1.3.3 Drug Repurposing or Identification of Side-Effects 16

1.4 Conclusion 16

1.5 Materials and Methods 17

1.5.1 Datasets 17

1.5.2 Probabilistic Matrix Factorization-based Chemical-Target Interaction Prediction 18

1.5.3 Evaluation of Chemical-Target Interaction Prediction Performance 21

1.5.4 Chemical-Target Interaction Pattern-based Similarity 23

1.5.5 Ligand Structure-based Similarity 24

1.5.6 Enrichment Analysis 24

1.5.7 Other Supporting and Visualization Tools 26

1.6 Acknowledgement 26

2.0 Predicting Large Scale Protein-Protein Interactions Using Symmetric Logistic Matrix Factorization 27

2.1 Introduction 28

2.2 Materials and Methodology 31

2.2.1 Problem Formalization 31

2.2.2 Data Collection and Datasets Construction 32

2.2.3 Symmetric Logistic Matrix Factorization (symLMF) 35

2.2.4 Other Related Matrix Factorization Methods 39

2.2.4.1 Symmetric Probabilistic Matrix Factorization (symPMF) 39

2.2.4.2 Symmetric Non-Negative Matrix Factorization (symNMF) 39

2.2.4.3 Non-Negative Matrix Tri-Factorization (NMTF) 40

2.2.5 Hyperparameter Selection 40

2.2.6 Cross-Validation and Performance Evaluation 41

2.3 Results and Discussion 42

2.3.1 Performance Comparisons with State-of-the-Art Algorithms on Benchmark Datasets Demonstrate the Superiority of the Proposed Method 42

2.3.2 Applications to Extended Datasets Support the Utility of the Proposed Method in Predicting Large-scale PPIs 49

2.3.3 Performance of symLMF on Tissue/Disease-Specific Datasets 50

2.3.4 Ability of symLMF To Recapitulate Hidden Protein-Protein Interactions with Limited Data 51

2.3.5 *De Novo* Predictions of Protein-Protein Interactions 53

2.4 Conclusion 57

2.5 Acknowledgements **58**

3.0 Application of Quantitative Systems Pharmacology Methods to three complex diseases/disorders **59**

3.1 Huntington's Disease: Connecting Neuronal Cell Protective Pathways and Drug Combinations **61**

3.1.1 Introduction **61**

3.1.2 Results **66**

3.1.2.1 Characterization of neuronal cell protective compounds in the *STHdh*^{Q111} model **66**

3.1.2.2 Combinations show enhanced protective effects **71**

3.1.2.3 Inferring protection-relevant pathways from the compounds' canonical mechanisms **76**

3.1.2.4 Some compounds may be protecting by non-canonical mechanisms **87**

3.1.3 Discussion **89**

3.1.4 Materials and Methods **94**

3.1.4.1 Computational predictions of drug-target binding **94**

3.1.4.2 Pathway analysis **95**

3.1.5 Acknowledgement **96**

3.2 Drug Abuse: Addiction Progression Mechanism and the Effector Role of mTORC1 ..... **97**

3.2.1 Introduction **97**

3.2.2 Results **101**

3.2.2.1 Functional similarity of drugs of abuse does not imply structural similarity, consistent with the multiplicity of their actions 101

3.2.2.2 The Selected Drugs and Identified Targets Are Highly Diverse and Promiscuous 105

3.2.2.3 Pathway enrichment analysis reveals the major pathways implicated in various stages of addiction development 111

3.2.2.4 Selected targets shared by dominant pathways emerge as common mediators of drug addiction 117

3.2.2.5 Pathway Enrichment Highlights the Interference of Drugs of Abuse with Synaptic Neurotransmission 119

3.2.2.6 mTORC1 emerges as a potential downstream-effector activated by drugs abuse 122

3.2.3 Discussion 124

3.2.3.1 Persistent restructuring in neuronal systems as a feature underlying drug addiction 125

3.2.3.2 ANS may mediate the negative-reinforcement of drug addiction ... 126

3.2.3.3 mTORC1 appears as a key mediator of cellular morphological changes elicited in response to continued drug abuse 127

3.2.3.4 Drug repurposing opportunities for combating drug addiction 129

3.2.4 Materials and Methods 132

3.2.4.1 Selection of drugs of abuse and their known targets 132

3.2.4.2 Probabilistic matrix factorization (PMF) based drug-target interaction prediction 134

3.2.4.3 Pathway Enrichment Analysis 134

3.2.5 Acknowledgments 134

3.3 Non-Alcoholic Fatty Liver Disease (NAFLD): Identification of Repurposable Drugs 135

3.3.1 Introduction 135

3.3.2 Results 139

3.3.2.1 Patient clustering based on pathway variation is consistent with NAFLD clinical stages 139

3.3.2.2 The differentially regulated pathways (DRPs) identified among different patient clusters reveals the major pathways implicated in NAFLD progression 142

3.3.2.3 Drug predictions via connectivity map (CMap) 149

3.3.2.4 Predicted drug prioritization using network proximity analysis 151

3.3.2.5 Comparison of the predicted drugs and NAFLD clinical trial drugs 153

3.3.2.6 Selection of the final list of drugs to test in Liver MPS 154

3.3.3 Discussion.....157

3.3.3.1 QSP approach complemented the target-centric drug design for NAFLD..... 158

3.3.3.2 Evidences support the potential effect of our proposed drugs 159

3.3.3.3 Drug combinations have the potential to advance the NAFLD treatment 161

3.3.4 Materials and Methods 161

3.3.4.1 Generation of gene expression profiles 161

3.3.4.2 KEGG pathway analysis and identification of differentially expressed genes (DEGs) and differentially regulated pathways (DRPs)..... 162

3.3.4.3 Drug predictions via CMap 163

3.3.4.4 Drug prioritization via network proximity 164

3.3.4.5 Identification of targets and pathways for predicted and clinical trial drugs..... 166

3.3.5 Acknowledgments 166

4.0 Future Directions 167

4.1 Future Development of QuartataWeb 168

4.2 Future Development and Application of PPI Prediction 169

4.3 Future Work of Three Applications of QSP 170

Appendix A Supporting Materials for the Huntington's Disease Study 174

Appendix A.1 Relative solubility of protective compounds..... 174

Appendix A.2 Targets from DrugBank and STITCH for 32 identified probes..... 176

Appendix A.3 Mapping of 32 identified probes and targets in KEGG human pathways 178

Appendix A.4 Combination Pairs 181

Appendix A.5 Synergistic Pairs..... 184

Appendix B Supporting Materials for the Drug Abuse Study 186

Appendix B.1 Dataset of 50 addictive drugs and their corresponding groups and identifiers..... 186

Appendix B.2 2D structures of the dataset of 50 addictive drugs 189

Appendix C Supporting Materials for the NAFLD Study 193

Appendix C.1 Categorization of differentially regulated pathways 193

Appendix C.2 Differentially regulated pathways of each comparison 199

Appendix C.3 Predicted drugs for 12 input signatures using CMap..... 210

Appendix C.4 Predicted small molecules for 12 input signatures using CMap..... 215

Appendix C.5 Ranking of the 49 drugs predicted by CMap using network proximity z-score 220

Appendix C.6 Drug information of the selected 8 drugs for modulating steatosis..... 229

Appendix C.7 Drugs in clinical trials for NAFLD and NASH 236

Appendix C.8 Targets and pathways of 23 predicted drugs and 20 clinical trial drugs for NAFLD 239

Bibliography 245

List of Tables

Table 1.1 Existing web servers for drug-target interactions 4

Table 1.2 Data sources used in QuartataWeb 5

Table 1.3 Parameters used for training latent factor models in QuartataWeb 21

Table 1.4 Performance of QuartataWeb observed in cross-validation tests 22

Table 2.1 Description of the eight datasets used in the present study and corresponding symLMF model parameters 35

Table 2.2 Comparison of the performance of all methods using the *S. cerevisiae-benchmark* dataset 46

Table 2.3 Comparison of the performance of all methods using the *H. sapiens-benchmark* dataset 46

Table 2.4 Performance of MF models tested againist different benchmarking datasets 48

Table 2.5 Performance of symLMF on tissue/disease-specific datasets 51

Table 3.1 Synergistic compound pairs that coverge on PKA/PKG signaling 77

Table 3.2 Enrichment of the 190 targets of addictive drugs in the brain and others 110

Table 3.3 Fifteen drugs for liver and lung diseases 131

Table 3.4 Distribution of 182 patient samples in each patient cluster based on GSVA clustering 142

List of Figures

Figure 1.1 Promiscuity of drugs/chemicals and targets in DrugBank and STITCH database 6

Figure 1.2 Detailed description of QuartataWeb server workflow for three input types 9

Figure 1.3 Illustration of QuartataWeb output on Input Type I and Input Type III examples 11

Figure 1.4 Illustration of QuartataWeb outputs on pathways and GO term enrichment released by QuartataWeb 13

Figure 1.5 A snapshot of the results page from QuartataWeb 15

Figure 1.6 Distribution of interaction confidence scores 20

Figure 1.7 Performance of QuartataWeb 22

Figure 2.1 Distribution of the protein degrees (number of connections at the nodes, each node representing on protein) in the PPI network described by eight datasets 37

Figure 2.2. Comparison of the performance of different methods tested against two benchmark datasets. 45

Figure 2.3 Comparison of the performance of MF methods on benchmark and extended datasets 47

Figure 2.4 Ability of the symLMF to recapitulate the hidden protein-protein interactions in the top 1,000 predicted pairs and improved performance with increasing size of dataset 53

Figure 2.5 Comparison of the top 1,000 predictions between symLMF and SPRINT models 56

Figure 3.1 Chemogenomics component of the QSP platform 64

Figure 3.2 Heterogeneity in mutant huntingtin (mHTT) induced neurotoxicity in *STHdh*^{Q111} cells 67

Figure 3.3 Compounds with confirmed neuroprotective activity in the *STHdh*^{Q111} model.. 68

Figure 3.4 Change in total and dead cell numbers with compound treatment relative to DMSO..... 70

Figure 3.5 Protective compounds are not quenching the PI signal 71

Figure 3.6 Combinations of probes with different canonical mechanisms provide enhanced protection of *STHdh*^{Q111} cells 74

Figure 3.7 Combinations of probes show synergistic protection in STHdh Q111 cells..... 75

Figure 3.8 Protective compounds can induce cAMP 79

Figure 3.9 PKA inhibitor H89 inhibits the protective effects of several probes 81

Figure 3.10 H89 inhibits phosphorylation of CREB at Ser 133..... 83

Figure 3.11 PKA activation is heterogeneous within the *STHdh*^{Q111} cell population 84

Figure 3.12 Protective compounds can activate PKA..... 86

Figure 3.13 Correlation between Percent Recovery from mHTT toxicity and pPKA levels 87

Figure 3.14 Comparison of carbonic anhydrase inhibitors in protecting *STHdh*^{Q111} cells .. 88

Figure 3.15 Ethoxzolamide may not work through the canonical carbonic anhydrase mechanism 89

Figure 3.16 Workflow of the QSP analysis 100

Figure 3.17 Dataset of 50 drugs of abuse: structure and interaction similarities, and classification of their targets 103

Figure 3.18 Promiscuity of drugs of abuse and their targets, and major families of proteins targeted by drugs of abuse 104

Figure 3.19 Prediction of new targets for known drugs of abuse 108

Figure 3.20 Results from pathway and target enrichments analysis 112

Figure 3.21 Pathways distinguished by the high propensity of targets of abused drugs ... 113

Figure 3.22 Pathway and target enrichments in five functional categories and the overlap of targets in different categories 114

Figure 3.23 The impact of drugs of abuse on synaptic neurotransmission 121

Figure 3.24 A unified signaling network mediates the effects of drugs of abuse 123

Figure 3.25 The enriched overlapping pathways for lung/liver drugs and drugs of abuse 132

Figure 3.26 Overview of workflow used to predict drugs for NAFLD via clinical gene expression profiles 138

Figure 3.27 Unsupervised clustering of individual patients based on KEGG pathway enrichment scores gene set variation analysis (GSVA) broadly follows disease stage 141

Figure 3.28 Distribution of the enriched KEGG pathways in KEGG pathway groups and NAFLD pathway categories 143

Figure 3.29 Top 10 differentially regulated pathways (A) and overlapping pathways (B) among three comparisons 148

Figure 3.30 Subnetwork of the steatosis-related targets and pathways of the 8 prioritized drugs 156

List of Abbreviations

Abbreviation	Definition
5-HT	Serotonin
AC	Auto covariance
AC	Adenylate cyclase
ACC	Auto cross covariance
ACE	Angiotensin I converting enzyme
ANS	Autonomic nervous system-innervation pathways
AUC	Area under the receiver operating curve
AUPR	Area under the precision-recall curve
BCI	Bliss combination index
BMI	Body mass index
CACNA1A	Calcium voltage-gated channel subunit α1 A
CHRM1	Cholinergic receptor muscarinic 1
CMap	Connectivity map
CNN	Convolution neural network
CNS	Central nervous system
COX2	Prostaglandin-endoperoxide synthase 2
CRHR1	Corticotropin-releasing factor receptor 1
CTI	Chemical-target interaction
DA	Dopamine

DAT	Dopamine transporter
DB	Database
DEG	Differentially expressed gene
DNL	*De novo* lipogenesis
DNN	Deep neural network
DRP	Differentially regulated pathways
DS	Disease pathways
DTI	Drug-target interaction
ER	Endoplasmic reticulum
FAA	Free fatty acid
FDR	False discovery rate
FN	False negative
FP	False positive
GABA	γ-aminobutyric acid
GO	Gene Ontology
GSVA	Gene set variation analysis
HCC	Hepatocellular carcinoma
HD	Huntington's disease
HTT	Huntingtin
IGF1R	Insulin-like growth factor 1 receptor
IID	Integrated Interaction Database
IKKβ	IκB kinase β
IP3	Inositol triphosphate

KNN	K-nearest neighbor
LAMPS	Human liver acinus MPS
LFM	Latent factor model
LMF	Logistic matrix factorization
LTD	Long-term depression
LTP	Long-term potentiation
mACh	Muscarinic acetylcholine
MAPPIT	Mammalian protein-protein interaction trap
mAUC	Mean area under the receiver operating curve
mAUPR	Mean area under the precision-recall curve
MF	Matrix factorization
mHTT	Mutant huntingtin
ML	Machine learning
MMI	Multivariate mutual information
MPS	Microphysiological system
mTOR	Mechanistic target of rapamycin kinase
mTORC1	Mammalian target of rapamycin complex 1
N&S	Normal & steatosis
nACh	Nicotinic acetylcholine
NAFL	Simple steatosis (fatty liver)
NAFLD	Non-alcoholic fatty liver disease
NASH	Nonalcoholic steatohepatitis
NE	Noradrenaline/norepinephrine

NET	Noradrenaline/norepinephrine transporter
NET/SLC6A2	Norephinephrine transporter
NLP	Natural language processing
NMBAC	Normalized Moreau-Broto Autocorrelation
NMDA	N-methyl-d-aspartate
NMDAR	N-methyl-d-aspartate receptor
NMDAR	Glutamate receptor
NMTF	Non-negative matrix tri-factorization
NP	Neuroplasticity related pathways
NT	Neurotransmission related pathways
OFS	Outward-facing state
OPR	Opioid receptor
OPRM1	μ-type opioid receptor
PCA	Principle component analysis
PDPK1	3-phosphoinositide dependent protein kinase 1
PF	Predominately fibrosis
PIK3CA	Phosphatidylinositol 3-kinase class 1A catalytic subunit α
PLI	Lobular inflammation
PMF	Probabilistic matrix factorization
PPARG	Peroxisome proliferator activated receptor gamma
PPI	Protein-protein interaction
QSP	Quantitative Systems Pharmacology
RAC1	Ras-related C3 botulinum toxin substrate 1

RF	Random forest
ROC	Receiver operating curve
ROS	Reactive oxygen species
SERT	Serotonin transporter
SG	Signal transduction pathways
SLC52A2	Riboflavin transporter 2A
SLC6A1	Na+/Cl−-dependent GABA transporters
SLC6A9	Glycine transporter
SPRINT	Scoring Protein INTeractions
SVMs	Support vector machines
symLMF	Symmetric logistic matrix factorization
symNMF	Symmetric non-negative matrix factorization
symPMF	Symmetric probabilistic matrix factorization
T2D	Type 2 diabetes
TAAR1	Trace amine-associated receptor 1
TN	True negative
TP	True positive
TRPA1	Transient receptor
UPR	Unfolded protein response
VGCC	Voltage-dependent P/Q-type calcium channel
vLAMPS	Vascularized liver microphysiology system
VMAT	Vesicular monoamine transporters
WSRC	Weighted sparse representation classifier

β1AR	β1 adrenergic receptor

Background

Over the past decade, the field of drug discovery and development has transitioned from a target-centric and phenotypic discovery area to include complementary system-level approaches (Sorger, et al., 2011; Stern, et al., 2016). The traditional target-centric and phenotypic drug discovery and development approach follows a linear set of steps, usually starting with the investigation of the basic science of a certain disease, followed by the identification and validation of a druggable target or a disease-relevant phenotype. The next step would be to develop target- or phenotype-specific assays, to screen against selected compound libraries, then identify hits and generate leads. After obtaining the leads, medicinal chemistry methods are used to optimize the leads, and ensure safety and acceptable ADMET (absorption, distribution, metabolism, excretion and toxicity) properties. Then the efficacy and pharmacokinetics of the candidates are tested using preclinical animal models. If a drug candidate passes all these steps, it gradually moves from phase I, to II, to III human clinical trials.

Although traditional drug discovery and development approach has been successful in a number of cases, it usually suffers from high cost and low success rate at advanced phases. In a recent analysis of data between 2009 and 2018 (Wouters, et al., 2020), the estimated median cost of getting a new drug into market was $985 million, and the average cost was $1.3 billion. A previous study placed the average cost as $2.8 billion and the typical investigation time was over 10 years (DiMasi, et al., 2016). The overall failure rate in drug development is as high as 96%, including 90% failure during the clinical development (Hingorani, et al., 2019). In addition, in the case of complex diseases including neurodegenerative diseases, metabolic syndromes and cancer,

with poorly understood pathogenesis, the failure rate is even higher. For example, there is no FDA-approved drug for non-alcoholic fatty liver disease (NAFLD).

Therefore, there is a great need for improving the efficiency of developing therapeutics and for advancing personalized medicine strategies by optimizing the process of drug discovery and development. In humans, proteins perform most of their complex functions via interactions with other proteins, forming cellular pathways or protein-protein interaction (PPI) networks. It is also widely known that many drugs or compounds bind to more than one protein target, and the target proteins interact with many other proteins. These interactions modulate the disease sub-network. Therefore, the effect of a drug should not by evaluated based on its interaction with a certain target only, but based on its overall effect on a disease sub-network. This systems-level approach is aimed to replace the traditional "one-gene, one-target, one-mechanism" hypothesis with an in-depth understanding of complex networks underlying the disease mechanisms (Bai, et al., 2019; Berg, et al., 2010; Bradshaw, et al., 2019; Kiyosawa and Manabe, 2016; Leung, et al., 2013; Perez-Nueno, 2015; Sorger, et al., 2011; Stern, et al., 2016) and to develop drugs with desired system-level effects.

With advances in experimental techniques, the generation of omics data (genomics, transcriptomics, proteomics, and metabolomics), as well as the development of computational power and sophisticated algorithms, multi-scale systems analysis has become possible (Cheng, 2019; Leung, et al., 2013; Li, et al., 2017; Perez-Nueno, 2015). The Quantitative Systems Pharmacology (QSP) platform (Stern, et al., 2016) at the Drug Discovery Institute of the University of Pittsburgh provides an innovative pipeline for integrated and iterative drug discovery combining quantitative experimental and computational tools, instead of the traditional linear process. Most of the components in the QSP platform have been applied in traditional approaches, such as

identification of drug-target interaction, identification of hits and generation of leads, or lead optimization. In addition, system level analyses such as inference of pathways involved in disease progression and mathematical modeling of disease progression form important elements of the platform.

In this thesis, I focus on the development and application of computational methods that are coupled with experimental methods, to generate QSP hypotheses and candidate solutions to treat selected diseases/disorders. The thesis contains both QSP method development and QSP applications. In the method development part, I introduced or adopted new computational algorithms and tools for QSP studies, such as those required for predicting drug-target interactions (DTIs) and PPIs and evaluating gene-set enrichment. In the application part, we designed and implemented customized workflows for Huntington's disease, drug abuse, and NAFLD. Specifically, we analyzed the targets and pathways of a list of active compounds identified through phenotypic screening assays in a Huntington's disease model, to generate a hypothesis on disease mechanisms, and then verified the hypothesis with subsequent biomarker assays. We also performed a comprehensive system level drug-target-pathway analysis for drugs of abuse across six different categories, and discovered the potential networks of interactions underlying the drug addiction process. Finally, we integrated the information retrieved from the connectivity map (CMap) tool and a network proximity analysis method to identify repurposable drugs for NAFLD patients, using RNA-seq data from NAFLD patients.

Most of the work presented in this thesis has been published. As the copyrights permit, some of the materials from previous publications are reused or quoted with proper citations in the following chapters. Some of the studies were accomplished in collaboration with other people whose contributions are acknowledged where appropriate. All presented studies were conducted

under the supervision and guidance of my doctoral advisors, Drs. Ivet Bahar and D. Lansing Taylor.

1.0 Development of QuartataWeb: a New Interface for Integrated Chemical-Protein-Pathway Mapping for Polypharmacology and Chemogenomics

In this section, we introduce the development of QuartataWeb, a user-friendly server developed for polypharmacological and chemogenomics analyses. Using QuartataWeb, users can easily obtain information on experimentally verified (known) and computationally predicted (new) interactions between 5,494 drugs and 2,807 human proteins compiled in DrugBank (Wishart, et al., 2017), and between 315,514 chemicals and 9,457 human proteins in the STITCH (Szklarczyk, et al., 2016) database. The predictions are based on a highly efficient machine learning (ML) algorithm, probabilistic matrix factorization (PMF) (Cobanoglu, et al., 2013). In addition, QuartataWeb links targets to KEGG pathways (Kanehisa, et al., 2017) and Gene Ontology (GO) annotations (Huntley, et al., 2015), completing the bridge from drugs/chemicals to function via protein targets and cellular pathways. It allows users to query a series of chemicals, drug combinations, or multiple targets, to enable multi-drug, multi-target, multi-pathway analyses, toward facilitating the design of polypharmacological treatments for complex diseases. QuartataWeb is a useful component of our QSP platform and is freely accessible at http://quartata.csb.pitt.edu.

1.1 Introduction

Lorem ipsum dolor sit amet, It is now widely accepted that many complex diseases are associated with multiple targets, which in turn affect multiple pathways, requiring the adoption of

QSP approaches for assessing the mechanisms of disease etiology, progression, and treatment (Leung, et al., 2013; Ma'ayan, et al., 2014; Perez-Nueno, 2015; Stern, et al., 2016) . The possibility of exploiting the promiscuity of drugs via drug repurposing and polypharmacological treatments also emerged in recent years as a means of reducing risk and cost in drug development (Ashburn and Thor, 2004; Pantziarka, et al., 2018; Sachs, et al., 2017). In parallel, chemogenomics studies assist in improving our understanding of disease mechanisms and developing therapeutic strategies by providing phenotypic information on ensembles of active compounds screened against families of targets. Such studies underscore the importance of developing computational tools that would harness the rapidly accumulating data to predict new chemical-target interactions (CTIs) over a broad space of chemicals and enable their mapping to pathways and function.

Several ML approaches have been developed for predicting CTIs over the past decade. Most of them are supervised learning methods, such as kernel regression-based method (Yamanishi, et al., 2008), correlation-based method (Yamanishi, et al., 2010), random forest algorithm (Cao, et al., 2014), bipartite local models (Bleakley and Yamanishi, 2009), and kernelized sparse learning SVM (Shi, et al., 2013). Those methods use criteria such as chemical-chemical similarities based on 2D fingerprints or 3D conformations, protein-protein similarities based on sequence or structure properties. These criteria heavily rely on the similarity-representing methods and are limited to proteins with structural data. To extend this limitation, other supervised methods such as restricted Boltzmann machines (Wang and Zeng, 2013) and Gaussian interaction profile kernels (van Laarhoven, et al., 2011) have been developed, which are based on drug-target interaction networks without using structural data on chemicals or proteins. The prediction accuracy of supervised learning methods might, however, be biased by the inaccurate negative sample selection, since the chemical-protein pairs without interactions lack experimental

verification. To meet this challenge, semi-supervised methods such as Laplacian regularized least square (Xia, et al., 2010) and matrix factorization methods including PMF (Cobanoglu, et al., 2013) have been developed. PMF is chosen in QuartataWeb because: (i) it does not need negative samples to train the model, avoiding bias generated from negative sample selection; (ii) it does not rely on the structure information of chemicals or proteins, rendering it applicable and efficient to predict large-scale CTIs; and (iii) PMF shows better performance than several other ML methods that use structural information on either ligand or target, and the predictions are complementary to those methods (Cobanoglu, et al., 2013; Ezzat, et al., 2017).

Several resources (see **Table** 1.1) have been developed in the last decade to address different aspects of the emerging needs but an integrated server designed to automate the association of multiple CTIs with enriched pathways and function remains to be developed. For example, the servers SEA (Keiser, et al., 2007), SwissTargetPrediction (Daina, et al., 2019), and SuperPred (Nickel, et al., 2014) predict new CTIs, but not corresponding pathways. DINIES (Yamanishi, et al., 2014) and DT-Web (Alaimo, et al., 2015) incorporate pathway information, but not large-scale CTIs, their respective data being limited to KEGG and DrugBank. Furthermore, existing interfaces are not designed to use as input multiple drugs/targets for polypharmacological strategies and/or for complementing chemogenomics efforts.

Table 1.1 Existing web servers for drug-target interactions

Web-servers	Chemicals	Targets	Prediction method	Pathways
SEA (Keiser, et al., 2007)	PubChem	MDDR (246 targets)	physicochemical similarities	no
DINIES (Yamanishi, et al., 2014)	KEGG drugs	KEGG targets	similarity based ML models	KEGG pathways
SuperPred (Nickel, et al., 2014)	query-based	SuperTarget, ChEMBL, BindingDB	2D and 3D similarities	no
PharmMapper (Wang, et al., 2017)	query-based	PDB	pharmacophore models	no
DR. PRODIS (Zhou, et al., 2015)	DrugBank	human proteome	structure information on targets	no
iDrug-Target (Xiao, et al., 2015)	the original benchmark dataset	the original benchmark dataset	chemical fingerprints and a ML model	no
DT-Web (Alaimo, et al., 2015)	DrugBank	DrugBank	network-based inference	no
BalestraWeb (Cobanoglu, et al., 2015)	DrugBank approved drugs	DrugBank	latent factor models	no
SwissTargetPrediction (Daina, et al., 2019)	280,318 bioactive small molecules	2,686 targets	combined 2D and 3D similarities	no

We developed the QuartataWeb server to address those needs. QuartataWeb uses known (experimentally verified) CTIs from DrugBank and STITCH (**Table** 1.2 and **Figure** 1.1) in a PMF algorithm (Cobanoglu et al. 2013) to predict new CTIs in the extended space of more than 300,000 chemicals and 9,000 human proteins. The engine parameters have been optimized to ensure high CTI prediction accuracy. The outputs are linked to KEGG pathways and GO Annotations (GOAs) (Huntley et al. 2015) to predict the most probable pathways, functions and processes affected by

one or more chemicals and to efficiently assist in interpreting and/or guiding chemogenomics and polypharmacological studies.

Table 1.2 Data sources used in QuartataWeb

Data type and properties	**DrugBank 5.1 (Wishart, et al., 2017)**		**STITCH 5 (human) (Szklarczyk, et al., 2016)**	
	Approved	**All**	**Experimental**	
# of drugs/chemicals (N)	1,883	5,494	315,514	
# of targets (M)	2,244	2,807	9,457	
Interaction space (NxM)	4,225,452	15,421,658	2.98 billion	
known interactions (S)	9,253	14,983	5,364,673	
Occupancy of $\boldsymbol{R}^{(a)}$ (S/[NxM])	0.0022	0.0010	0.0018	
Data type and properties	**KEGG (human) (Kanehisa, et al., 2017)**	**GOA (human) (Huntley, et al., 2015)**		
		Molecular Function	**Biological Process**	**Cellular component**
# of genes	19,124	2,807		
# of pathways/GO terms	323	2,346	6,892	897
# of associations	28,664	11,475	39,510	15,160

(a) ***R*** represents the chemical-target interaction matrix, described in the Method below.

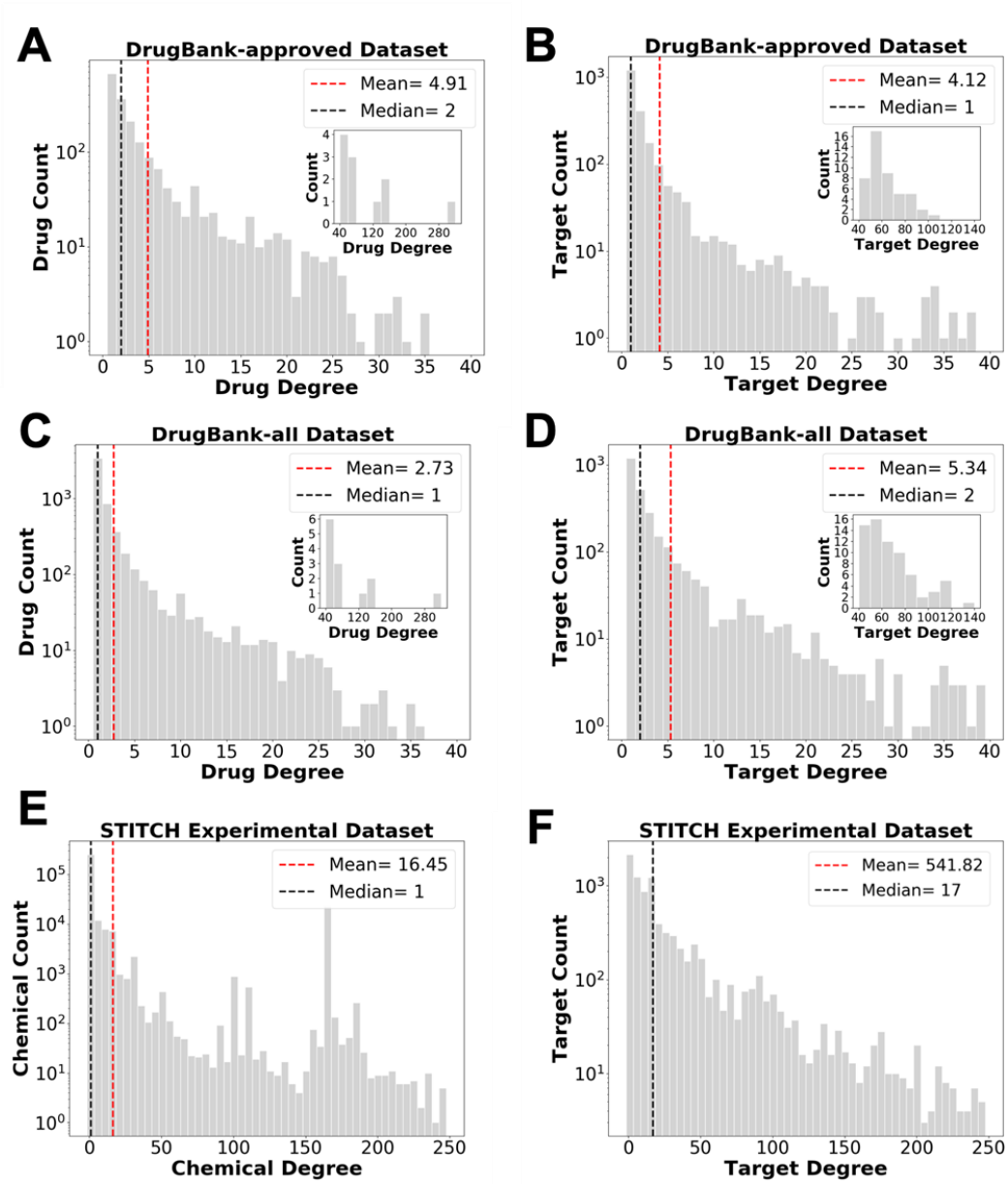


Figure 1.1 Promiscuity of drugs/chemicals and targets in DrugBank and STITCH database

Histogram of the degrees of drugs/chemicals and targets in the bipartite network of CTIs, computed for DrugBank-approved (A-B), DrugBank–all (C-D) and STITCH-experimental (E-F) datasets. The degree of a given node represents the number of links emanating from that node, connected to first neighbors in the network, each node representing a drug/chemical or a target protein. Most drugs and targets in DrugBank have degrees below 40. Chemicals and targets in STITCH have much larger numbers of interactions (E-F). The distributions of drugs or targets

with degree higher than 40 are shown in the insets of panels A-D. The dashed vertical lines indicate the mean (*red*) and median (*black*) in each case. The corresponding values are given in the insets**.**

1.2 Implementation and Pipeline of QuartataWeb

The QuartataWeb server pipeline is schematically depicted in **Figure** 1.2. The server can be flexibly queried with three types of input: (I) a list of chemicals (or targets) for chemogenomics-like screening in silico (**Figure** 1.2**A**); (II) one or more pairs of chemicals to be administered in combination for polypharmacological purposes (**Figure** 1.2**B**); and (III) a single chemical and/or a single target to be characterized (**Figure** 1.2**C**). In response to a list of chemicals entered in type I query, QuartataWeb releases newly predicted CTIs and chemical-chemical similarities based on pre-computed latent factor models (LFMs) learned from DrugBank or STITCH data, in addition to retrieving known CTIs from these datasets, as schematically described in **Figure** 1.2**A**. The outputted targets are then subjected to target enrichment, which also lead to enrichment scores for associated pathways and GOAs (p-values) (See Methods). The same sequence of tasks can be carried out for a list of targets entered as input. In the case of Type II input, the same tasks are carried out for pairs of chemicals to obtain shared targets, and their enrichment, along with enriched pathways and GOAs. Type III input is the simplest query where users enter one chemical, one target or a chemical-target pair to identify associated CTIs, similar chemicals or targets, and enriched pathways and GOAs. Furthermore, the secondary interactions (2˚, beyond the immediate neighbors) in the bipartite network of chemical/targets can be visualized. In all cases, outputs are presented as tables with several specifications (e.g., drug/chemical identifiers, Gene IDs and names, PDB IDs, confidence scores, and enrichment p-values), in addition to visuals such as

network representations, bar plots, or heatmaps. The force-directed layout in JavaScript D3 package has been customized and designed to interactively display the CTIs and pathways networks.

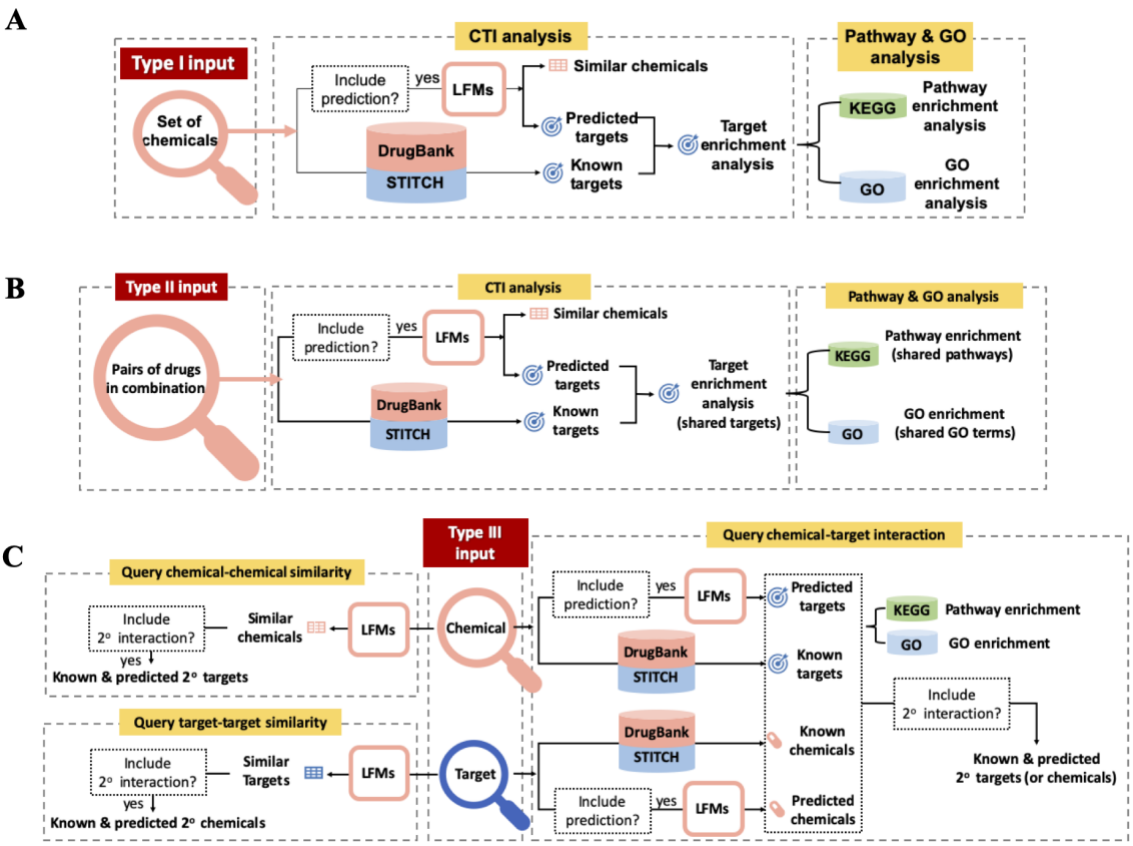


Figure 1.2 Detailed description of QuartataWeb server workflow for three input types

(A) In type I input, users enter either a list of chemicals, or a list of targets of interest, details are descripted in the text. (B) In Type II input, one or more pairs of chemicals are administrated in combination (and queried in combination), and in Type III one chemical, target or a chemical-target pair is given as query. (C) In type II input, a workflow similar to that of type I holds, with the exception that, targets, pathways and GO terms that are *shared* by pairs of chemicals (administered in combination, rather than serially) are released as outputs. In type III, the user can retrieve information on known and predicted interactions as well as similar chemicals-based on CTI patterns. The space of targets is that of human proteins in DrugBank or STITCH. The user can select either database. Both known and predicted targets are mapped to KEGG pathways and GO terms to perform an enrichment analysis. The secondary chemicals that interact with these targets can be viewed if the "secondary interactions" option button is selected. Likewise, the targets of the similar chemicals can be viewed by a similar option. When users enter a target in type III query, they can retrieve information on CTIs or on target-target similarities, and explore the secondary interactions in a similar way.

1.3 Applications of QuartataWeb

1.3.1 Chemogenomics Analysis for a List of Chemicals

In many cases, a set of chemicals exhibiting comparable phenotypes are analyzed in phenotypic screening. Suppose we are interested in finding out the common mechanism of action of this set of chemicals (type I input). **Figure** 1.3**A** illustrates such a case where four drugs are inputted having the same phenotype. QuartataWeb identifies the common targets along with the interaction confidence scores and enrichment scores of the targets. One may further learn about the pathways associated with the shared targets (**Figure** 1.3**B**) and the corresponding GOAs (**Figure** 1.3**C**). The interface also provides tables with detailed information on the pathways and GOAs, including their p-values (**Figure** 1.4), which could help assessing the dominant pathways and processes that underlie the shared phenotype. Our recent QSP analysis of 50 drugs of abuse serves as an example of the utility of this type of integrated studies (Pei, et al., 2019), as will be described in Chapter 3.2.

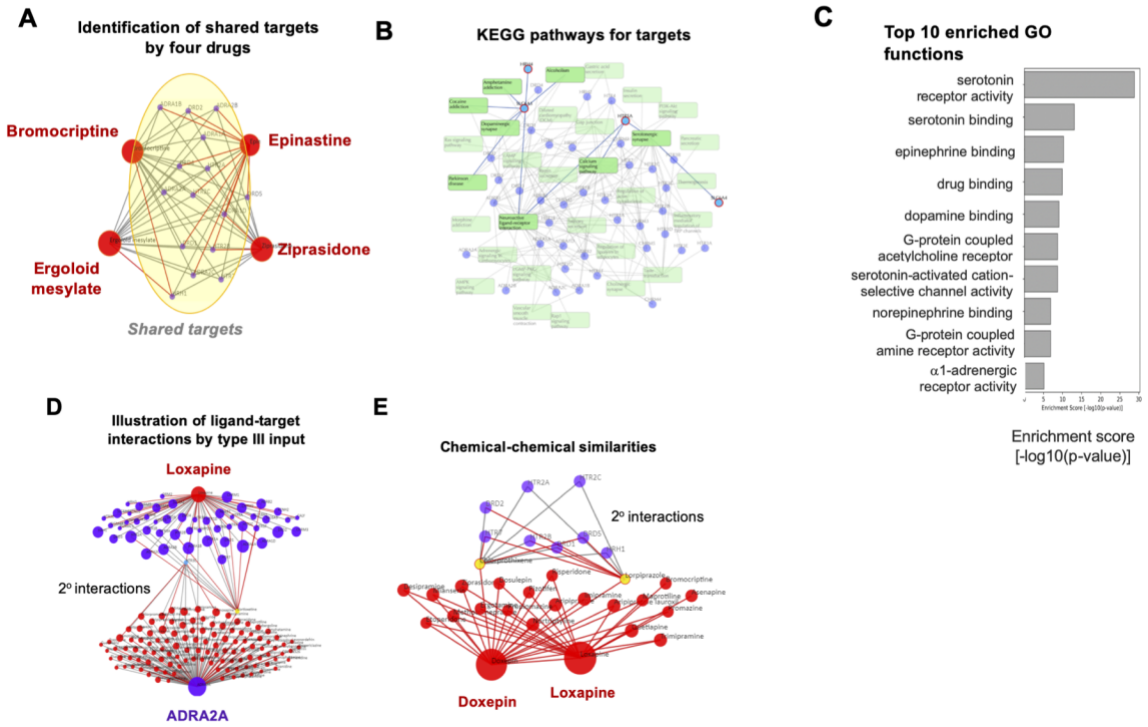


Figure 1.3 Illustration of QuartataWeb output on Input Type I and Input Type III examples

(A) Identification of targets (*dark violet dots, in yellow ellipse*) shared by four drugs (Input Type I) indicated by *red nodes*. (B) target-pathway network view of KEGG pathways (*green boxes*) corresponding to tragets in A. (C) Top 10 enriched GO molecular function for targets in B. Bar plot shows enrichment p-values. (D) Illustration of ligand-target intereactions obtained by Type III input. Second generation of nodes with degrees less than 3 are hidden by applying "*Trim 2nd generation nodes*" button. (E) Chemical-chemical similarities. The option "*Display secondary interactions*" displays targets shared by selected drugs (*yellow*).

A

30 pathways for known targets of input drugs:

No.	Pathway ID	Pathway name	Pathway Class 1	Pathway Class 2	Targets	# of targets	*p*-value (EPD)	*p*-value (EPT)↑
1	hsa04080	Neuroactive ligand-receptor interaction	Environmental Information Processing	Signaling molecules and interaction	ADRA2B;DRD5;CHRM5;CHRM4;HTR7;ADRA1B;HTR2B;DRD1;HTR6;DRD3;HTR1B;ADRA1A;DRD2;HTR1D;HTR2C;HTR2A;HRH2;ADRA2A;CHRM2;CHRM1;HTR1A;HTR1E;CHRM3;HTR1F;ADRB2;ADRA2C;ADRB3;DRD4;HRH1;ADRB1;HTR4;ADRA1D	32	0.0086	1.0E-26
2	hsa04020	Calcium signaling pathway	Environmental Information Processing	Signal transduction	ADRB1;CHRM1;DRD5;HTR2C;HTR6;ADRA1B;HTR4;ADRA1A;CHRM3;CHRM2;ADRB2;CHRM5;HTR7;HTR2A;ADRA1D;HTR2B;HRH2;DRD1;ADRB3;HRH1	20	0.003	6.3E-14
3	hsa04726	Serotonergic synapse	Organismal Systems	Nervous system	HTR3C;HTR6;HTR1A;HTR3A;HTR3E;HTR1B;HTR4;HTR3B;HTR1E;HTR1F;HTR3D;HTR1D;HTR2C;HTR7;HTR2A;HTR2B	16	0.001	1.4E-12
4	hsa04742	Taste transduction	Organismal Systems	Sensory system	HTR3C;HTR1A;HTR3A;HTR1B;HTR3B;HTR1E;CHRM3;HTR1F;HTR3E;HTR1D;HTR3D	11	0.0039	2.7E-8
5	hsa04024	cAMP signaling pathway	Environmental Information Processing	Signal transduction	ADRB1;CHRM1;DRD5;HTR1E;HTR6;HTR1B;HTR4;HTR1A;ADRB2;HTR1F;HTR1D;DRD2;DRD1;CHRM2	14	0.029	4.1E-7
6	hsa04970	Salivary secretion	Organismal Systems	Digestive system	ADRB1;ADRA1A;ADRB2;CHRM3;ADRA1B;ADRB3;ADRA1D	7	0.00043	0.00065
7	hsa04022	cGMP-PKG signaling pathway	Environmental Information Processing	Signal transduction	ADRB1;ADRA2B;ADRA2C;ADRA1A;ADRB2;ADRA1B;ADRB3;ADRA2A;ADRA1D	9	0.0007	0.00078
8	hsa04540	Gap junction	Cellular Processes	Cellular community - eukaryotes	ADRB1;DRD2;HTR2C;HTR2A;HTR2B;DRD1	6	0.00043	0.0032
9	hsa04725	Cholinergic synapse	Organismal Systems	Nervous system	CHRM1;CHRM4;CHRM5;CHRM3;CHRM2	5	0.33	0.048
10	hsa04728	Dopaminergic synapse	Organismal Systems	Nervous system	DRD5;DRD1;DRD4;DRD2;DRD3	5	0.008	0.081

B

46 GO molecular function of known targets for input drugs:

No.	GO terms	Targets	# of targets	Drugs	# of drugs	*p*-value (EPT)↑	*p*-value (EPD)
1	serotonin receptor activity	HTR3C;HTR6;HTR1A;HTR3A;HTR3E;HTR1B;HTR4;HTR3B;HTR1E;HTR1F;HTR3D;HTR1D;HTR2C;HTR7;HTR2A;HTR2B	16	Bromocriptine;Ziprasidone;Epinastine;Ergoloid mesylate	4	1.5E-29	9.8E-6
2	serotonin binding	HTR3A;HTR1B;HTR1E;HTR1F;HTR1D;HTR2C;HTR2A;HTR2B	8	Bromocriptine;Ziprasidone;Epinastine;Ergoloid mesylate	4	8.4E-14	9.2E-6
3	epinephrine binding	ADRB1;ADRA2B;ADRB2;ADRA2C;ADRB3;ADRA2A	6	Bromocriptine;Ziprasidone;Epinastine;Ergoloid mesylate	4	5.3E-11	9.2E-6
4	drug binding	CHRM1;DRD3;HTR1B;DRD2;CHRM3;ADRB2;HTR2C;HTR2A;DRD4;DRD1;HTR2B;CHRM2	12	Bromocriptine;Ziprasidone;Epinastine;Ergoloid mesylate	4	1.1E-10	0.0019
5	dopamine binding	DRD5;DRD3;DRD2;ADRB2;DRD4;DRD1	6	Bromocriptine;Ziprasidone;Ergoloid mesylate	3	8.7E-10	0.00081
6	serotonin-activated cation-selective channel activity	HTR3C;HTR3D;HTR3A;HTR3B;HTR3E	5	Ziprasidone;Ergoloid mesylate	2	2.0E-9	0.00065
7	G-protein coupled acetylcholine receptor activity	CHRM1;CHRM4;CHRM5;CHRM3;CHRM2	5	Ziprasidone	1	2.0E-9	0.12
8	norepinephrine binding	ADRB1;ADRB3;ADRA2A;ADRB2	4	Bromocriptine;Ziprasidone;Epinastine;Ergoloid mesylate	4	1.3E-7	8.1E-6
9	G-protein coupled amine receptor activity	DRD5;DRD1;DRD4;DRD3	4	Bromocriptine;Ziprasidone;Ergoloid mesylate	3	1.3E-7	4.5E-5
10	alpha1-adrenergic receptor activity	ADRA1A;ADRA1B;ADRA1D	3	Bromocriptine;Ziprasidone;Epinastine;Ergoloid mesylate	4	7.5E-6	2.4E-6

Figure 1.4 Illustration of QuartataWeb outputs on pathways and GO term enrichment released by QuartataWeb

(A) KEGG pathways in which all targets are tabulated for four inputted drugs "epinastine, ergoloid mesylate, bromocriptine and ziprasidone". The list is ordered here based on enrichment *p*-values. (B) GO molecular functions for the same targets listed in a sortable table, which is sorted here based on enrichment *p*-values calculated for the targets. 30 KEGG pathways and 46 GO molecular functions are displayed on the results pages (only top 10 are shown here in each case). EPT: enrichment p-value based on targets, to describe if a pathway or GOA is overrepresented

among the targets; EPD: enrichment p-value based on drugs/chemicals, to describe if a pathway or GOA is overrepresented among the chemicals, through the CTIs.

1.3.2 Polypharmacological Evaluation of Drug Pairs

Similarly, **Figure** 1.5 illustrates the output from QuartataWeb for identifying common targets given pairs of chemicals (type II input) that trigger comparable responses, e.g., aripiprazole/olanzapine, clozapine/trimipramine, methotrimeprazine/epinastine and cabergo-line/mianserin. The corresponding CTIs are listed in tables, and also displayed in a network viewer with an interactive control panel. Links to pathways and GOAs result pages are indicated. In this example, 15 among 192 known and 80 predicted (confidence scores > 0.9) targets were identified as common targets. Pathways shared by each chemical pair are also listed in the pathway enrichment table. This type of analysis applied to drug combinations used in a Huntington's disease model helped elucidate the origin (shared pathways) of observed synergistic effects (Pei, et al., 2017).

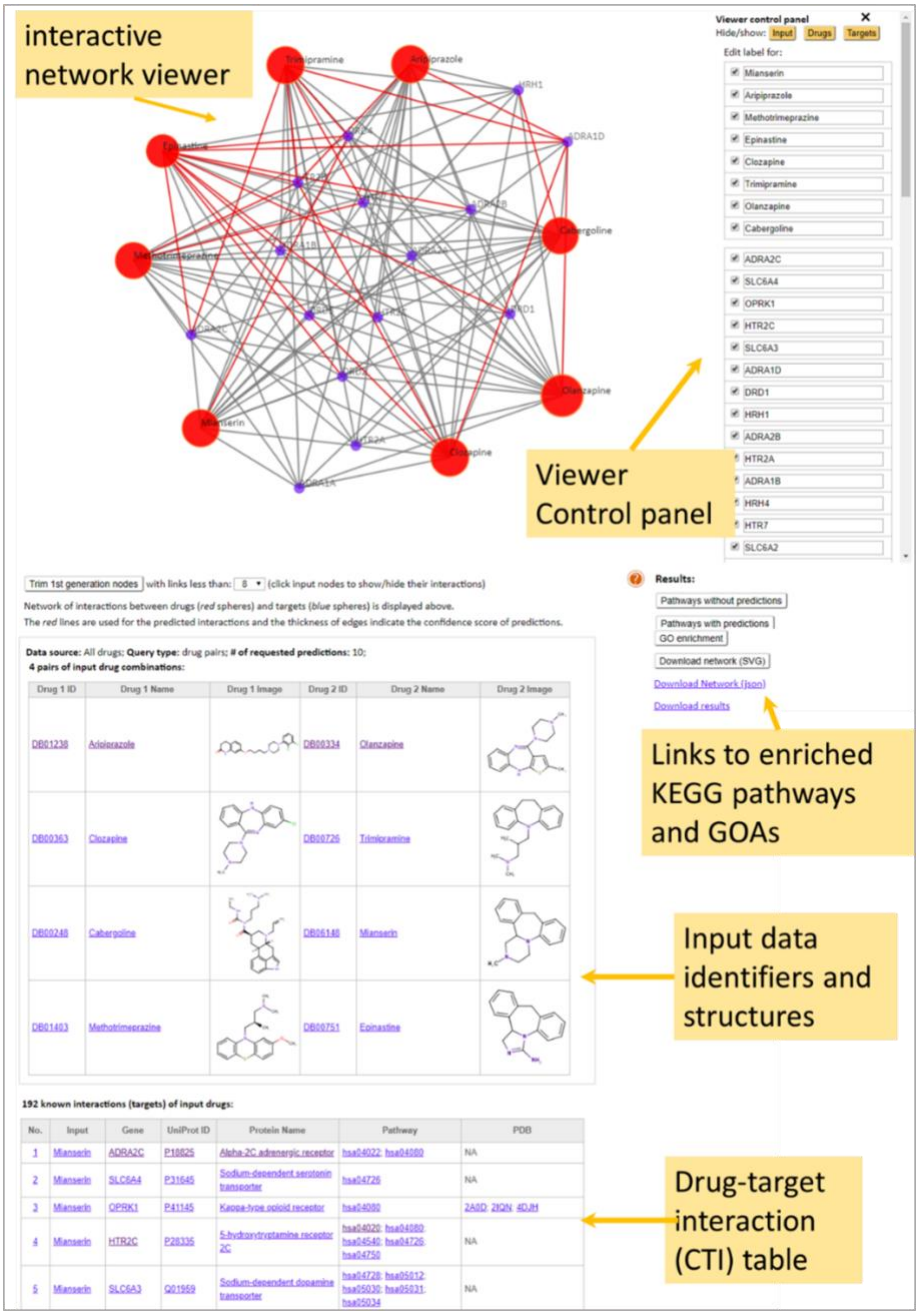


Figure 1.5 A snapshot of the results page from QuartataWeb

Here four pairs of drugs are fed as an input of type II (see text), displayed by the *red spheres* in the interactive network viewer. Their targets are represented as *blue* spheres. Targets nodes with coordination number less than eight are trimmed (hidden) by clicking "*Trim 1st generation nodes"* with a cutoff of eight. The viewer control panel on the *right top* of the page is displayed upon clicking the menu button on the *top*. 2D structures and other information of input drug combinations are presented in a table below the network viewer. All CTIs associated with the input drugs are

listed in a table at the *bottom*. Links to pathways- and GO enrichments-analysis pages are provided, as well as the option of downloading the entire page and data. The figure displays the outputs from QuartataWeb, except for the *yellow labels* and *arrows* included here for facilitating the description.

1.3.3 Drug Repurposing or Identification of Side-Effects

Consider loxapine and its target, α2A adrenergic receptor (gene name: ADRA2A) as an example. CTIs corresponding to both drug (*red sphere*) and target (*blue sphere*) can be viewed in peacock representation (**Figure** 1.3**D**), where known and predicted CTIs being distinguished by the gray and red edges, respectively. Users can interactively select nodes to view primary and 2˚ interactions. Loxapine is an antipsychotic agent approved for treating schizophrenia, whose primary targets are dopamine and serotonin receptors. The figure displays the 2˚ interactions of a serotonin receptor (HTR3C, node colored cyan) which turns out to be a target of many drugs associated ADRA2A, some of which are repurposable. Finally, type III input also permits to identify similar drugs and shared targets as illustrated for the pair doxepin and loxapine (**Figure** 1.3**E**).

1.4 Conclusion

We presented QuartataWeb, an integrated server that offers multiple capabilities for QSP analyses using both known associations and machine-learning predictions. We showed that the interface can help identify repurposable drugs, side-effects, enriched pathways, as well as shared functions, cellular processes and environment for different types of queries. QuartataWeb is

expected to serve as a first filter toward designing more effective phenotypic screens and polypharmacological strategies.

1.5 Materials and Methods

1.5.1 Datasets

DrugBank and STITCH. The outputs released by QuartataWeb are based on LFMs generated by a PMF scheme for three sets of data: approved-drugs in DrugBank 5.1, experimentally verified data in DrugBank 5.1, and experimentally verified data on human in STITCH 5, shortly referred to as *DrugBank-approved*, *DrugBank-all* and *STITCH-experimental*. **Table** 1.2 lists the content of these data sources. STITCH-experimental contains information on more than S = 5.3 million CTIs. This number is only a small percent (0.18%) of the entire space of interactions potentially existing between the N = 315,514 chemicals and M = 9,457 targets contained in STITCH. DrugBank-approved is considerably smaller, with N = 1,883 drugs and M = 2,244 targets. Yet, the occupancy of the interaction space, $S/[N \times M]$, is comparable (0.22%). **Figure** 1.1 display the histograms of the numbers of interactions for chemicals and targets in the STITCH and DrugBank datasets, respectively. Notably, some chemicals and targets have more than 250 interactions in STITCH, and more than 40 in DrugBank.

KEGG. QuartataWeb uses the 28,664 gene-pathway associations between 19,124 human genes and 323 human pathways extracted from KEGG Pathway DB (**Table** 1.2). Proteins are mapped to genes following UniProt (The UniProt, 2017) annotations, and then mapped to

pathways through gene-pathway associations. Based on the known and predicted CTIs, drugs/chemicals can then be connected to affected pathways.

GOA. A total of 2,807 target genes in our datasets were mapped to GO terms, comprising 11,475 molecular functions, 39,150 biological processes, and 15,160 cellular components (**Table** 1.2**)**. The connection between targets and GO annotations is established through UniProt gene identifiers. Likewise, drugs/chemicals are mapped to GOA via the associated known and predicted targets, which enables users to assess the molecular functions, biological processes, and cellular components potentially affected by the chemicals.

1.5.2 Probabilistic Matrix Factorization-based Chemical-Target Interaction Prediction

PMF models the CTI matrix $\boldsymbol{R}_{N \times M}$ between N chemicals and M targets by two lower-rank chemical and target matrices: $\mathbf{U}^{\mathbf{T}}{}_{N \times D}$ and $\mathbf{V}_{D \times M}$, where each chemical is represented by a D-dimensional latent vectors (LVs) $\boldsymbol{u}_i$, and each target is represented by a D-dimensional LV $\boldsymbol{v}_j$. The PMF adopts a probabilistic linear model with Gaussian distribution noise, resulting in the conditional distribution

$$\mathrm{p}(\mathbf{R}|\mathbf{U},\mathbf{V},\sigma^2) = \prod_{i=1}^{N}\prod_{j=1}^{M}\left[\mathcal{N}\left(R_{ij}\middle|\boldsymbol{u}_i^{\boldsymbol{T}}\boldsymbol{v}_{\boldsymbol{j}},\sigma^2\right)\right]^{I_{ij}} \qquad (1.1)$$

where $\mathcal{N}(x|\mu,\sigma^2)$ is the probability density function of the Gaussian distribution with mean μ and variance σ^2, and I_{ij} is the indicator function equal to 1 if chemical i and target j interact with each other, and 0 otherwise. The log-likelihood of **U** and **V** is given by

$$\ln\left(\mathrm{p}(\mathbf{U},\mathbf{V}|\mathbf{R},\sigma^2,\sigma_{\mathrm{u}}^2,\sigma_{\mathrm{v}}^2)\right) = -\frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}\left(R_{ij}-\boldsymbol{u}_i^T\boldsymbol{v}_j\right)^2 - \frac{1}{2\sigma_{\mathrm{u}}^2}\sum_{i=1}^{N}\boldsymbol{u}_i^T\boldsymbol{u}_i - \frac{1}{2\sigma_{\mathrm{v}}^2}\sum_{i=1}^{N}\boldsymbol{v}_j^T\boldsymbol{v}_j + C \quad (1.2)$$

where C is a constant independent of parameters. Maximizing the log-posterior over chemical and target features is equivalent to minimizing the sum of squared errors loss function with quadratic regularization terms:

$$E = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{M} I_{ij}\left(R_{ij}-\boldsymbol{u}_i^T\boldsymbol{v}_j\right)^2 + \frac{\lambda}{2}\sum_{i=1}^{N}\|\boldsymbol{u}_i\|^2 + \frac{\lambda}{2}\sum_{j=1}^{M}\|\boldsymbol{v}_j\|^2 \quad (1.3)$$

To learn an optimal LFM means to find the **U** and **V** matrices that minimize the loss function. Once the optimal **U** and **V** are obtained, the product $\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{V}$ yields the reconstructed CTI matrix $\widehat{\boldsymbol{R}}_{N\times M}$,

$$\widehat{\boldsymbol{R}}_{N\times M} = \mathbf{U}_{N\times D}^{\mathrm{T}}\mathbf{V}_{D\times M} \quad (1.4)$$

each entry of which (other than those already known) represents the *confidence score* for the occurrence of the corresponding CTI. The histograms of confidence scores computed for known and predicted interactions are displayed in **Figure** 1.6.

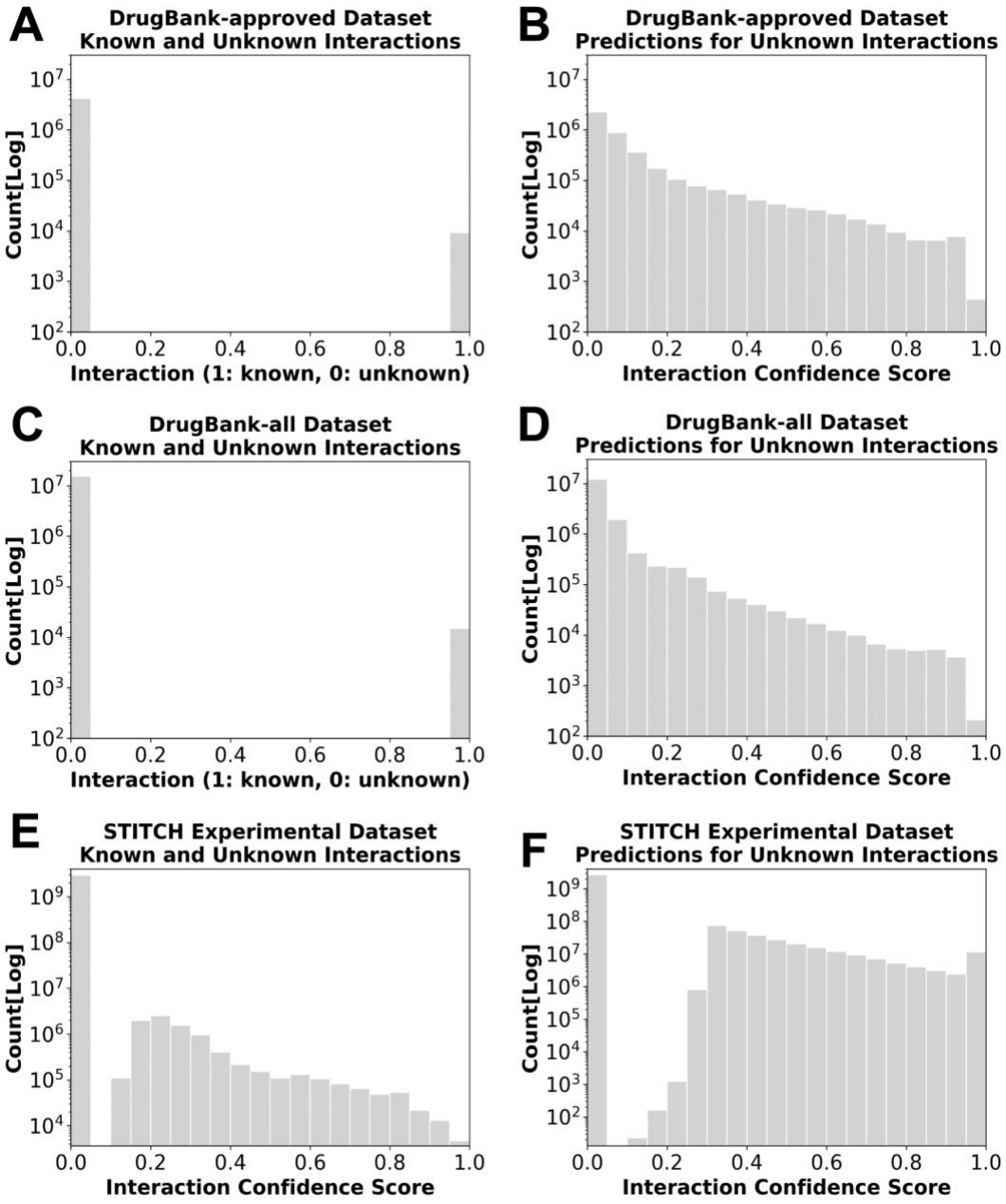


Figure 1.6 Distribution of interaction confidence scores

DrugBank-approved (A-B), DrugBank-all (C-D) and STITCH experimental (E-F). The distributions are presented for known (A, C and E) and unknown/predicted (B, D and F) interactions for the three datasets. The data for the drug-target interactions in DrugBank (A and C) are binary (equal to one or zero) since only known interactions are recorded without confidence scores, hence the bimodal distributions in panels A and C. Similarly, score of zero is assigned to unknown interactions in STITCH dataset (E). The reported confidence scores for the known interactions in STITCH

range from 0 to 1, where a higher score means a higher probability of interaction confirmed in experiments. The confidence scores for the predictions are evaluated using the PMF method with the optimal training parameters listed in Table 1.3.

Table 1.3 Parameters used for training latent factor models in QuartataWeb

Parameters	DrugBank-approved	DrugBank-all	STITCH-experimental
Latent vector dimensionality (D)	50	50	100
Regularization term (λ)	0.01	0.0001	0.5

(a) Latent factor models (LFMs) are evaluated using the method described in the Supplementary Theory and Methods. Details on the evaluation of these optimal parameters can be found on the Theory webpage of QuartataWeb.

1.5.3 Evaluation of Chemical-Target Interaction Prediction Performance

The prediction performance of QuartataWeb has been evaluated using 10-fold cross-validations for DrugBank and STITCH (see **Table** 1.4). The validation dataset is not included in training dataset. As a further test, we hid 70% of known interactions in DrugBank and used the remaining 30% in training. The precisions of the predictions for DrugBank-approved and DrugBank-all datasets are 0.684 and 0.706, respectively, based on top 1,000 predictions (**Figure** 1.7).

Table 1.4 Performance of QuartataWeb observed in cross-validation tests

Dataset	cross-validation	Training set	Test set			Results			
		Size[d]	Size[d]	Positive[e]	Negative[f]	AUC[g]	Sensitivity[h]	Specificity[i]	Precision/ RMSE[j]
DrugBank-approved [a]	10-fold	8,328	1,850	925	925	0.836 ± 0.010	0.582	0.992	0.987 (Precision)
DrugBank-all[b]	10-fold	13,485	2,996	1,498	1,498	0.836 ± 0.009	0.492	0.991	0.984 (Precision)
STITCH-experimen-tal[c]	10-fold	4,828,203	1,072,934	536,467	536,467	0.870 ± 0.015	N/A	N/A	0.0328 (RMSE)

[a] approved drugs and their targets in DrugBank v5; [b] all experimentally verified drugs and targets in DrugBank v5.1; [c]experimentally confirmed chemicals and proteins corresponding to human targets in STITCH v5; [d]Size: number of interactions in the training and validation data sets; [e]Positive: number of known interactions; [f]Negative: number of false interactions; [g]AUC: area under the receiver operating characteristics curve; [h]Sensitivity: TP / (TP + FN); [i]Specificity: TN / (FP + TN); [j]Precision=TP/(TP+FP); root-mean-square error (RMSE) compared to the confidence scores reported in STITCH for known interactions. TP: true positive, FP: false positive, FN: false negative, TN: true negative. Sensitivity and Specificity results are not available for STITCH as the input training data are continuous confidence score instead of binary input. Instead, we use RMSE as metric.

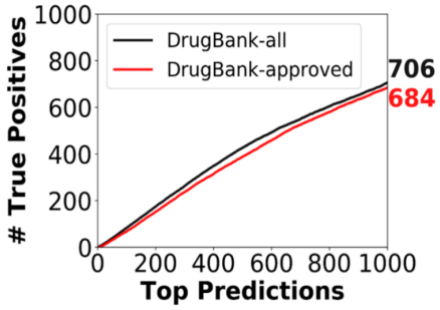


Figure 1.7 Performance of QuartataWeb

Results refer to the approved (red) and all (black) drugs in DrugBank. The abscissa indicates the rank m (1 < m < 1,000) of top-ranking predictions, among all potential CTIs. The total numbers of interactions are [N x M - 0.3S] = 4,222,676 and 15,417,163, in the two respective cases, using the values reported in Table 1.2). The ordinate indicates the average number of recaptured hidden interactions (TPs). For the 1,000 top ranked predictions average precisions of 0.684 and 0.706 are attained in the two respective datasets.

In addition, the top 1,000 predicted drug-target pairs from the LFM generated for the full DrugBank-approved (1,883 drugs and 2,244 targets) and DrugBank-all dataset (5,494 drugs and 2,807 targets) were examined using other CTI DBs STITCH, ChEMBL (Mendez, et al., 2019) and TTD (Li, et al., 2018), respectively. 376 of DrugBank-approved LFM 'predicted' pairs were

actually listed among the experimentally confirmed pairs in STITCH, 255 pairs were reported in ChEMBL, and 14 pairs in TTD, ending up with 459 confirmed pairs (because of the overlaps between those confirmed in different databases). 341 of DrugBank-all LFM 'predicted' pairs were listed in STITCH, 260 pairs in ChEMBL, and 16 pairs in TTD, ending up with 438 confirmed pairs. This demonstrated the predictive power of the current tool. The list of confirmed drug-target pairs, along with their IDs in different databases can be found here. It remains to be seen if the remaining pairs are confirmed in the future.

Other details on PMF-based evaluations are presented in earlier work (Cobanoglu, et al., 2013). PMF models were trained using our in-house MATLAB codes (Cobanoglu, et al., 2013; Cobanoglu, et al., 2015) and the collaborative filtering toolkit GraphChi (Kyrola, et al., 2012), respectively. The parameters adopted in the final LFM for each dataset are listed in **Table** 1.3**.**

1.5.4 Chemical-Target Interaction Pattern-based Similarity

Chemical-chemical (or target-target similarities) based on CTI patterns are calculated by evaluating the correlation cosine between the latent vectors, LVs, generated for drugs (or for targets).

Chemical-chemical or target-target similarities based on CTI patterns are calculated by evaluating the correlation cosine between two chemical LVs ($\boldsymbol{u}_i$ and $\boldsymbol{u}_j$) or two target LVs ($\boldsymbol{v}_i$ and $\boldsymbol{v}_j$) as

$$S_{u_i u_j} = \frac{\boldsymbol{u_i} \cdot \boldsymbol{u_j}}{|\boldsymbol{u_i}||\boldsymbol{u_j}|} \qquad (1.5)$$

$$S_{v_i v_j} = \frac{\boldsymbol{v_i} \cdot \boldsymbol{v_j}}{|\boldsymbol{v_i}||\boldsymbol{v_j}|} \qquad (1.6)$$

The CTI-pattern-based similarities range from -1 to 1, where 1 represents the highest similarity. Histograms of the chemical-chemical or target-target similarities based on CTI patterns for each dataset can be found on the Theory page of QuartataWeb.

1.5.5 Ligand Structure-based Similarity

The *Tanimoto coefficient* T_{ab} between the 2D fingerprints of two chemicals is calculated to provide a metric of the structural similarity between two ligands (Bajusz, et al., 2015) as

$$\boldsymbol{T_{ab}} = \frac{\boldsymbol{a \cdot b}}{\|\boldsymbol{a}\|^2 + \|\boldsymbol{b}\|^2 - \boldsymbol{a \cdot b}} \qquad (1.7)$$

where **a** and **b** represent 2D fingerprint binary vectors, the Tanimoto coefficient ranges from 0 to 1, and 1 is the highest similarity. We generated 2D circular fingerprints based on the Morgan algorithm with feature invariants similar to the FCFP (Rogers and Hahn, 2010) using RDKit (http://rdkit.org).

1.5.6 Enrichment Analysis

We provide enrichment *p*-values to determine if a protein is overrepresented among the targets of a set of chemicals, or if a chemical is overrepresented among the small molecules that target a set of proteins. We also provide enrichment values for pathways and GOAs if a pathway or GO term is enriched in a list of targets. The *p*-values are calculated by the hypergeometric test,

then adjusted by False Discovery Rate (FDR) correction using the Benjamini-Hochberg method (Benjamini, et al., 2001).

The hypergeometric *p-value* (P^A) of an item *A* (e.g. a pathway) in a list of items *Bs* (e.g. targets), is the probability of randomly drawing k_0 or more *Bs* that associate with the evaluated *A*

$$P^A = \sum_{k_0 \le k \le m} \frac{\binom{K}{k}\binom{M-K}{m-k}}{\binom{M}{m}} \qquad (1.8)$$

where *M* is the total number of the background items of type *A*, *m* is the total number of *B*s we identified, and *K* is the number of *Bs* that associated with the evaluated *A*, while k_0 is the number of *Bs* we identified that associated with the evaluated *A*.

For multiple testing, Benjamini-Hochberg method is applied to correct FDR. Giving *T* as the total number of the evaluated *As*, the hypergeometric *p-values* are sorted from smallest to largest, and the adjusted *p-value* of the ith item (p_i^*) is calculated as

$$p_i^* = min_{k=i\ldots m}[\min\left(\frac{p_k T}{i}, 1\right)] \qquad (1.9)$$

The adjustment limits the FDR to a selected cutoff level α (e.g. adjusted *p-value* < 0.05), which indicates that the fraction of false significant *As* among all significant *As* identified is expected to be less than α.

The enrichment score used in the provided enrichment output bar plots is defined as

$$Enrichment\ Score_i = -\log10(p_i^*) \qquad (1.10)$$

where higher enrichment score represents higher significance.

1.5.7 Other Supporting and Visualization Tools

Known and predicted CTIs, KEGG pathways, GO annotations, PDB data and other information from public DBs are constructed in a PostgreSQL DB. Modern JavaScript libraries D3 with force-directed graphs and jQuery were employed for developing the customized interactive network viewers. Users can view results with user-friendly interactive interfaces and/or download tabulated data. The website is compatible with modern browsers (Chrome, Firefox, Microsoft Edge and Safari) and operates in Windows, Linux, MacOS and iOS environments.

1.6 Acknowledgement

The presented work is published as Li, Hongchun, Fen Pei, D. Lansing Taylor, and Ivet Bahar. "QuartataWeb: integrated chemical–protein-pathway mapping for polypharmacology and chemogenomics." *Bioinformatics* 36, no. 12 (2020): 3935-3937. Dr. Hongchun Li contributed to the design and development of the web server interface, Dr. Bing Liu helped modify the web interface. I performed the data curation, database construction, ML model training and evaluation, and server backend support. Dr. Ivet Bahar and Dr. Lansing Taylor supervised the project.

2.0 Predicting Large Scale Protein-Protein Interactions Using Symmetric Logistic Matrix Factorization

Protein-protein interactions (PPIs) play an essential role in enabling and sustaining cellular activities. Their accurate assessment at a systems levels is critical to deciphering disease mechanisms and developing novel drugs. Computational methods to predict PPIs proved useful in complementing expensive experiments and helping reduce false positives, even though with growing PPI data the need for more efficient methods emerged. In this chapter, we propose a novel symmetric logistic matrix factorization (symLMF)-based approach to predict PPIs, especially useful for large PPI networks. The method utilizes data on experimentally confirmed PPIs, projected onto a relatively low-dimensional matrix used for evaluating the pairwise probabilities of PPIs. Benchmarking of predictions against two widely used datasets (*S. cerevisiae*-benchmark and *H. sapiens*-benchmark) demonstrated the utility of the new method which outperformed most of the state-of-the-art methods applied to human PPIs, and exhibited a performance comparable to those of deep learning models despite its conceptual and technical simplicity and efficiency. Comparative benchmarking against large datasets (*S. cerevisiae*-extended and *H. sapiens*-extended) further revealed the higher performance of symLMF compared to other matrix factorization methods. Tests performed on human, yeast, and tissue (brain and liver)- and disease (neurodegenerative and metabolic disorders)-specific datasets after hiding 50% of known interactions showed that 235 to 327 of the top 1000 predictions capture the hidden interactions in those specific databases. Notably, many '*de novo* predictions' made by symLMF are verified to actually exist in other PPI databases that were not used for training/testing the

method. The new method is expected to be of broad utility as a simple and highly efficient, yet effective and accurate, tool for discovering new PPIs, using large-scale data.

2.1 Introduction

PPIs play critical roles in various cellular processes, including signal transduction, immune response, cellular organization, and cell regulation and death. Uncovering new PPIs is of great importance to understanding disease mechanisms and developing novel therapeutic strategies (Scott, et al., 2016; Skrabanek, et al., 2008). A plethora of experimental methods including high-throughput technologies such as yeast two-hybrid screens (Y2H) (Fields and Song, 1989), mass spectrometric protein complex identification (MS-PCI) (Ho, et al., 2002), protein microarrays (Melton, 2004), mammalian protein-protein interaction trap (MAPPIT) (Lievens, et al., 2016), and BioID (Roux, et al., 2018) have been developed for large-scale PPI identification, which generated extensive data compiled in public PPI databases, such as BioGRID (Stark, et al., 2006), STRING (Szklarczyk, et al., 2019), and DIP (Xenarios, et al., 2002). However, these methods are inherently subject to different types of noise and suffer from relatively high false positive and false negative rates (Collins, et al., 2007). Besides, most experimental data are biased toward certain protein types or cellular localizations, providing an incomplete description of the protein-protein interactome. Efficient and robust computational methods, especially those rooted in fundamental theory and concepts of machine learning (ML), emerge as powerful tools for facilitating and accelerating the consolidation of the data of PPIs including the discovery of new interactions.

Over the past decade, various ML algorithms have been developed to predict novel PPIs based on protein sequence (An, et al., 2019; Chen, et al., 2019; Chen, et al., 2019; Guo, et al.,

2008; Huang, et al., 2015; Romero-Molina, et al., 2019; Sun, et al., 2017; Wang, et al., 2017; Wang, et al., 2019; Wang, et al., 2017; Yao, et al., 2019; You, et al., 2013; Zhang, et al., 2014), structure (Johansson-Åkhe, et al., 2019; Sacca, et al., 2014; Zhang, et al., 2012), function annotation (Bandyopadhyay and Mallick, 2017), and evolutionary relationship (Emamjomeh, et al., 2014; Hamp and Rost, 2015; Kamada, et al., 2014; Xu, et al., 2011). Most of them are supervised classification algorithms, where a set of positive/known PPIs and a set of negative entries are used to train the model and predict whether protein pairs interact or not. Specifically, Guo et al. (Guo, et al., 2008) combined a feature representation using autocovariance and support vector machines (SVM) for predicting yeast PPIs; Zhang et al. (Zhang, et al., 2014) used pairwise kernel SVMs to avoid the concatenation of protein features; You et al. (You, et al., 2013) used a combination of principal component analysis (PCA) and ensemble extreme learning machine model to enable better generalization performance and fast learning speed from protein sequence data; and Du et al. (Du, et al., 2014) used many physicochemical or biochemical features of the proteins in a random forest (RF) algorithm to predict PPIs.

These ML methods have enabled efficient predictions of PPIs compared to traditional experimental detection or computational docking methods. Yet, there are still several inherent drawbacks that await to be resolved. First, most approaches rather focus on feature extraction methods while the improvements on prediction accuracy have been limited. Second, the selection of negative cases may also bring errors since most of the negative PPIs are not available in public domains and the lack of observation/report on a PPI does not necessarily mean that those proteins do not interact. In recent years, deep learning methods including stacked autoencoder (Sun, et al., 2017; Wang, et al., 2017), convolution neural networks (Wang, et al., 2019), and feature embedding (Yao, et al., 2019), aimed to tackle the first problem by using data-driven features,

resulting in increased prediction accuracy. However, these methods still suffer from the difficulty in choosing negative samples, and usually include heavy parameter tuning which becomes inefficient for large-scale PPI predictions. In addition, the "black box" nature of deep learning models results in low interpretability of the model and outputs.

Matrix factorization (MF) models present several advantages that may overcome these limitations. MF models have been highly popular in recommender systems due to their simplicity and superior performance, and they have been shown to be efficiently parallelizable and highly scalable to large scale datasets. MF models have been successfully applied to predictions of drug-target interactions (Cobanoglu, et al., 2013; Cui, et al., 2019; Hao, et al., 2017; Shi, et al., 2018; Xia, et al., 2019) and PPIs (Wang, et al., 2013) in previous studies. A major advantage of MF is that it recommends top interactions purely based on known patterns of interactions for each protein, without dependence on physicochemical, structural, or functional features associated with the proteins. In addition, it performs better with large sparse matrices compared to deep learning models, due to its simplicity, flexibility, and scalability.

In this work, we propose a symLMF algorithm to enable efficient and accurate prediction of large-scale PPIs. We evaluated the performance of the method using the widely used *S. cerevisiae* and *H. sapiens* benchmark datasets, and two extended datasets (*S. cerevisiae*-extended and *H. sapiens*-extended) extracted from BioGRID (Stark, et al., 2006) and STRING (Szklarczyk, et al., 2019), as well as four tissue- or disease-specific PPI datasets (brain, liver, neurodegenerative disorders, metabolic disorders) extracted from the Integrated Interaction Database (IID) (Kotlyar, et al., 2019). The results show that symLMF outperforms most classification methods including support vector machines (SVMs) (Guo, et al., 2008; Zhou, et al., 2011), K-nearest neighbor (KNN) (Yang, et al., 2010), principle component analysis (PCA) (You, et al., 2013), random forests (RFs)

(Ding, et al., 2016), and performs comparably well with respect to the latest deep learning methods (Du, et al., 2017; Wang, et al., 2019; Yao, et al., 2019) on the *human*-benchmark dataset. Comparison of the performance of symLMF with those of other matrix factorization models, including symmetric probabilistic matrix factorization (symPMF), symmetric non-negative matrix factorization (symNMF) and non-negative matrix tri-factorization (NMTF) (Wang, et al., 2013) on two extended datasets further reveals the higher accuracy of symLMF. Applications of symLMF to the *S. cerevisiae*-extended, *H. sapiens*-extended, and brain, liver, neurodegenerative disorders and metabolic disorders datasets where half of the interactions were hidden further confirmed the ability of the method to capture hidden data with approximately 280 hits among the top 1,000 predictions in each case.

Finally, the method was trained on the entire *S. cerevisiae*-extended and *H. sapiens*-extended. Compared to the sequence-based interactome prediction program SPRINT (Li and Ilie, 2017), symLMF shows a higher ability to predict potential PPIs, supported by the fact that about half of these top 1,000 predictions are verified to be listed in external PPI databases GPS-Prot (Fahey, et al., 2011) and APID (Alonso-Lopez, et al., 2019).

2.2 Materials and Methodology

2.2.1 Problem Formalization

The set of proteins is denoted as $\boldsymbol{P} = (p_i)_{i=1}^{m}$, where m is the total number of proteins. PPIs are represented by a binary matrix $\boldsymbol{Y} \in \mathbb{R}^{m \times m}$, where each element $y_{ij} \in \{0, 1\}$. If a protein p_i has been experimentally verified to interact with p_j, y_{ij} is set to 1; otherwise it is set to 0. The non-

zero elements of **Y** are "interaction pairs" and regarded as positive observations. The zero elements of **Y** are "unknown pairs", where a pre-defined subset is considered as true negatives in the evaluation process. The object of the study is to predict the interaction probability of a protein-protein pair and subsequently rank the candidate protein-protein pairs according to their predicted probabilities in descending order, such that the top-ranking pairs are predicted to be those most likely to interact, and the personalized top-*n* recommendations are provided for each individual protein.

2.2.2 Data Collection and Datasets Construction

We used eight different datasets for performance evaluation in this study (**Table** 2.1): two benchmark datasets, two extended datasets, and two tissue-specific and two disease-specific datasets. Specifically, *S. cerevisiae*-benchmark was extracted from the database of interacting proteins (DIP, version 2007.02.19) (Xenarios, et al., 2002), and *H. sapiens*-benchmark from the Human Protein References Database (HPRD, v2010.04.13) (Keshava Prasad, et al., 2009; Peri, et al., 2003). Two extended datasets were collected from publicly available PPI databases BioGRID (v2019.03.25) (Stark, et al., 2006) and STRING (v2019.01.19) (Szklarczyk, et al., 2019), and four tissue/disease-specific datasets were collected from the Integrated Interaction Database (IID, v2018.1.1) (Kotlyar, et al., 2019).

We constructed the extended and tissue/disease-specific datasets following the rules suggested earlier (Guo, et al., 2008). Specifically, to construct positive datasets (datasets of interacting protein-protein pairs), the protein pairs with sequence identities of 40% or higher were removed because these are considered as homologous, and protein fragments with < 50 amino acids were removed. For the negative datasets (datasets of non-interacting protein-protein pairs),

the proteins were randomly paired with other proteins in the positive dataset subject to the three requirements: (1) the negative pairs cannot appear in the positive dataset; (2) the number of negative pairs is equal to the number of positive pairs; and (3) the two proteins in a negative pair do not share subcellular localization. For matrix factorization models, negative examples are not required in the training process; the negative datasets were constructed mainly for evaluation purposes. Protein-protein sequence similarity was calculated using the Protr R package (Xiao, et al., 2015), and the cellular component (CC) annotations were from Gene Ontology (GO) Consortium CC terms. Pairs that shared one or more GO CC terms were excluded from the set of negative pairs. Details on the generation of each dataset are presented below, and statistical data for each dataset are shown in **Table** 2.1.

Benchmark datasets. The *S. cerevisiae*-benchmark dataset (Guo, et al., 2008) was collected from the core subset of DIP. It contains 11,188 protein pairs including 5,594 positive and 5,594 negative pairs. The *H. sapiens*-benchmark dataset (Huang, et al., 2015) was collected from HPRD, and consisted of 3,899 positive and 4,262 negative pairs. These two datasets have been widely used for benchmarking state-of-the-art-methods (An, et al., 2019; Chen, et al., 2019; Huang, et al., 2015; Wang, et al., 2017; Wang, et al., 2019; Wang, et al., 2017; You, et al., 2013). The benchmark datasets comprised interactions detected by multiple small-scale screens, and might be biased and only represent a small fraction of the complete PPI networks.

Extended datasets. Since the sizes of the benchmark datasets were limited, we constructed two extended datasets for *S. cerevisiae* and *H. sapiens* based on the overlaps between the PPIs from the BioGRID and STRING databases, namely *S. cerevisiae*-extended, *H. sapiens*-extended. The physical interaction subsets from BioGRID and experimental subsets from STRING were selected, and their overlapping PPIs were used in order to reduce the noise from individual

databases and obtain more reliable PPI data. Following the rules described above, we ended up with 5,142 proteins, that formed 56,316 positive and 56,316 negative pairs in the *S. cerevisiae*-extended dataset, and 14,455 proteins, with 285,618 positive and 285,618 negative pairs in the *H. sapiens*-extended dataset.

Tissue-specific datasets and disease-specific datasets. The human tissue-specific datasets were extracted from IID, where PPIs are assigned to the tissues that predominantly express the proteins. IID has a total number of 4,927,742 PPIs for 18 species, based on three types of evidence: experimental detection (from nine curated databases), orthology, and *in silico* predictions. Only the human subset with experimental support was used as source data to extract tissue-specific PPIs. Most PPIs are annotated with one or more tissues/conditions. We selected 2 tissues (brain and liver) and 2 disease conditions (neurodegenerative and metabolic disorders) as examples to investigate the tissue-specific and condition-specific PPI networks. Positive and negative pairs were generated following the requirements as described above.

Table 2.1 Description of the eight datasets used in the present study and corresponding symLMF model parameters

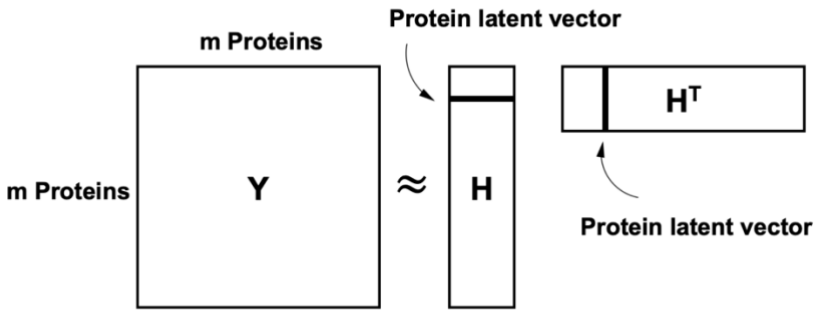
Properties\Dataset	*S. cerevisiae*-benchmark	*H. sapiens*-benchmark	*S. cerevisiae*-extended	*H. sapiens*-extended
Number of proteins	2,526	2,835	5,142	14,407
Number of positive pairs	5,594	3,899	56,316	157,967
Number of negative pairs	5,594	4,262	56,316	157,967
maximal degree	88	71	810	807
median degree	2	2	11	10
sparsity	99.82%	99.90%	99.57%	99.85%
percentage of proteins with one interaction	32.80%	44.84%	9.33%	9.90%
Model parameters (r, λ, c)	60, 10^{-4}, 2	40, 10^{-4}, 1	130, 10^{-6}, 1	110, 10^{-6}, 3
Properties\Dataset	**Brain**	**Liver**	**Neurodegenerative disorders**	**Metabolic disorders**
Number of proteins	11,167	10,627	820	1,063
Number of positive pairs	225,200	218,239	5,881	5,131
Number of negative pairs	225,200	218,239	5,881	5,131
maximal degree	1,838	1,831	191	168
median degree	21	21	7	5
sparsity	99.64%	99.61%	98.25%	99.09%
percentage of proteins with one interaction	2.90%	2.76%	8.78%	14.09%
Model parameters (r, λ, c)	100, 10^{-6}, 1	90, 10^{-6}, 1	90, 10^{-4}, 2	100, 10^{-4}, 2

r: dimensionality of latent vector; λ: regularization parameter; c: weight of positive pairs relative to negative pairs

2.2.3 Symmetric Logistic Matrix Factorization (symLMF)

In this work, we develop PPI prediction models based on a symmetric logistic matrix factorization (symLMF) algorithm. Logistic matrix factorization (LMF) (Johnson, 2014) has been demonstrated to be effective for personalized recommendations. Considering the symmetric nature of **Y** for PPI space, we adopted a symmetric version of LMF, factorizing the observation matrix **Y**

to a lower dimensional matrix $\mathbf{H} \in \mathbb{R}^{m \times r}$, where r ($r << m$) is the number of latent factors, schematically described as



Each protein p_i is described by a latent vector $\boldsymbol{h}_i$, and the interaction probability $p(l_{ij})$ between p_i and p_j is modeled by a logistic function l_{ij} parameterized by the inner product of $\boldsymbol{h}_i$ and $\boldsymbol{h}_j$

$$p\left(l_{ij} \middle| \boldsymbol{h_i}, \boldsymbol{h_j}, \beta_i, \beta_j\right) = \frac{\exp\left(\boldsymbol{h_i h_j^T} + \beta_i + \beta_j\right)}{1 + \exp\left(\boldsymbol{h_i h_j^T} + \beta_i + \beta_j\right)} \tag{2.1}$$

where β_i and β_j represent the biases of proteins i and j that accounts for the variations in interaction behaviors of different proteins. As we can see in Table 2.1 and Figure 2.1, some proteins are highly promiscuous (i.e. they interact with a large number of proteins), while others are more specific (i.e. interact with one or a few proteins). The bias terms are latent factors associated with every protein that are meant to offset these interaction biases. The final logistic function helps constraint the interaction probability between 0 and 1.

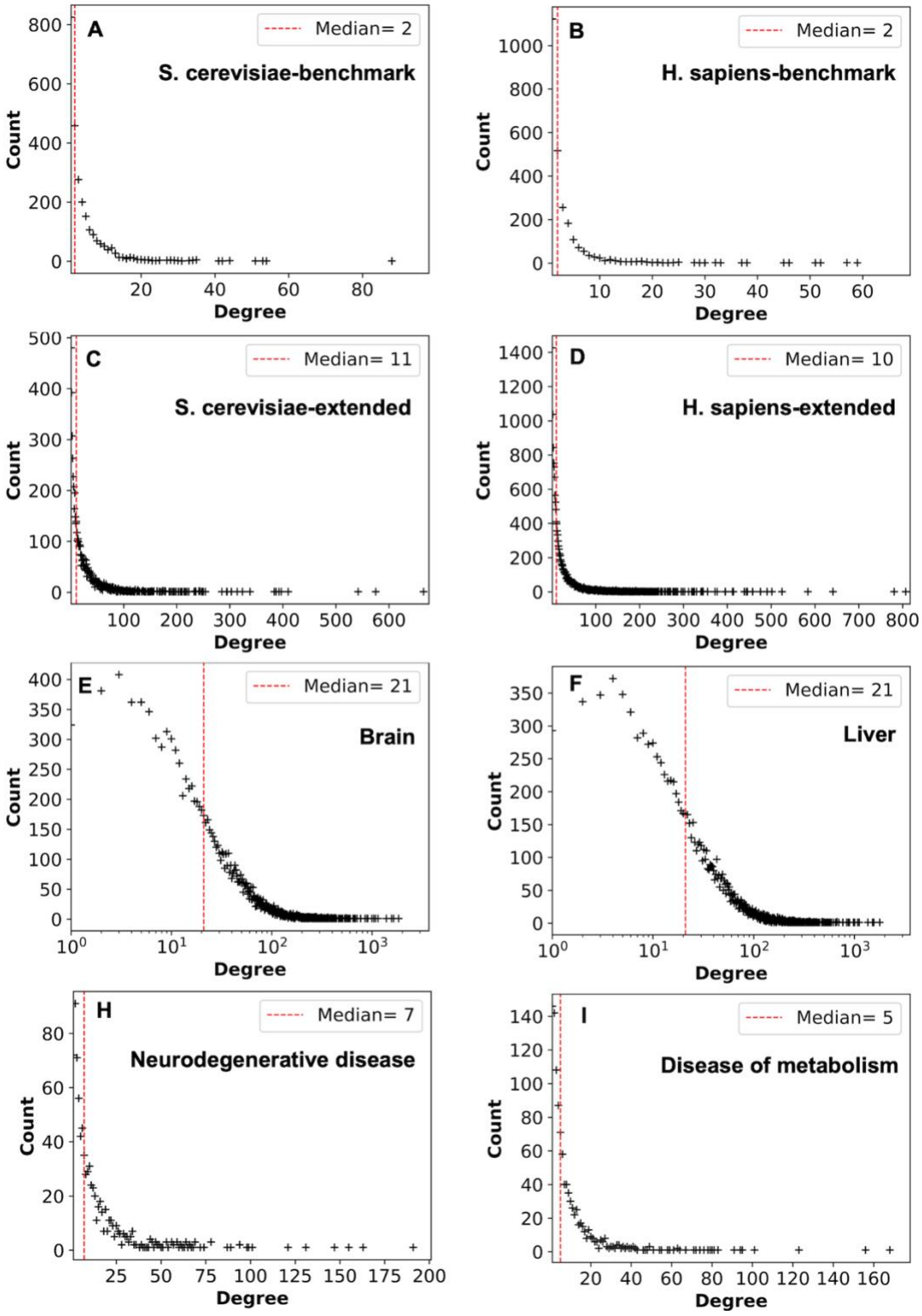


Figure 2.1 Distribution of the protein degrees (number of connections at the nodes, each node representing on protein) in the PPI network described by eight datasets

(A) *S. cerevisiae*-benchmark, (B) *H. sapiens*-benchmark, (C) *S. cerevisiae*-extended, (D) *H. sapiens*-extended, (E) Brain, (F) Liver, (H) Neurodegenerative disease, (I) Metabolic disorders. The x-axis represents the protein degree (number of interacting partners for a given protein), y-axis corresponds to the count number of proteins with a certain degree. The plots show that the degrees of the proteins are right-skewed, where a small number of outliers have extremely large degrees. In this case, the median values are better than the mean values to describe the data averages.

In the PPI datasets, the positive pairs have been experimentally verified, while the negative pairs are sampled examples without experimental verification. In order to ensure more accurate predictions of PPIs, we assign a weight c to each positive example ($c \geq 1$); whereas each negative pair is assigned the weight $c = 1$. Increasing c places more weight on the experimentally verified PPIs. Thus, c serves as a hyperparameter that can be tuned to yield the best results.

By assuming all the training examples are independent, the likelihood of the observations $\mathbf{Y}$ given the parameters $\mathbf{H}$ and β is

$$\mathcal{L}(\mathbf{Y}|\mathbf{H},\boldsymbol{\beta}) = \prod_{i,j} p\left(l_{ij}\middle|\boldsymbol{h}_i,\boldsymbol{h}_{\boldsymbol{j}},\beta_i,\beta_j\right)^{cy_{ij}} \left(1 - p\left(l_{ij}\middle|\boldsymbol{h}_{\boldsymbol{i}},\boldsymbol{h}_{\boldsymbol{j}},\beta_i,\beta_j\right)\right)^{1-y_{ij}} \tag{2.2}$$

We assume zero-mean spherical Gaussian priors with variance σ_i^2 $(= \sigma^2$ for all $i)$ on protein latent factor vectors to help regularize the model and avoid over fitting.

$$p(\mathbf{H}|\sigma^2) = \prod_i \mathcal{N}(\boldsymbol{h}_{\boldsymbol{i}}|0,\sigma_i^2\mathbf{I}) = \prod_i \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\boldsymbol{h}_{\boldsymbol{i}}}{\sigma}\right)^2} \tag{2.3}$$

where $\mathbf{I}$ is the identity matrix of order r. Using Bayesian inference $p(\mathbf{H}\,|\mathbf{Y},\sigma^2,\boldsymbol{\beta}) = p(\boldsymbol{Y}|\boldsymbol{H},\boldsymbol{\beta})p(\boldsymbol{H}|\sigma^2\mathbf{I})$, the log of the posterior distribution is derived as

$$log\ p(\mathbf{H}|\mathbf{Y},\sigma^2) = \sum_{i,j}[cy_{ij}(\boldsymbol{h}_{\boldsymbol{i}}\boldsymbol{h}_{\boldsymbol{j}}^{\boldsymbol{T}} + \beta_i + \beta_j) - (1 - y_{ij} + cy_{ij})\log\left(1 + \exp\left(\boldsymbol{h}_{\boldsymbol{i}}\boldsymbol{h}_{\boldsymbol{j}}^{\boldsymbol{T}} + \beta_i + \beta_j\right)\right) - \frac{1}{2\sigma^2}\sum_i \|\boldsymbol{h}_{\boldsymbol{i}}\|^2\ ] + C \tag{2.4}$$

where C is a constant. Our goal is to learn $\mathbf{H}$ and $\boldsymbol{\beta}$ that maximize the log posterior using the regularization parameter $\lambda = 1/2\sigma^2$. This is equivalent to minimizing the objective function

$$\min_{\boldsymbol{H},\boldsymbol{\beta}}\{ \sum_{i,j}[(1 + cy_{ij} - y_{ij}) \log(1 + \exp(\boldsymbol{h_i h_j^T} + \beta_i + \beta_j)) - cy_{ij}(\boldsymbol{h_i h_j^T} + \beta_i + \beta_j)] + \lambda \sum_i \|\boldsymbol{h_i}\|^2 \} \quad (2.5)$$

which is solved using Adam stochastic gradient descent method (Kingma and Ba, 2014).

2.2.4 Other Related Matrix Factorization Methods

2.2.4.1 Symmetric Probabilistic Matrix Factorization (symPMF)

PMF is a machine learning technique widely used in recommender systems, and has been applied in predicting drug-target interactions (Cobanoglu, et al., 2013), RNA-disease associations (Ha, et al., 2020; Xuan, et al., 2019), drug-disease associations (Yang, et al., 2014), and clustering of microarray data (Dueck, et al., 2005). Its symmetric version, symPMF, directly models the probability of interaction between two proteins without taking the logistic function, as

$$p(l_{ij} | \boldsymbol{h_i}, \boldsymbol{h_j}, \beta_i, \beta_j) = \boldsymbol{h_i h_j^T} + \beta_i + \beta_j \quad (2.6)$$

Like symLMF, symPMF assumes zero-mean spherical Gaussian priors on protein latent factor vectors, such that the objective function becomes:

$$\min_{\boldsymbol{H},\boldsymbol{\beta}} \sum_{i,j}(1 + cy_{ij} - y_{ij})(y_{ij} - \boldsymbol{h_i h_j^T} - \beta_i - \beta_j)^2 - \lambda \sum_i \|\boldsymbol{h_j}\|^2 \quad (2.7)$$

2.2.4.2 Symmetric Non-Negative Matrix Factorization (symNMF)

In the NMF method, the input non-negative matrix is decomposed into two non-negative matrices. The non-negativity constraint of NMF makes the resulting matrices easier to interpret, it is also widely used in clustering problems due to its inherent clustering property. Here we adopted its symmetric version, symNMF, to predict PPIs. The probability of interaction and objective

function of symNMF has the same form as those of symPMF (**Equations 2.6** and **2.7**), with the only difference that the symNMF constraints the latent vectors to be non-negative.

2.2.4.3 Non-Negative Matrix Tri-Factorization (NMTF)

NMTF (Wang, et al., 2013) has been proposed to predict candidate PPIs by decomposeing the symmetric input PPI matrix $\mathbf{Y} \in \mathbb{R}^{m\times m}$ into two low-rank non-negative factor matrices, $\mathbf{H} \in \mathbb{R}^{m\times r}$ and $\mathbf{S} \in \mathbb{R}^{r\times r}$, that approximate the input matrix as $\mathbf{Y} = \mathbf{HSH^T}$. Each row in **H** represents the latent factor vector for a protein, and **S** encodes the latent factor interactions. The corresponding objective function is simply

$$\min_{\mathbf{H}\geq\mathbf{0},\ \mathbf{S}\geq\mathbf{0}} \|\mathbf{Y} - \mathbf{HSH^T}\|^2 \qquad (2.8)$$

Note that symPMF, symNMF and NMTF are MF methods similar to symLMF, with differences mainly residing in the latent factor constraints.

2.2.5 Hyperparameter Selection

The dimensionality r of the latent vectors, the regularization parameter λ, and the positive sample weight c are three hyperparameters that need to be optimized for training the symLMF model for each dataset. Empirically, we varied r from 10 to 150 at intervals of $\Delta r = 10$, λ as [10^{-10}, 10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}], and c as integers from 1 to 5 (inclusive) using a random search. This helped us to narrow down λ to 10^{-6} or 10^{-4} and perform a grid search/screening for r and c. The resulting optimal hyperparameters can be found in the last row corresponding to each dataset in **Table** 2.1.

2.2.6 Cross-Validation and Performance Evaluation

To evaluate the performance of the proposed method, we used five-fold cross-validations, i.e. we randomly divided the dataset into five even subsets, and each subset was selected as a test set while the remaining four were used for training. The whole procedure was repeated 10 times using different dataset distributions to eliminate any bias. The following metrics were used to evaluate the model predictions compared to those obtained by state-of-the-art methods: prediction accuracy, recall, precision, and Matthews correlation coefficient (MCC), defined as

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (2.9)$$

$$Recall = \frac{TP}{TP+FN} \qquad (2.10)$$

$$Precision = \frac{TP}{TP+FP} \qquad (2.11)$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \qquad (2.12)$$

where TP (true positive) is the number of the predicted PPIs found in the positive data set; FP (false positive) is the number of the falsely predicted PPIs that are not actually in the positive dataset; FN (false negative) is the number of noninteracting pairs that are falsely predicted to interact; TN (true negative) is the number of true noninteracting pairs predicted correctly. MCC is a measure of the global quality of binary classification, which is a correlation coefficient between

observed and predicted results. It ranges from -1 to 1, where 0 represents completely random prediction, -1 means consistently wrong prediction and 1 is perfectly accurate prediction. For most recommendation applications of MF, ranking helps us to select the most promising PPIs or suggest the top n binding partners for a certain protein. Therefore, we added four metrics to complement the comparison of MF models: (i) area under the receiver operating curve (AUC), where the receiver operating curve (ROC) plot shows the TP rate plotted against FP rate; (ii) area under the precision-recall curve (AUPR), a plot of precision against recall at different thresholds. Both AUC and AUPR vary from 0 to 1, with 0.5 representing completely random prediction, and 1.0 referring to perfect prediction; and (iii – iv) mean AUC and AUPR, denoted as mAUC and mAUPR, obtained from the averages over the interactions of each individual protein, as well as the predicted results. These two metrics give us the average prediction performance per protein, to make an assessment on the prediction accuracy rate for each protein.

The data and code to generate/reproduce the results reported in this chapter are available at: https://github.com/Fengithub/symLMF-PPI

2.3 Results and Discussion

2.3.1 Performance Comparisons with State-of-the-Art Algorithms on Benchmark Datasets Demonstrate the Superiority of the Proposed Method

We compared the performance of our method to several state-of-the-art methods, in addition to three other MF models (symPMF, symNMF, NMTF).

The methods used with the *S. cerevisiae*-benchmark dataset are as follows: (a) three SVM classifiers, the first two (Guo, et al., 2008) utilizing feature extraction techniques auto cross covariance (ACC) and auto covariance (AC), and the third, local descriptors of protein sequence (Zhou, et al., 2011); (b) a KNN algorithm-based model (Yang, et al., 2010) that also uses local descriptors as features; (c) a combination of PCA with ensemble-extreme-learning-machine- based on protein sequences (You, et al., 2013); (d) two classifiers that use deep neural networks (DNNs) based on protein descriptors (DNN-1) (Du, et al., 2017) and protein sequences (DNN-2) (Li, et al., 2018); and (e) two natural language processing (NLP) methods, Bio2Vec (Wang, et al., 2019) algorithm developed by Wang et al. that converts protein sequences into vectors and uses them in convolution neural networks (CNNs) to model PPIs, and Res2vec (Yao, et al., 2019) where vectors obtained from protein residues are fed as input for downstream deep learning model.

In the case of the *H. sapiens*-benchmark dataset, in addition to the above mentioned DNN-2, CNN-Bio2vec and DNN-Rec2vec methods, three RF models (Ding, et al., 2016) and a weighted sparse representation classifier (WSRC) (Huang, et al., 2015) were evaluated. The RF algorithms use three different methods for extracting protein features, namely multivariate mutual information (MMI), normalized Moreau-Broto Autocorrelation (NMBAC) and a combination of them. WSRC uses discrete cosine transform on substitution matrix representation of the protein sequences to generate features.

To generate comparable results, we used the same five-fold cross-validation as adopted in previous studies, as described in Cross-Validation and Performance Evaluation section. The results are presented in **Figure** 2.2, **Table 2.2** and **Table 2.3**. **Figure** 2.2 displays the accuracy, recall, precision and MCC values obtained with different methods (abscissa) in the respective panels **A-D**, each shown in pairs (for *S. cerevisiae*-benchmark (left), and *H. sapiens*-benchmark (right)

datasets. symLMF results (*red bars*) generally stand out in comparison to other MF methods (symPMF, symNMF, NMTF; *light red bars*) and most of the classification methods that include variations of SVMs, KNNs and PCA-EELM (*blue bars*). symLMF also outperforms the deep learning model DNN-1, however, its performance with respect to the other three deep learning-based methods DNN-2, CNN-Bio2vec and DNN-Res2vec (*green bars*) depends on the dataset, being lower/higher in yeast/human datasets. In fact, in the application to *H. sapiens*-benchmark dataset, symLMF reaches an accuracy percent of **98.88 ± 0.41**, recall of **99.66 ± 0.20** and MCC of **97.76 ± 0.80** (see **Table 2.3**) outperforming all other models based on these metrics. As to precision, WSRC shows the highest performance 99.59 even though that of symLMF (**98.03 ± 0.82**) is close.

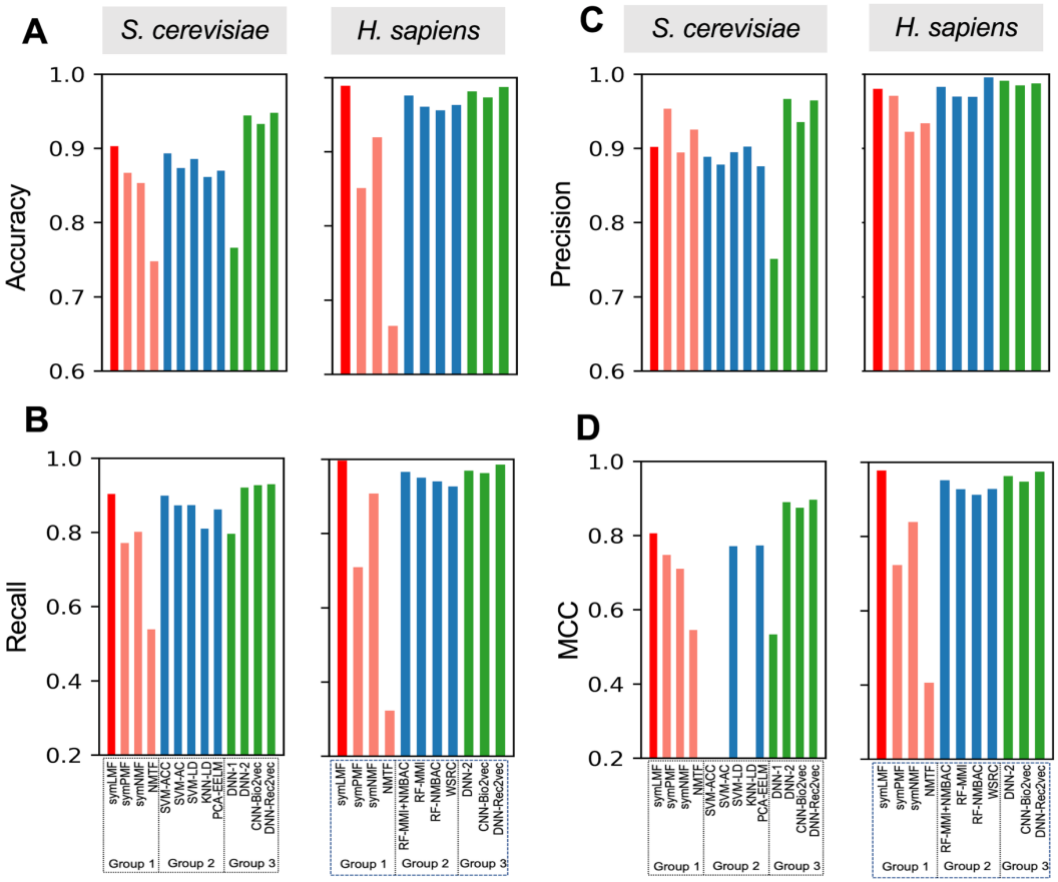


Figure 2.2. Comparison of the performance of different methods tested against two benchmark datasets. Panels A-D show the results (accuracy, recall, precision and MCC, respectively) from 5-fold cross-validations, using different metrics. Each bar refers to a different method, indicated along the lower abscissa. The methods are organized in three groups: MF-based (Group 1; *red/orange bars*), including the proposed symLMF (*red bars*), deep-learning or NN-based methods (Group 3; *green bars*); and others (Group 2; *blue bars*). In each panel, two sets of results are presented, referring to the performance of 13 methods against the *S. cerevisiae*-benchmark dataset (*left*) and the performance of 11 methods against the *H. sapiens*-benchmark dataset (*right*).

Table 2.2 Comparison of the performance of all methods using the *S. cerevisiae-benchmark* dataset

Model	Accuracy (%)	Recall (%)	Precision (%)	MCC (%)
symLMF	**90.30 ± 1.03**	**90.40 ± 1.10**	**90.22 ± 1.03**	**80.61 ± 2.07**
symPMF	86.69 ± 0.99	77.16 ± 1.93	95.34 ± 0.59	74.76 ± 1.78
symNMF	85.33 ± 0.80	80.12 ± 1.38	89.45 ± 0.89	71.06 ± 1.56
NMTF	74.76 ± 0.97	53.88 ± 0.65	92.52 ± 2.34	54.51 ± 1.42
SVM-ACC (Guo, et al., 2008)	89.33 ± 2.67	89.93 ± 3.68	88.87 ± 6.16	N/A
SVM-AC (Guo, et al., 2008)	87.36 ± 1.38	87.30 ± 4.68	87.82 ± 4.33	N/A
SVM-LD (Zhou, et al., 2011)	88.56 ± 0.33	87.37 ± 0.22	89.50 ± 0.60	77.15 ± 0.68
KNN-LD (Yang, et al., 2010)	86.15 ± 1.17	81.03 ± 1.74	90.24 ± 1.34	N/A
PCA-EELM (You, et al., 2013)	87.00 ± 0.29	86.15 ± 0.43	87.59 ± 0.32	77.36 ± 0.44
DNN-1 (Li, et al., 2018)	76.61 ± 0.51	79.63 ± 1.43	75.10 ± 0.66	53.32 ± 1.05
DNN-2 (Du, et al., 2017)	94.43 ± 0.30	92.06 ± 0.36	96.65 ± 0.59	88.97 ± 0.62
CNN-Bio2vec (Wang, et al., 2019)	93.30	92.70	93.55	87.49
DNN-Res2vec (Yao, et al., 2019)	94.78 ± 0.61	92.99 ± 0.66	96.45 ± 0.87	89.62 ± 1.23

Table 2.3 Comparison of the performance of all methods using the *H. sapiens-benchmark* dataset

Model	Accuracy (%)	Recall (%)	Precision (%)	MCC (%)
symLMF	**98.88 ± 0.41**	**99.66 ± 0.20**	**98.03 ± 0.82**	**97.76 ± 0.80**
symPMF	85.10 ± 0.71	70.94 ± 1.50	97.09 ± 0.46	72.30 ± 1.23
symNMF	91.95 ± 0.58	90.79 ± 1.00	92.25 ± 1.20	83.88 ± 1.16
NMTF	66.52	32.22	93.38	40.56
RF-MMI (Ding, et al., 2016)	96.08	95.05	96.97	92.71
RF-NMBAC (Ding, et al., 2016)	95.59	94.06	96.94	91.21
RF-MMI+NMBAC (Ding, et al., 2016)	97.56	96.57	98.30	95.13
WSRC (Huang, et al., 2015)	96.30	92.63	99.59	92.82
DNN-2 (Du, et al., 2017)	98.14	96.95	99.13	96.29
CNN-Bio2vec (Wang, et al., 2019)	97.31	96.28	98.48	94.76
DNN-Res2vec (Yao, et al., 2019)	98.71 ± 0.30	98.54 ± 0.55	98.77 ± 0.53	97.43 ± 0.61

Since the relative ranking is usually used to suggest top predictions in MF recommendation applications, in addition to the above metrics, we added four new metrics for the comparison of matrix factorization models: AUC, AUPR, mAUC, mAUPR. **Figure** 2.3**A-B** and **Table 2.4**, show

that symLMF performs outperforms all MF methods, with regard to all comparison metrics, except precision on the *S. cerevisiae*-benchmark dataset.

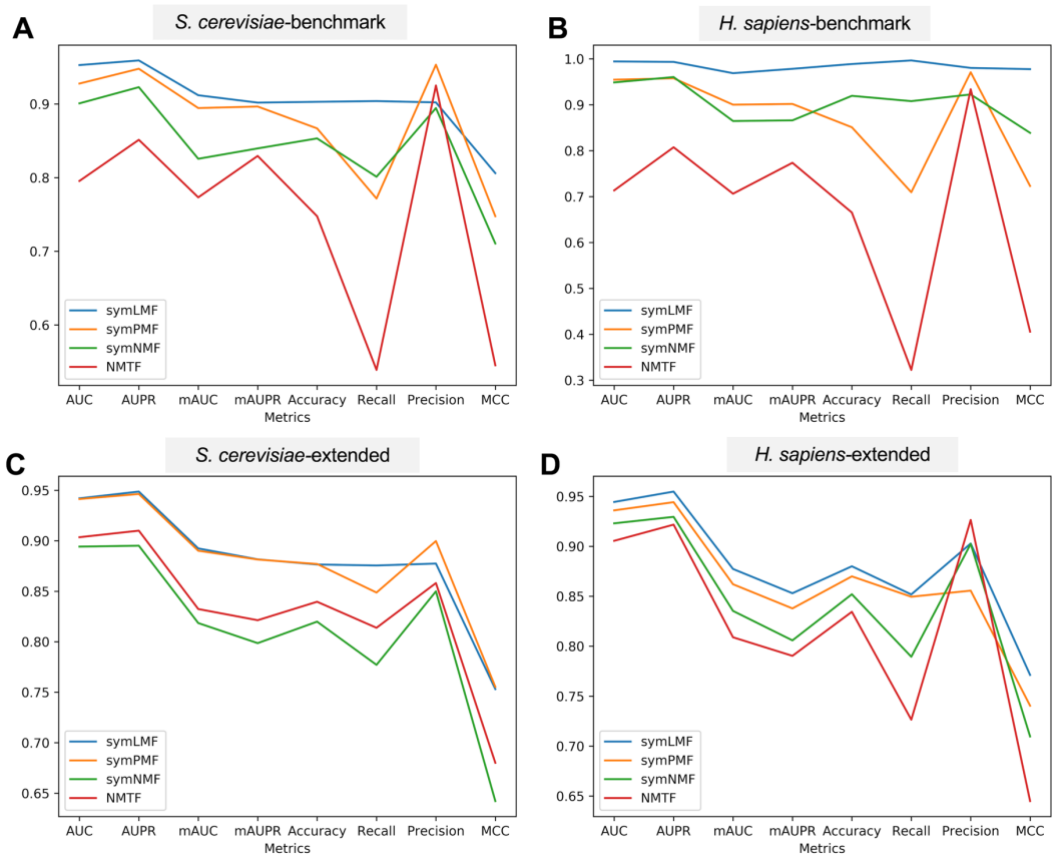


Figure 2.3 Comparison of the performance of MF methods on benchmark and extended datasets

(A) *S. cerevisiae*-benchmark (B) *H. sapiens*-benchmark (C) *S. cerevisiae*-extended (D) *H. sapiens*-extended. The abscissa lists the eight performance metrics, and the ordinate represents the corresponding mean values over five-fold cross-validations.

Table 2.4 Performance of MF models tested againist different benchmarking datasets

Dataset	Model	AUC	AUPR	mAUC	mAUPR
S. cerevisiae-benchmark	**symLMF**	**0.953 ± 0.007**	**0.959 ± 0.008**	**0.912 ± 0.010**	**0.902 ± 0.011**
	symPMF	0.928 ± 0.007	0.948 ± 0.005	0.895 ± 0.009	0.897 ± 0.010
	symNMF	0.901 ± 0.003	0.923 ± 0.004	0.826 ± 0.008	0.006 ± 0.017
	NMTF	0.796 ± 0.008	0.851 ± 0.009	0.773 ± 0.023	0.830 ± 0.021
H. sapiens-benchmark	**symLMF**	**0.994 ± 0.002**	**0.993 ± 0.005**	**0.969 ± 0.016**	**0.978 ± 0.011**
	symPMF	0.955 ± 0.008	0.958 ± 0.008	0.900 ± 0.029	0.902 ± 0.021
	symNMF	0.949 ± 0.002	0.960 ± 0.002	0.865 ± 0.019	0.866 ± 0.017
	NMTF	0.714 ± 0.010	0.807 ± 0.009	0.706 ± 0.025	0.774 ± 0.025
Dataset	**Model**	**AUC**	**AUPR**	**mAUC**	**mAUPR**
S. cerevisiae-extended	**symLMF**	**0.942 ± 0.001**	**0.949 ± 0.001**	**0.892 ± 0.003**	**0.882 ± 0.002**
	symPMF	0.941 ± 0.001	0.946 ± 0.001	0.890 ± 0.001	0.8814 ± 0.002
	symNMF	0.894 ± 0.002	0.895 ± 0.003	0.819 ± 0.002	0.799 ± 0.003
	NMTF	0.904 ± 0.003	0.910 ± 0.003	0.832 ± 0.005	0.821 ± 0.006
H. sapiens-extended	**symLMF**	**0.944 ± 0.001**	**0.955 ± 0.001**	**0.877 ± 0.002**	**0.853 ± 0.003**
	symPMF	0.936 ± 0.001	0.944 ± 0.001	0.862 ± 0.001	0.838 ± 0.002
	symNMF	0.923 ± 0.001	0.930 ± 0.001	0.835 ± 0.002	0.806 ± 0.002
	NMTF	0.906 ± 0.001	0.922 ± 0.001	0.809 ± 0.002	0.790 ± 0.001
Dataset	**Model**	**Accuracy (%)**	**Recall (%)**	**Precision (%)**	**MCC (%)**
S. cerevisiae-extended	**symLMF**	**87.65 ± 0.31**	**87.55 ± 0.22**	**87.74 ± 0.45**	**75.31 ± 0.63**
	symPMF	87.70 ± 0.21	84.87 ± 0.22	89.97 ± 0.29	75.53 ± 0.43
	symNMF	81.99 ± 0.28	77.71 ± 0.43	84.99 ± 0.47	64.22 ± 0.56
	NMTF	83.96 ± 0.28	81.38 ± 0.53	85.81 ± 0.19	68.01 ± 0.54
H. sapiens-extended	**symLMF**	**88.00 ± 0.08**	**85.18 ± 0.18**	**90.28 ± 0.12**	**77.12 ± 0.16**
	symPMF	86.99 ± 0.11	84.95 ± 0.15	85.56 ± 0.14	74.04 ± 0.23
	symNMF	85.20 ± 0.07	78.94 ± 0.31	90.24 ± 0.19	70.96 ± 0.11
	NMTF	83.44 ± 0.09	72.64 ± 0.29	92.64 ± 0.18	64.49 ± 0.16

Taken together, these results demonstrate the strong performance of our proposed symLMF. Although DNN-Res2vec shows higher performance with the yeast (but not human) dataset, the complexity of the model makes it less efficient to train, which may be a challenge for large-scale datasets. Precisely, DNN-Res2vec method includes two training processes: (a) unsupervised representation learning (Res2vec) to generate residue representations, where each protein sequence ends up with a vector representation of 17,000 elements; (b) two separate DNNs with four layers for feature extracting, a merged layer for feature integration, and two layers for

classification. Apparently, MF models are much simpler, resulting in much more efficient training than deep learning models, which is particularly useful on large datasets. We further evaluated the performance of MF models on two extended datasets.

2.3.2 Applications to Extended Datasets Support the Utility of the Proposed Method in Predicting Large-scale PPIs

With continual development of high-throughput technologies, a large number of PPIs have been detected recently, revealing PPI networks that are much larger and more extensive than the two widely used benchmark datasets considered above. Therefore, we constructed two extended datasets by integrating two comprehensive databases, BioGrid and STRING (see Materials and Method section for details). As summarized in **Table** 2.1, the number of proteins in the *S. cerevisiae*-extended dataset is twice as large as that of the *S. cerevisiae*-benchmark dataset, and the number of interactions is higher by 10-fold; the number of proteins in the *H. sapiens*-extended dataset is five times as large as that of the *H. sapiens*-benchmark dataset, and the number of interactions increased forty times. To evaluate the scalability of symLMF, we compared its performance with respect to the other three MF models on these two extended datasets.

As shown in **Figure** 2.3**C-D** and **Table 2.4**, among the four MF models, symLMF ranks first based on all metrics except precision where it ranks second. This detailed analysis further corroborates the superior performance of symLMF over the three other MF methods, and significantly, it establishes its utility as an effective approach for large datasets.

In the next section, we further validate its performance on additional tests with four tissue/disease-specific datasets.

2.3.3 Performance of symLMF on Tissue/Disease-Specific Datasets

To test the predictive ability of our proposed method symLMF on different datasets, we applied it to four tissue/disease-specific datasets. In this experiment, we selected protein-protein interactomes for brain, liver, metabolic disorders, and neurodegenerative disorders as examples of tissue/disease-specific datasets. As described in **Table** 2.1, the numbers of proteins in the brain and liver datasets are significantly high (comparable to that of *H. sapiens*-extended dataset) while the population of proteins with one interaction only is much lower, resulting in a high number of interactions (positive pairs) compared to all four datasets used above. As to the metabolic and neurodegenerative disorders datasets, they are smaller than all others in terms of the number of proteins/nodes in the PPI network even though their connectivity (number of positive pairs) is comparable to those of the yeast and human benchmark sets, due to the higher degree of the individual nodes.

Table 2.5 summarizes the detailed performance of symLMF on these tissue/disease-specific datasets. The AUC values are 0.9515 (brain), 0.9523 (liver), 0.9107 (metabolic disorders) and 0.9407 (neurodegenerative disorders), and accuracy values are 88.66% (brain), 88.78% (liver), 81.37% (metabolic disorders) and 86.11% (neurodegenerative disorders). Compared to the *H. sapiens*-extended dataset, the brain, liver, and neurodegenerative disorders datasets exhibit a similar or slightly better overall performance compared to the metabolic disorders dataset. These results further validate that symLMF performs well on extended, tissue-specific, and disorder-specific datasets.

Table 2.5 Performance of symLMF on tissue/disease-specific datasets

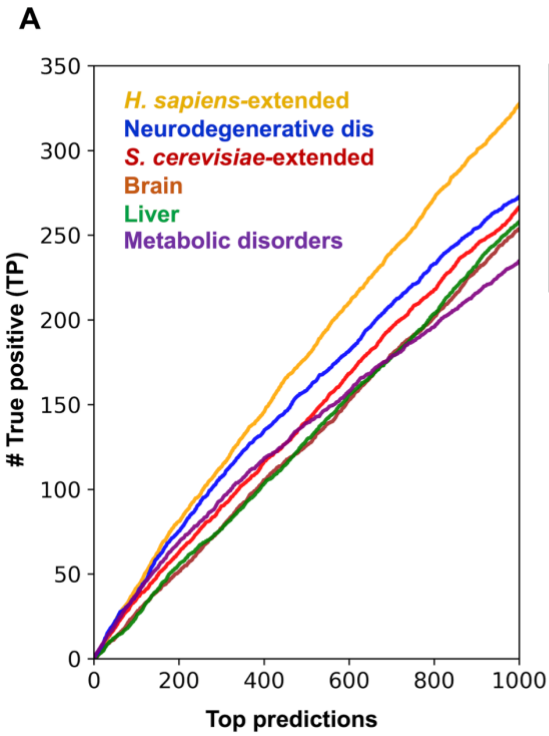
Dataset	AUC	AUPR	mAUC	mAUPR
H. sapiens-extended	0.944 ± 0.001	0.955 ± 0.0008	0.877 ± 0.0024	0.853 ± 0.0025
Brain	0.952 ± 0.0005	0.957 ± 0.0006	0.901 ± 0.0019	0.862 ± 0.0032
Liver	0.952 ± 0.0004	0.958 ± 0.0004	0.903 ± 0.0008	0.864 ± 0.0024
Neurodegenerative disorders	0.941 ± 0.0052	0.952 ± 0.0039	0.897 ± 0.0077	0.885 ± 0.0117
Metabolic disorders	0.911 ± 0.0058	0.926 ± 0.0054	0.860 ± 0.0074	0.863 ± 0.0130
Dataset	**Accuracy (%)**	**Recall (%)**	**Precision (%)**	**MCC (%)**
H. sapiens-extended	88.00 ± 0.08	85.18 ± 0.18	90.28 ± 0.12	77.12 ± 0.16
Brain	88.66 ± 0.05	87.48 ± 0.09	89.59 ± 0.07	77.33 ± 0.10
Liver	88.78 ± 0.07	87.52 ± 0.09	89.78 ± 0.11	77.58 ± 0.14
Neurodegenerative disorders	86.11 ± 1.05	90.44 ± 0.81	83.24 ± 1.28	72.49 ± 2.07
Metabolic disorders	81.37 ± 1.04	88.01 ± 1.30	77.70 ± 1.07	63.31 ± 2.07

2.3.4 Ability of symLMF To Recapitulate Hidden Protein-Protein Interactions with Limited Data

As a more stringent test, 50% of the known interactions in each of the datasets (*S. cerevisiae*-extended, *H. sapiens*-extended, brain, liver, neurodegenerative disorders, metabolic disorders) were randomly hidden; and the resulting interaction matrix was used to predict the hidden interactions using symLMF. The predicted interactions were rank-ordered by their confidence score, and each of the top 1,000 predictions was checked to assess whether it is a TP (a hidden known interaction) or a FP (or an interaction not present in the original dataset). Note that the method gives us a lower bound for precision (TP/ (TP + FP)) because the predictions are labeled as TP only if they are annotated in our source datasets, although they can be true but not yet observed experimentally or annotated in the datasets. The experiments were repeated 5 times with different randomly selected hidden parts, and in each simulation, the model was trained 10 times with different random initializations, resulting in overall 50 runs per dataset.

Figure 2.4, presents the results averaged over these independent runs for each dataset. The curves in panel **A** display the number of TPs captured as a function of size of the predicted (rank-ordered) interactions. The panel **B** lists the fraction of TPs and recall for each dataset, compared to random predictions. We note that 327 hidden interactions are captured among the 1,000 predictions for *H. sapiens-extended*, i.e. the precision is 0.327, and becomes the lowest in the case of metabolic disorders (0.235). Compared to random predictors symLMF yields an 818-fold improvement in this case. The performance for tissue/disorder specific predictions is lower. Yet, the TPs or recall percentage are 38-fold (neurodegenerative disorders) and 51-fold (metabolic disorders) enhanced over random.

These results permit us to draw two conclusions. First, a precision of 28 ±5% is attained in the top 1,000 predictions upon adopting a symLMF method for identifying hidden/unknown interactions in a sparse dataset (half of the original data is hidden), irrespective of the size of the interaction space. As we can see, the interaction space of *S. cerevisiae*-extended, *H. sapiens*-extended, brain or liver datasets are about 2 orders of magnitude larger those of the neurodegenerative- and metabolic-disorders datasets. Second, the symLMF method outperforms a random predictor by 38 to 818 folds at the level of top 1,000 predictions. Notably, the enhancements over random increases exponentially with the size of the dataset, as illustrated in panel **C**. This increase in predictive power with increasing number of proteins supports the utility of the method in the applications to large datasets.



B

Dataset \ Results	TP		Fold change
	symLMF	random	symLMF/random
S. cerevisiae-extended	266.6	1.2	222
H. sapiens-extended	327.4	0.4	818
Brain	253.6	1.4	181
Liver	257.8	2.8	92
Neurodegenerative dis	272.8	7.2	38
Metabolic disorders	234.8	4.6	51

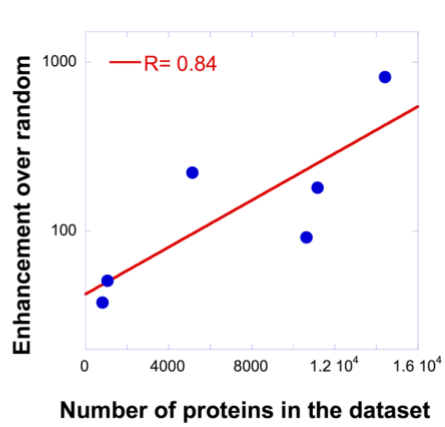


Figure 2.4 Ability of the symLMF to recapitulate the hidden protein-protein interactions in the top 1,000 predicted pairs and improved performance with increasing size of dataset

(A) Results from *in silico* experiments performed by randomly hiding 50% of entries in each dataset (see text). The ordinate shows the number of hidden interactions captured as a function of the number of $1 < m < 1{,}000$ rank-ordered predictions (x-axis) for the six different datasets. The curves are color-coded by the corresponding datasets: *S. cerevisiae*-extended (*red*), *H. sapiens*-extended (*orange*), brain (*brown*), liver (*green*), neurodegenerative disorders (*blue*), and metabolic disorders (*purple*). (B) True positives and enhancement with respect to random tabulated for the six datasets; (C) Increase in enhancement over random as a function of the size (number of proteins) of the datasets.

2.3.5 *De Novo* Predictions of Protein-Protein Interactions

As a final test, we used our method to predict new (potential) protein-protein interactions after training with the whole dataset of *S. cerevisiae*-extended and *H. sapiens*-extended. The final

models were trained using the optimal hyperparameters, with 20 repeated runs with random initialization. We selected the top 1,000 predicted protein-protein pairs obtained from each model, and cross-checked the possible occurrence of these PPIs in recent integrated resources, mainly the web-based PPI platform GPS-Prot (updated on Feb 2019) (Fahey, et al., 2011) and a comprehensive protein interactome database APID (Alonso-Lopez, et al., 2016). The GPS-Prot houses different HIV-host interaction datasets as well as PPIs between human proteins derived from six publicly accessible databases, MINT (Licata, et al., 2012), BioGRID (Stark, et al., 2006), DIP (Xenarios, et al., 2002), IntAct (Kerrien, et al., 2007), MIPS (Mewes, et al., 2006) and HPRD (Keshava Prasad, et al., 2009). GPS-Prot assigns a score for each PPI based on the number of independent publications supporting the PPI, and the reliability of the related experimental techniques. We examined the scores from GPS-Prot for the top 1,000 predictions from our *H. sapiens-extended* model (yeast PPIs are not included in GPS-Prot). APID is also a comprehensive database of PPIs obtained from several of the abovementioned databases including BioGRID, DIP, HPRD, IntAct and MINT. APID offers a pipeline to identify the PPIs with "experimental evidences". These were used to evaluate our top predictions made for both *S. cerevisiae*-extended and *H. sapiens*-extended datasets.

As illustrated in **Figure** 2.5**A**, among the top 1,000 predictions made for the *S. cerevisiae*-extended dataset (*symLMF-yeast*), 498 are found in APID along with corresponding experimental evidences; and among the top 1,000 predictions from *H. sapiens*-extended model (*symLMF-human*), a total number of 559 PPIs are found in either GPS-Prot or APID, mainly 538 in GPS-Prot and 512 in APID, 491 of which are shared between the two datasets. The hit ratio of symLMF-yeast model and symLMF-human model reaches 0.498 and 0.559 respectively, supporting the ability of symLMF models to make new predictions.

Besides that, we also compared symLMF with the sequence-based PPI prediction program SPRINT (Scoring PRotein INTeractions) (Li and Ilie, 2017), which can effectively predict the entire human interactome. We extracted the top 1,000 predictions of SPRINT using the human dataset from BioGRID (*SPRINT-human*), and found that 124 of them appear in GPS-Prot, and 118 in APID with experimental evidences, ending up with 159 PPIs in total (due to 83 shared PPIs). Our symLMF method thus yields a 3.5-fold improvement over SPRINT, in terms of the PPIs verified among the top 1,000 predictions for the human interactome. This result further highlights the importance of learning from protein-protein interaction patterns (as in symLMF) rather than sequence-based properties (as in SPRINT) for accurate assessment of potential PPIs.

It was interesting to observe that the set of 1,000 PPIs predicted by symLMF and SPRINT were somewhat different and involved in different pathways. Among the top 1,000 PPIs predicted by symLMF and SPRINT, only 8 overlapped (**Figure** 2.5**B**). The two PPI lists contained 370 and 378 proteins respectively, with 85 overlapping proteins (**Figure** 2.5**C**). The top 20 enriched pathways obtained upon mapping them to KEGG (Kanehisa, et al., 2017) human pathways is shown in **Figure** 2.5**D**. Among them, only four (ubiquitin-mediated proteolysis, HIV-1 infection, pathways in cancer, and human cytomegalovirus infection) are shared between symLMF and SPRINT predictions. The most enriched pathways deduced from symLMF and SPRINT predictions show different compositions: symLMF is dominated by pathways from immune system (6/20) and infectious diseases (5/20), while SPRINT is dominated by infectious diseases (8/20) and cancer (6/20) pathways.

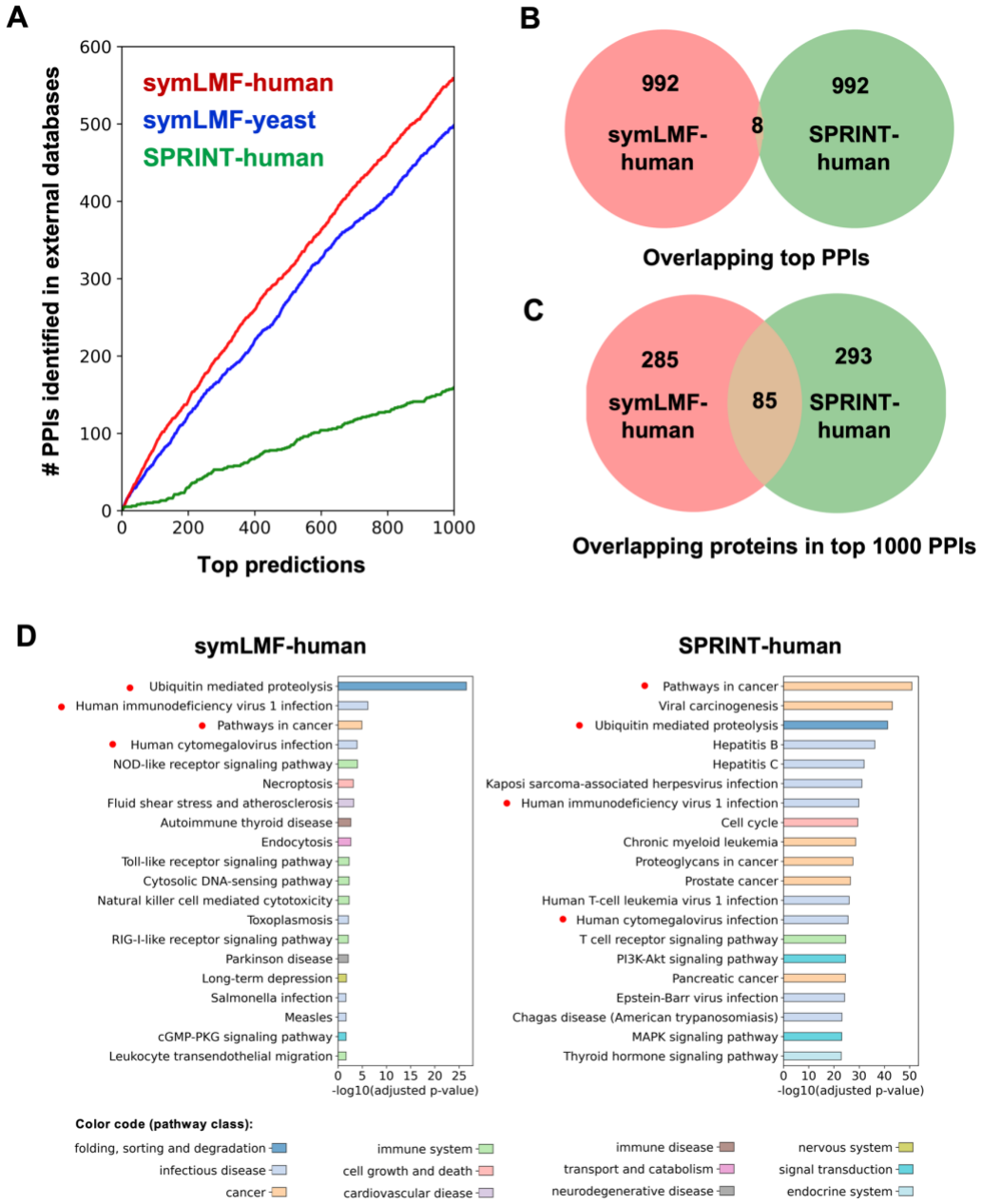


Figure 2.5 Comparison of the top 1,000 predictions between symLMF and SPRINT models

(A) The number of PPIs identified in external databases among the top 1,000 predictions for symLMF and SPRINT models. The ordinate shows the number of PPIs identified in external databases as a function of the number of $1 < m < 1{,}000$ rank-ordered predictions (*x*-axis) for the three different models, symLMF-human (*red*), symLMF-yeast (*blue*), SPRINT-human (*green*). (B) Overlapping predicted PPIs between the top 1,000 PPIs predicted by symLMF-human and SPRINT-human. (C) Overlapping proteins corresponding to the proteins in top 1,000 predictions from symLMF-human and SPRINT-human. (D) Most enriched 20 pathways corresponding to the top 1,000 PPIs predicted by

symLMF-human (*left*) and SPRINT-human (*right*) Y-axis lists the top 20 enriched pathways, x-axis is the corresponding -$\log_{10}$(adjusted p-value). Four pathways marked by *red circles* overlap between symLMF-human and SPRINT-human. Pathways are color-coded by different biological functions shown in the lower key.

2.4 Conclusion

A plethora of ML models have been developed in the last two decades for identifying PPIs, many of them being sequence-based binary classifiers. The major challenge in sequence-based methods has been to find a suitable way to extract features from protein sequence. Pioneering studies include SVM classifiers using a conjoint triad feature extraction method (Shen, et al., 2007), or auto-covariance (AC) with predefined physical chemical properties (Guo, et al., 2008). Recently, embedding techniques including Res2vec (Yao, et al., 2019) derived from natural language processing have been developed, aiming at learning protein representations from raw sequence data. In this work, we tackled the problem from a matrix completion perspective by proposing a symLMF-based methodology, which decomposed the observed PPI matrix into low-dimensional protein latent factors, without using the protein's sequence or structural information. The method is simple to implement, highly scalable, and has been shown to outperform many complex feature extraction-based classification approaches. Importantly, it can be advantageously used for large-scale PPI predictions for entire interactomes.

The applicability of the proposed method to large datasets is worth further attention, given that completing the whole interactome is important in understanding the mechanisms of PPI networks or corresponding biological processes. When training on benchmark datasets with limited number of proteins and interactions, deep learning models such as CNN-Bio2vec, DNN-

Res2vec were able to construct better models than symLMF upon using sequence information. However, when it comes to large datasets like the extended datasets we generated, modeling on 10^4 sequences and 10^5 interactions with deep learning becomes too expensive computationally. Our study demonstrated that a simple symLMF model, without using sequence or structure data, but simply known interaction patterns, can predict large-scale PPIs efficiently and accurately, even with very sparse input. The method is applicable to PPIs specific to particular tissues or diseases, as demonstrated by the applications to PPIs involved in the brain, liver, neurodegenerative disorders and metabolic disorders.

Another benefit of the new method is that it releases an estimate on the interaction probability of a protein pair, which enables recommending the most promising binding partners for a target protein of interest. A major application of it is to generate testable hypotheses, which can be further utilized in drug discovery for disease conditions. As demonstrated in our study, the top-ranking predictions exhibit a high precision, which are likely to provide useful guidance for experimental tests and help save time and cost. Finally, while we focused on the introduction of a new methodology, symLMF, and demonstrating its application to predicting binary PPIs, symLMF can be extended to analyzing ternary interactions and predicting such complex interactions, provided that there are sufficient training data to develop such an extension.

2.5 Acknowledgements

The design, implementation and analysis of the project are mainly conducted by me. Qingya Shi and Haotian Zhang contributed to the analysis of the results, the project was supervised under Dr. Ivet Bahar.

3.0 Application of Quantitative Systems Pharmacology Methods to three complex diseases/disorders

In the previous chapters, we presented several QSP models, methods, and tools that bridge drugs/chemicals to pathways and cellular functions through their targets. In the present chapter, I will discuss three applications of QSP towards better understanding the disease mechanisms and/or proposing novel drug candidates for complex diseases. These diseases share common challenges: 1. they exhibit heterogeneities in populations and disease stages, 2. the underlying molecular mechanisms that drive the disease progression are not fully understood, 3. existing treatments are ineffective due to the complexity of the underlying cellular networks. Therefore, there is an urgent need for understanding the molecular mechanisms that underlie these diseases toward designing novel preventive or therapeutic strategies. The rapidly accumulating data as well as advances in machine learning (ML) methods and computing technology discussed in the previous chapters present an opportunity to systematically mine existing data and draw inferences on potential new strategies.

In the first study (Pei, et al., 2017), we report the implementation of QSP to Huntington's disease (HD), with the application of a chemogenomics platform to identify strategies to protect neuronal cells from mutant huntingtin induced death. Using the $STHdh^{Q111}$ cell model, we investigated the protective effects of small molecule probes having diverse canonical modes-of-action to infer pathways of neuronal cell protection connected to drug mechanism. Several mechanistically diverse protective probes were identified, most of which showed less than 50% efficacy. Specific combinations of these probes were synergistic in enhancing efficacy. Computational analysis of these probes revealed a convergence of pathways indicating activation

of PKA. Analysis of phospho-PKA levels showed lower cytoplasmic levels in *STHdh*^{Q111} cells compared to wild *STHdh*^{Q7} cells, and these levels were increased by several of the protective compounds. Pharmacological inhibition of PKA activity reduced protection supporting the hypothesis that protection may be working, in part, through activation of the PKA network. This systems-level procedure can be broadly applied to any discovery strategy involving small molecule modulation of disease phenotype.

In the second study (Pei, et al., 2019), we carried out a comprehensive analysis of the cellular pathways implicated in a diverse set of 50 drugs of abuse using QSP methods. The analysis of the drug/ligand-target interactions compiled in DrugBank and STITCH databases revealed 142 known and 48 newly predicted targets, which have been further analyzed to identify the KEGG pathways enriched at different stages of drug addiction process, as well as those implicated in cell signaling and regulation events associated with drug abuse. Apart from synaptic neurotransmission pathways detected as a common upstream signaling module that 'senses' the early effects of drugs of abuse, pathways involved in neuroplasticity are distinguished as determinants of neuronal morphological changes. Notably, many signaling pathways converge on important targets such as mammalian target of rapamycin (mTOR) complex 1 (mTORC1). The latter is proposed to act as a universal effector of the persistent restructuring of neurons in response to continued use of drugs of abuse.

In the third study, we analyzed the gene expression profiles of a cohort of NAFLD patients, and identified the genes and pathways that are essential for stages of NAFLD progression defined by pathology reads (steatosis, inflammation, fibrosis). Based on the gene signature associated with NAFLD progression, we integrated the data from connectivity map (CMap) (Lamb, et al., 2006; Subramanian, et al., 2017) and results from network proximity analyses (Guney, et al., 2016) to

propose repurposable drug candidates for NAFLD. Instead of targeting a specific molecular target involved in NAFLD, our proposed drugs target multiple proteins that are involved in one or multiple networks dominating different stages of NAFLD. These mechanistically diverse drugs could serve as probes with the potential to be repurposed singly or in combinations for NAFLD treatment, and tested in the human liver microphysiological system (MPS) (Taylor, et al., 2019).

3.1 Huntington's Disease: Connecting Neuronal Cell Protective Pathways and Drug Combinations

3.1.1 Introduction

HD is a neurodegenerative disease characterized by personality changes, generalized motor dysfunction, and mental deterioration. Symptoms generally develop in the third to fifth decade of life, and the disease ends in dementia and death. HD is rare, affecting 4 to 10 cases in 100,000 people, yet its pathology is strikingly similar to other more common and complex neurodegenerative diseases including Parkinson's and Alzheimer's diseases. HD displays an autosomal-dominant inheritance and an abnormal extension of the number of glutamine repeats at the N-terminus of a single protein (huntingtin, HTT) (Zuccato, et al., 2010). Mutant HTT (mHTT) has been shown to satisfy Koch's postulates for causing this devastating neurological disorder in which striatal neuronal subtypes exhibit particular but not exclusive vulnerability (Zuccato, et al., 2010).

HTT (and mHTT) is a large protein that interacts with many binding partners (Clabough, 2013), and a number of key pathogenic mechanisms have been described in HD, including aberrant

caspase activation, mitochondrial dysfunction (Chen, et al., 2000; Ona, et al., 1999; Wang, et al., 2003; Wang, et al., 2008; Yano, et al., 2014), ER stress, transcriptional dysregulation, altered calcium signaling, proteasome inhibition, defects in vesicle transport, and altered neurotransmitter release and activity (Chen, et al., 2000; Ona, et al., 1999; Zuccato, et al., 2010). However, despite knowledge of the causal gene, and the existence of multiple rodent models that recapitulate key molecular, cellular, and behavioral phenotypes of the human disease (Zuccato, et al., 2010), drug-like molecules that can reduce mHTT protein expression, increase its clearance, or prevent mutant HTT-induced cell death have yet to be successfully identified in clinical trials. The slow progress toward effective therapy has been attributed to an insufficient knowledge of those biological functions of mHTT that are critical in HD. Furthermore, resulting pleiotropic effects have made it difficult to distinguish whether particular aspects of mHTT-associated dysregulation are actually mechanistically linked to disease progression (i.e., pathogenic), epiphenomena, or disease-ameliorating compensatory effects.

Treating HD, or any complex disease, requires a thorough understanding of its mechanisms of progression. Identifying disease mechanisms is hindered by epistasis, pleiotropy and heterogeneity (Gough, et al., 2017), all of which are intrinsic and often confounding characteristics in complex diseases (Chakravarti, et al., 2013). An attractive path to systematically understanding mechanisms of disease progression is QSP, an approach that integrates and iterates computational and experimental methods to determine molecular pathogenesis (Perez-Nueno, 2015; Stern, et al., 2016). A chemogenomics component of QSP involves perturbing disease phenotypes in clinically relevant assays with mechanistically annotated compounds, and using the known mode-of-action of active compounds to infer cellular pathways that are related to the disease and its modulation (see **Figure** 3.1). Concordance in the perturbation of a disease phenotype among a set of

structurally diverse chemical probes sharing an annotated common mechanism can provide compelling evidence for the role of a particular target/pathway in the molecular etiology (Wagner and Schreiber, 2016). In turn, a discordance with such a probe set could lead to the identification of a novel disease-specific mechanism. This finely tunable pharmacological approach is complementary to genetic approaches (Wagner and Schreiber, 2016).

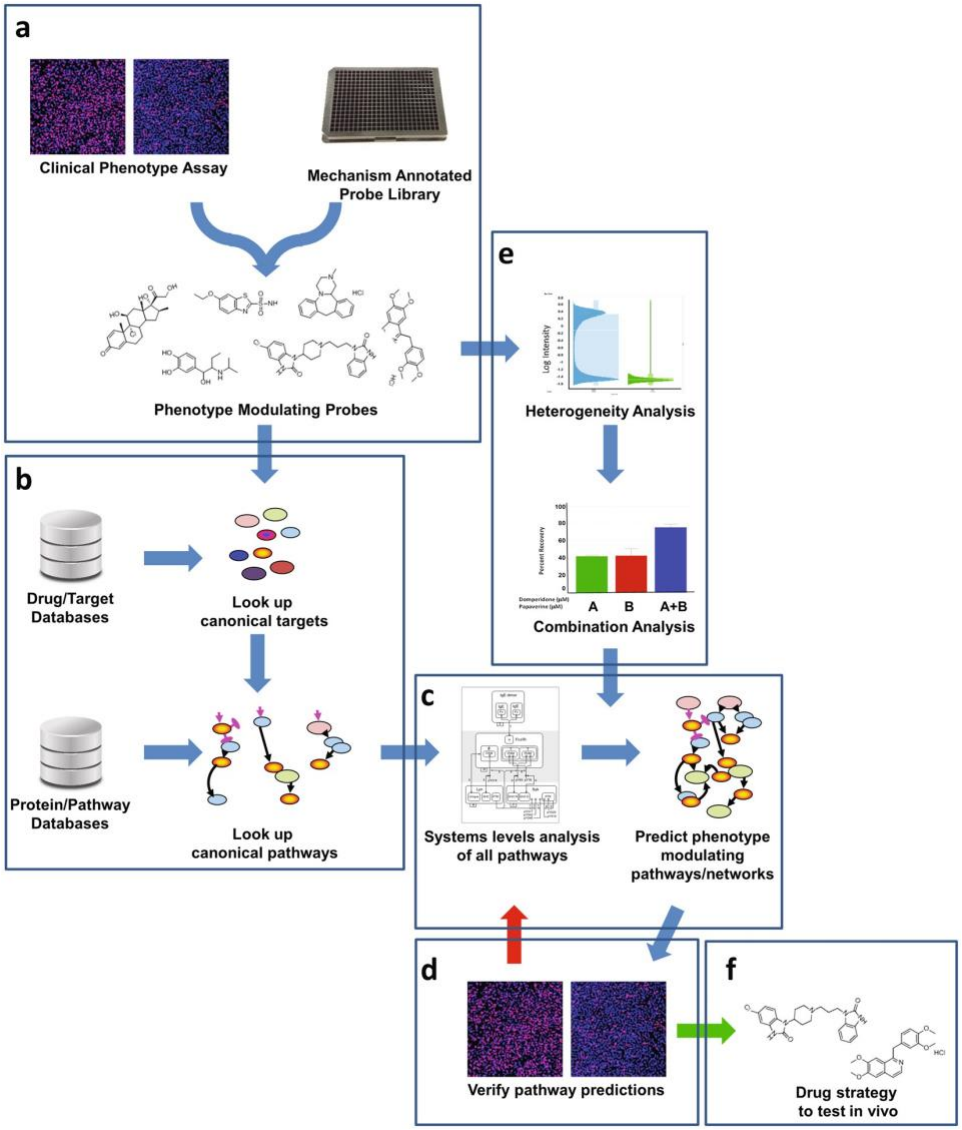


Figure 3.1 Chemogenomics component of the QSP platform

(a) Libraries of mechanism annotated probe compounds are screened in a clinically relevant phenotypic assay to identify phenotype modulating probes. (b) Targets for the active probes are identified from various drug-target databases and then are associated with biological pathways using information from protein-pathway databases. (c) Using a systems level analysis of all pathways identified, computational analysis is performed to predict the optimal modulating pathways/networks based on the activity of the respective probes (i.e., activation or inhibition of pathways in relation to the known effects of the pathway on the phenotype). (d) Predicted pathway/network hypotheses are tested in phenotypic assays by i) testing additional compounds known to modulate the pathways, ii) testing compounds predicted by advanced ML methods that will modulate the pathway, iii) modulate pathways by knock-down and

knock-in approaches, and/or iv) evaluate probes in pathway specific phenotypic assays. If pathways are not confirmed, then the hypothesis is refined with the new information gained from the testing, additional probes are identified, and the new hypothesis is tested. If pathways are confirmed, then the active probes are advanced to in vivo testing. (e) At the initial screening analysis stage, the heterogeneity of phenotype modulating response is assessed. If no heterogeneity is detected, then proceed as above. However, if heterogeneity is detected, then hypotheses are developed and tested to characterize the basis of the heterogeneity (e.g., effects of combinations of different compounds). The information gained from the heterogeneity analysis is used to inform the prediction of the phenotype modulating pathways/networks. (f) The outputs of this strategy are i) a systems level understanding of the pathways/networks involved in the clinically relevant phenotype which enables the design of optimal therapeutic strategies, and ii) probes/drugs that can be advanced to in vivo and clinical testing.

We initiated the QSP approach and implemented the chemogenomic strategy investigating the protective effects of small molecule probes with diverse canonical molecular mechanisms of action in a well-established striatal neuronal cell model (*STHdh*^{Q111}) for HD (Trettel, et al., 2000). The objective of this work is to generate testable hypotheses regarding disease mechanism and potential mechanisms involved in protection of neuronal cells from mHTT dependent toxicity. We report here on the first two iterations of the QSP approach. We identified a number of small molecule probes with a range of distinct canonical mechanisms that protect the *STHdh*^{Q111}cells from mHTT-induced death. We found that the response of the cell population to most of the compounds was heterogeneous, i.e., not all of the cells within a population were protected by the compounds, which was not unexpected since heterogeneous responses to compounds are common (Gough, et al., 2014). Interestingly, testing of combinations of moderately active compounds identified specific combinations that synergistically increased the efficacy of protection. Analysis of the canonical mechanisms of 10 compound pairs that synergistically protected *STHdh*^{Q111} cells showed a convergence of pathways leading to the activation of PKA and PKG. Cytoplasmic

phospho-PKA levels were lower in *STHdh*^{Q111} than in the wild type *STHdh*^{Q7} cells under stress conditions, and these levels were increased by several of the protective compounds. In addition, co-incubation with the PKA inhibitor H89 inhibited the protective effects of the compounds. Our results suggest that active PKA may have a role in the protective effects of these compounds. The information gained from the annotated compounds and combination analysis provided input for inference of neuronal cell protective pathways.

3.1.2 Results

3.1.2.1 Characterization of neuronal cell protective compounds in the *STHdh*^{Q111} model

We employed the well-established *STHdh*^{Q111} cell model for HD (Lu, et al., 2013; Trettel, et al., 2000) to identify compounds that would protect neuronal cells from mHTT-dependent cell death. In this model, serum deprivation (which mimics the clinical stress of growth factor deprivation) of the *STHdh*^{Q111} cells containing mHTT results in cell death, whereas under the same conditions the *STHdh*^{Q7} wild type cells are resistant to cell death. The propidium iodide (PI) readout enables an unbiased assessment of cell death by measuring an irreversible step that is common to all cytotoxic mechanisms (Kroemer, et al., 2009). Under serum-depleted conditions, ~50 percent of the *STHdh*^{Q111} cells underwent cell death as evident by positive nuclear PI staining, compared to less than 10 percent of the wild type *STHdh*^{Q7} cells (**Figure** 3.2). From screens of the LOPAC1280 library, the NCATS Pharmaceutical Collection17, and a library of 83 compounds computationally predicted to be neuroprotective (see Methods), we confirmed the activity of 32 compounds (**Figure** 3.3).

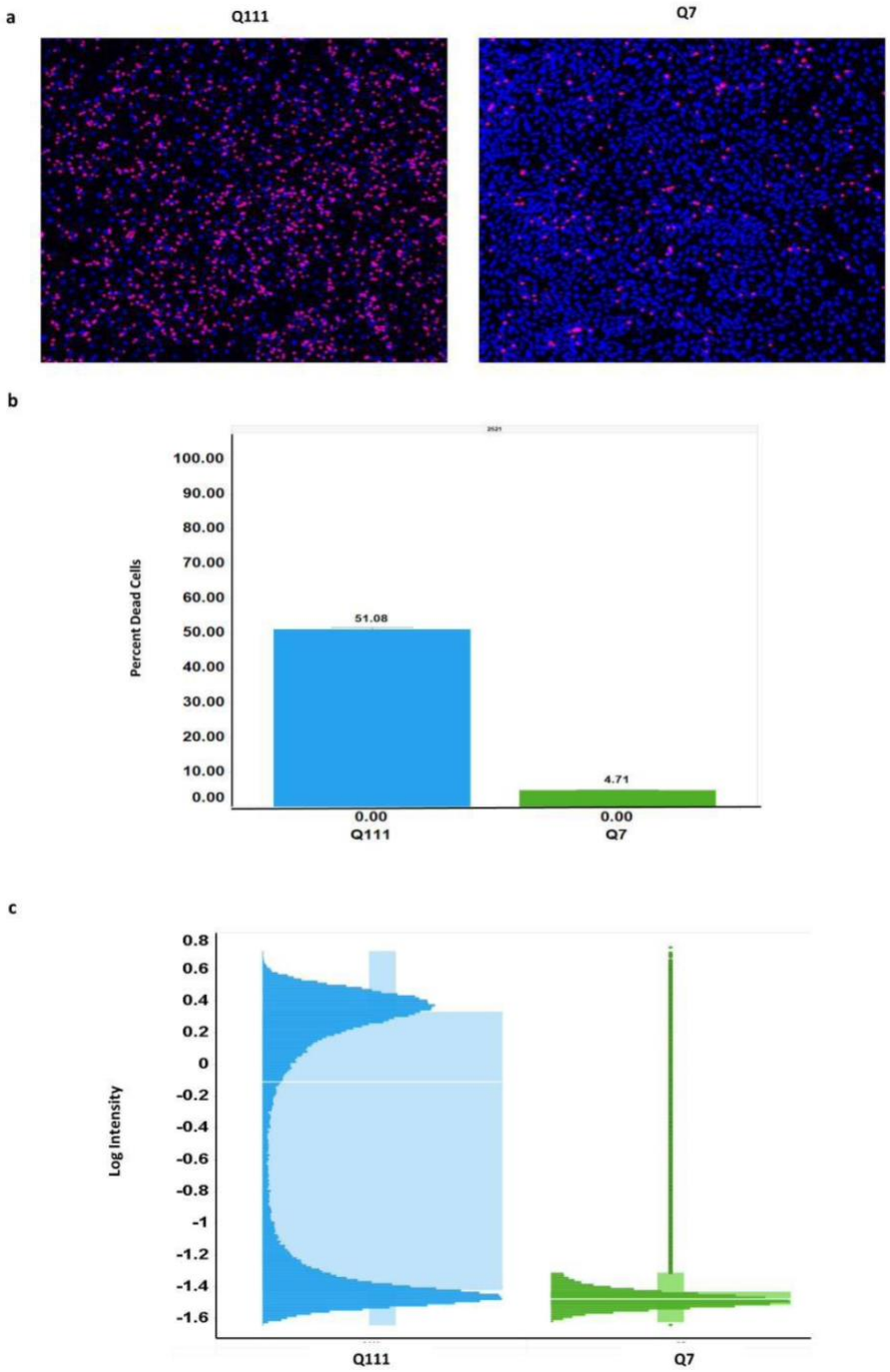


Figure 3.2 Heterogeneity in mutant huntingtin (mHTT) induced neurotoxicity in *STHdh^{Q111}* cells

a) *STHdh^{Q111}* and *STHdh^{Q7}* cells were incubated under serum free conditions for 24h at 37°C, labeled with Hoechest (blue) and PI (red) and imaged. B) Under stress conditions only ~ 50% of the *STHdh^{Q111}* cells die as evident by only half of the Hoechst positive cells labeling with PI. Only about 5-7% of the *STHdh^{Q7}* cells die under these conditions. C) Histograms quantifying the intensity of PI in the nucleus of *STHdh^{Q111}* cells (*left*) and *STHdh^{Q7}* cells (*right*).

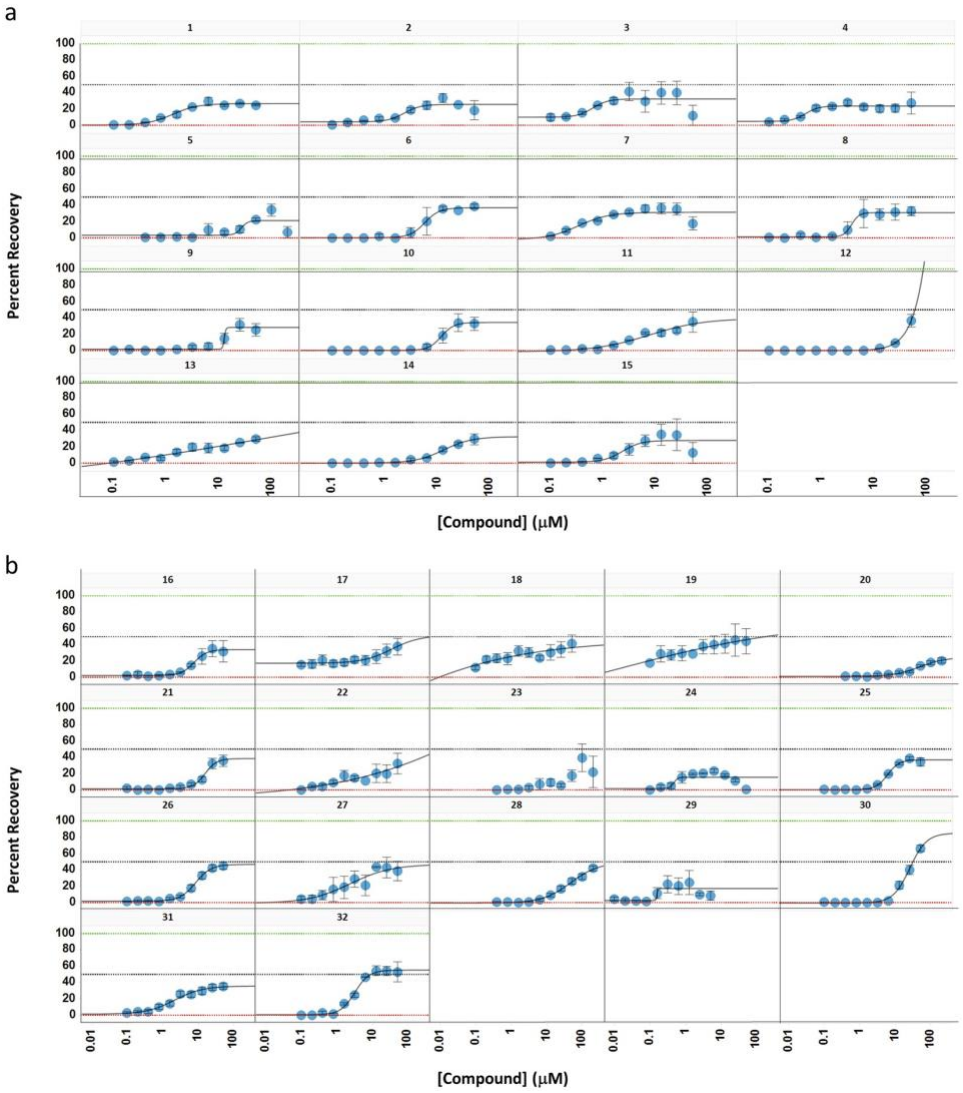


Figure 3.3 Compounds with confirmed neuroprotective activity in the *STHdh^{Q111}* model

Compound titrations were tested for protective activity in the 384-well PI assay. Compounds representing a diverse set of canonical mechanisms show only partial efficacy in protecting *STHdh^{Q111}* cells from mHTT induced cell death. (a) Compounds reported in the literature to be associated with central nervous system (CNS) activity: 1) 3-tropanyl-indole-3-carboxylate hydrochloride; 2) Benztropine mesylate; 3) Cyproheptadine hydrochloride; 4) Domperidone; 5) Isoetarine mesylate; 6) JWH-015; 7) Loxapine succinate; 8) Meclizine; 9) Mianserin hydrochloride; 10) PD 168,077 maleate; 11) Quipazine, N-methyl-,dimaleate; 12) Ruthenium red; 13) SB 203186; 14) Triprolidine hydrochloride; 15) Vinpocetine. (b) Compounds reported to be associated with non-CNS activity: 16) (Z)-Gugglesterone; 17) Beclomethasone; 18) Betamethasone; 19) Budesonide; 20) Ethoxzolamide; 21) Flutamide; 22) Hydrocortisone; 23) Lansoprazole; 24) Lonidamine; 25) m-Iodobenzylguanidine hemisulfate; 26) Papaverine hydrochloride; 27)

Prednisolone; 28) Sodium Nitroprusside; 29) Vorinostat; 30) Tetradecylthioacetic acid; 31) Triamcinolone; 32) U-83836 dihydrochloride. Results are from triplicate samples run in at least two independent experiments (Error bars are +/−SE).

Interestingly, the level of protection afforded by the majority of the compounds did not reach 100%, exhibiting plateaus in the dose-response curves between 30% and 50%. We verified that the neuronal cell protection observed was not an overestimate simply due to an undetectable loss of dead cells (**Figure** 3.4), and that partial protection was not simply due to limited solubility within the efficacious dose range (**Appendix Appendix A.1**). The spectral properties of PI are red shifted relative to the majority of small molecule compounds, thus avoiding compound interference (quenching). Preliminary analysis of the hit compounds in an LDH-based cell death assay with a format and readout distinct from that of PI showed similar curves for the hit compounds (data not shown) as seen in the PI assay. For a subset of compounds, we also examined the direct effect on quenching the PI signal and found that quenching did not occur (**Figure** 3.5). These results indicate that the partial protection was an outcome of compound perturbation of mHTT-induced biology under these experimental conditions.

We searched the DrugBank and STITCH DBs for the canonical targets of the 32 active compounds. Ten compounds had no known targets in either DB; the remaining set of 22 displayed a diverse range of canonical mechanisms of action targeting 75 proteins on a number of pathways (**Appendix Appendix A.2** and **Appendix A.3**). Many of the canonical targets have known functions that are critical to CNS activity. For example, histamine receptors, the target of 7 hit compounds, are associated with multiple neuropsychiatric disorders. Receptors of the neurotransmitters serotonin and dopamine are also targets of several of our hit compounds. Nine active compounds did not share any targets with other hits in the screen, suggesting that either

multiple mechanisms are capable of conferring neuronal cell protection or some of the active compounds operate through shared non-canonical mechanisms.

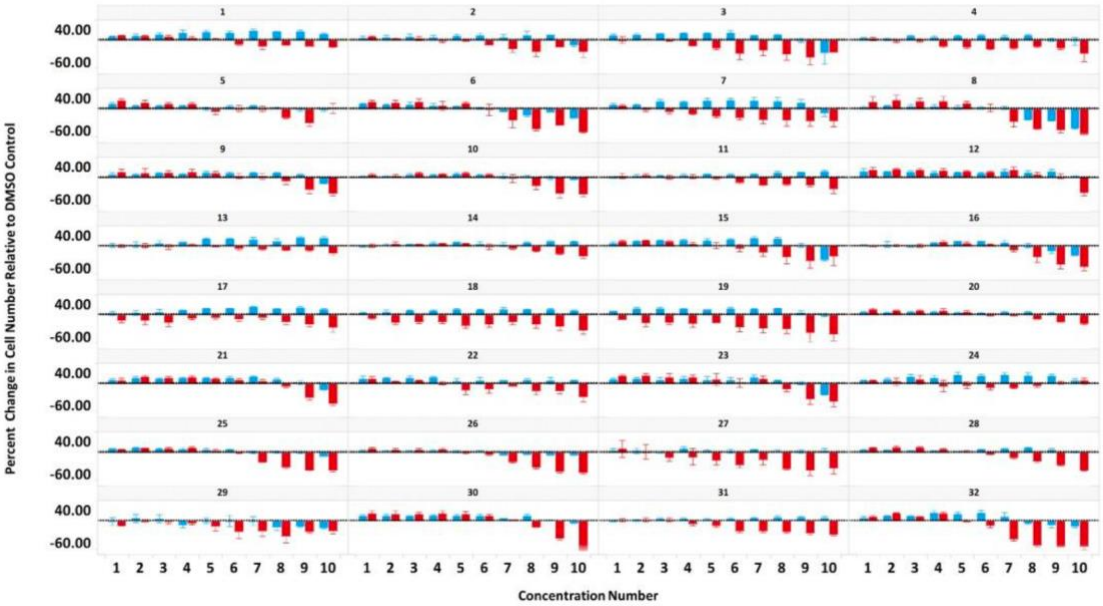


Figure 3.4 Change in total and dead cell numbers with compound treatment relative to DMSO

A greater percent change in the number of dead cells (PI positive, *red* bars) was seen compared to the change in total cell number (*blue* bars) indicating that the decrease in dead cell number was not simply due to loss of cells from the plate. Panel numbers are the compounds listed in Figure 3.3.

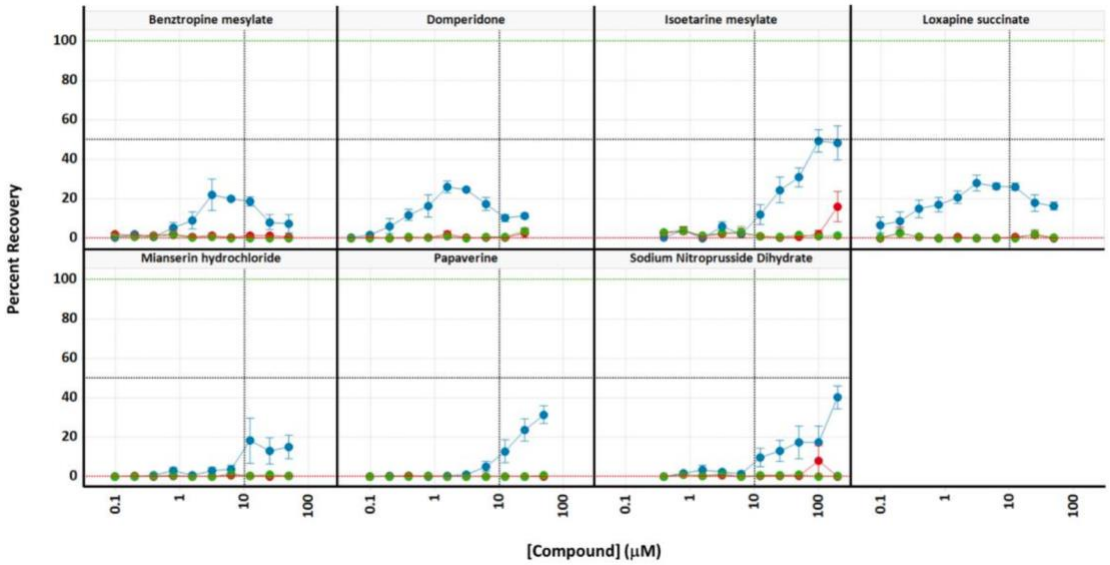


Figure 3.5 Protective compounds are not quenching the PI signal

Two sets of *STHdh*^{Q111} cells were treated with a compound titration (set 1) or DMSO (set 2) as per the standard 384-well protocol. At 24 hrs, both sets were labeled with PI and imaged. After imaging, to set 2 was added the compound titration and incubated at room temperature for 30 minutes and then imaged. The Percent Recovery was calculated for both sets. The blue curves show the compound titration curves of set 1 where the compounds were added before the PI. The red curves are the DMSO curves in set 2 before compound addition, and the green curves are the compound titration added to set 2 after PI addition. The compounds did not show the characteristic response curve when added after the PI indicating that the Percent Recovery seen with the compounds was not due to quenching of the PI signal. The analysis is from three independent runs (+/- S.E.).

3.1.2.2 Combinations show enhanced protective effects

The diversity of canonical mechanisms of the compounds exhibiting protection and the partial maximal protection for any one compound suggested the presence of more than one protective mechanism, where the sufficiency for any one mechanism to afford complete protection in an individual cell varied across the cell population. To explore this further, we asked if the efficacy of neuronal cell protection could be enhanced with pairwise combinations of compounds with different canonical mechanisms. We implemented the combination screen using 25 of the

confirmed LOPAC hits and ethoxzolamide, one of the computationally predicted hits. We screened 268 compound pairs with each compound at a single concentration that was on or near the plateau of the activity of the respective individual compound, and compared the percent recovery (i.e., protection from cell death) of compound combinations to that of the individual compounds (See **Figure** 3.6 as an example). From the 268 pairs tested, 109 pairs showed enhanced toxicity. Toxicity is defined as the loss of cells from the well using the criteria of total cell number being below 3 SD of the total number of cells in the DMSO controls. For the remaining 159 pairs of combinations (**Appendix Appendix A.4**), we determined if the combination effect was additive, synergistic, or antagonistic by calculating a combination index using the Bliss Independence Model (Bliss, 1939; Greco, et al., 1995). We found that 61 combination pairs in this screen had synergistic interactions (**Figure** 3.7**a**, **Appendix Appendix A.4**) while 90 pairs were calculated to be antagonistic and 8 appeared to be additive. We verified the synergistic assessment of the single point analysis by selecting 20 pairs of compounds, testing them in concentration response experiments, and calculating the combination index using the method of Chou and Talalay (Chou and Talalay, 1984). All of the pairs tested in this analysis were determined to be synergistic (**Figure** 3.7**b**). This test gave us confidence in the assessment of the other combinations used in the single point experiments.

Bliss independence (additivity) exists when the effects of compounds are statistically independent: applying one compound neither enhances nor diminishes the effects of the other. Whereas independence implies completely separate mechanisms, synergism and antagonism each imply a relationship between mechanisms, either within cells, across the population, or both. Antagonism at the population level can occur between compounds that share a therapeutic target and therefore compete with each other. Similarly, synergy can arise from mutually exclusive

mechanisms manifested in non-overlapping cell subpopulations. Any given cell will respond to only one compound in the synergistic pair, minimizing the number of cells that are redundantly protected by both compounds. The results of our combination screens support these mechanisms. Forty-five of the 90 antagonistic pairs of compounds identified in our screen have known targets. Fifteen of these pairs (33%) are compounds that share at least one target. In contrast, target sharing is observed in only 2 of the 41 synergistic pairs (5%) with known targets.

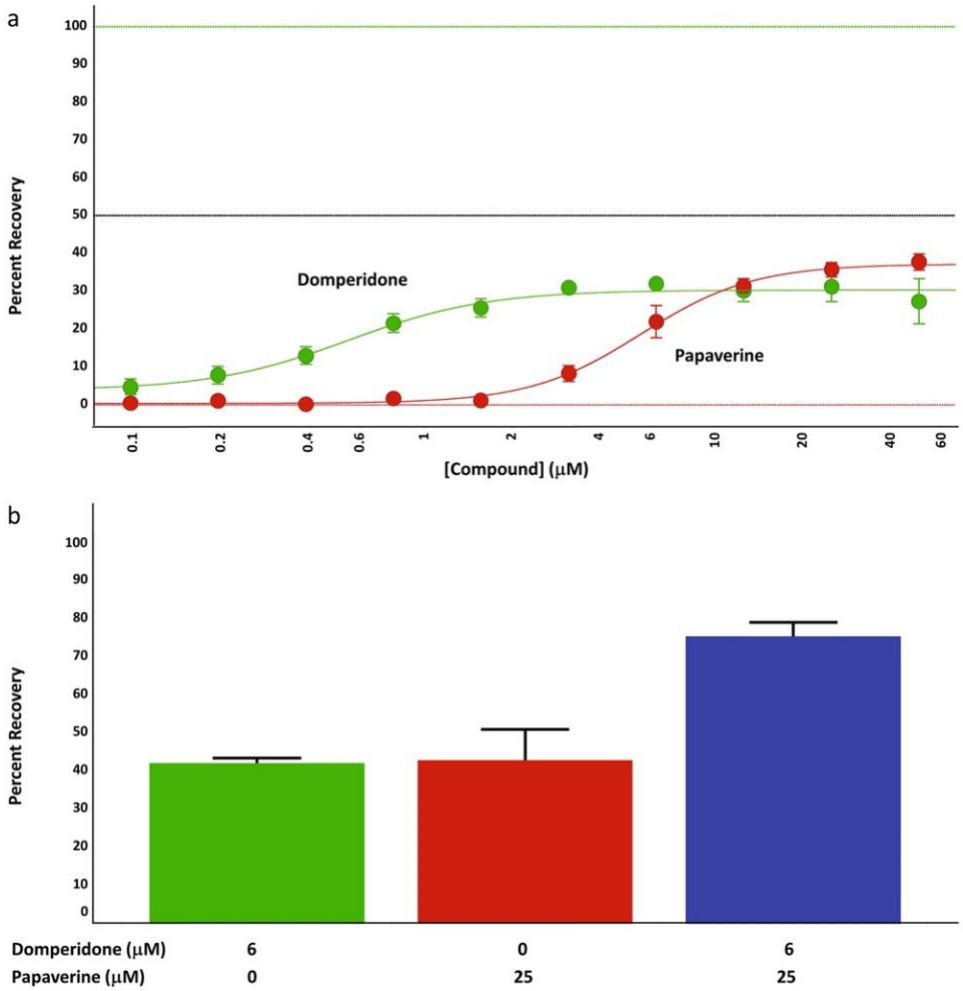


Figure 3.6 Combinations of probes with different canonical mechanisms provide enhanced protection of *STHdh^{Q111}* cells

(a) Using domperidone and papaverine as an example, concentrations of compounds that were on the plateau of the activity curve were chosen for combination experiments. In this example, 6 μM domperidone and 25 μM papaverine were selected. (b) Compounds were combined and tested in the 384-well PI assay. The percent activity of the combination was compared with the activity of the single compounds run in parallel, and the ratio of the combined activity to that of the single compound with the highest activity is taken as the combination ratio. For domperidone and papaverine the combination ratio shown here is 1.74 (n = 3 independent experiments, error bars are +/−SE). The combination experiments in panel b were run independently from the titration experiments in panel a.

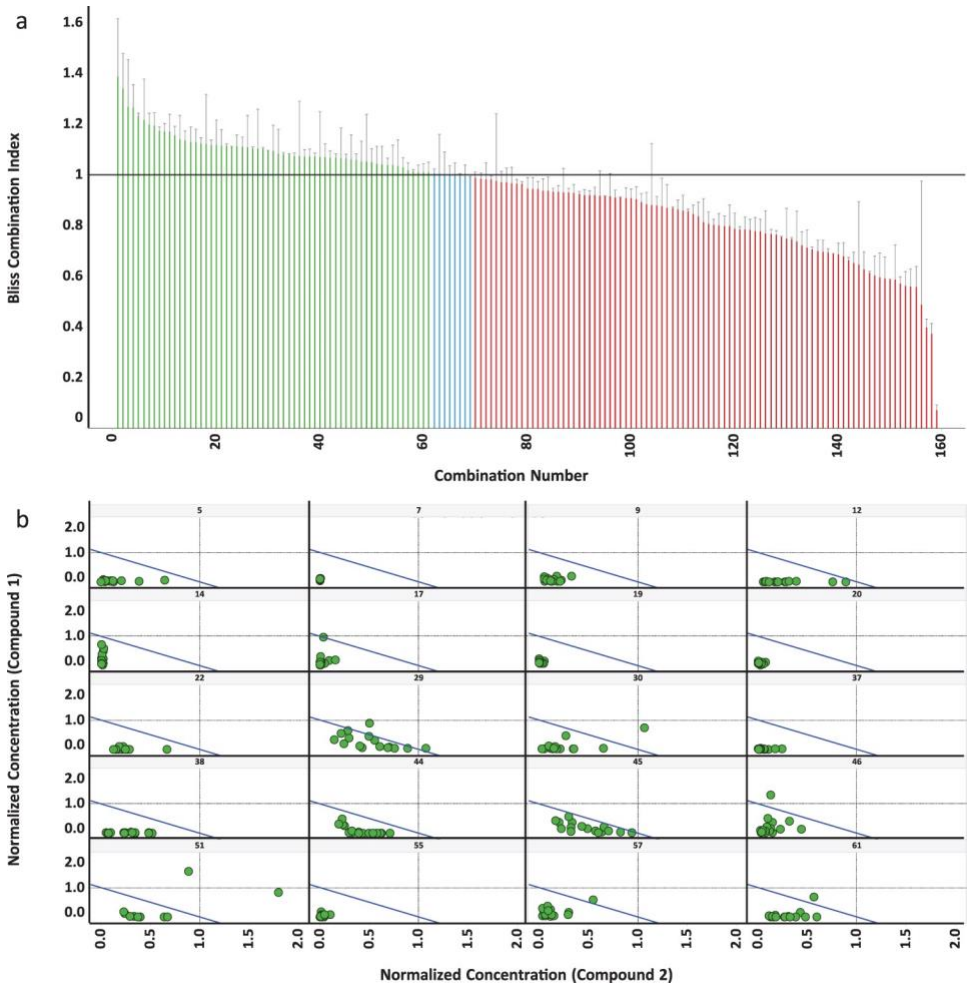


Figure 3.7 Combinations of probes show synergistic protection in STHdh Q111 cells

(a) Active LOPAC probes were screened in combinations using a single concentration of each probe. Combination numbers refer to the combinations listed in Appendix Appendix A.4. Bliss Independence Model analysis indicated 61 combinations to be synergistic in the single concentration combination screen. The Bliss Independence Model compares the predicted activity of probe combinations to the experimentally observed activity of the combination51. The Bliss Combination Index (BCI) is the ratio of the observed combination activity to the predicted combination activity based on the activity of the individual compounds. A BCI > 1 indicates synergy (green bars) and a BCI < 1 indicates antagonism (red bars), while BCI = 1 indicates additivity (blue bars). To accommodate additive BCI calculations not equaling 1 exactly, a cutoff of 0.99–1.01 was assigned to classify synergy and antagonism. (Results from at least 2 independent runs, error bars are the Median Absolute Deviation). (b) 20 probe pairs were selected and tested using 4 different concentrations, 2 each from the plateau and linear portions of the single compound concentrations curves. Curves were analyzed by the method of Chou and Talely (Chou and Talalay, 1984), and the

isobolograms are plotted. Points below the diagonal line represent synergistic activity of the two compounds (n = 2 independent runs). The panel numbers are the Combination Numbers for the combinations tested listed in Appendix Appendix A.4.

3.1.2.3 Inferring protection-relevant pathways from the compounds' canonical mechanisms

The mechanistic diversity and synergistic effects of the compounds affording protection from mHTT-induced cell death suggested functional interrelationships among their targets. Synergy can arise from mechanistic interactions within the cell if two compounds affect distinct upstream effectors of a common mechanism. Each provides partial protection to the cell, and both, when combined, may confer sufficient protection to permit survival. Alternatively, targets on the same pathway may be heterogeneously expressed in a correlated fashion within the population, causing some cells to modulate the targeted pathway in response to one compound, and other cells to modulate the same pathway through an alternative mechanism. Assays with binary readouts, such as the PI assay used here, mask the mechanistic origins of synergy. We therefore turn to pathway analysis to investigate whether the observed synergy results from pathway convergence within cells, or from mutually exclusive modulation of pathways across a heterogeneous population.

Sixteen compounds were associated with the 41 synergistic pairs that had known targets. In 21 of these synergistic pairs, the compound targets shared at least one pathway as annotated in the KEGG database. The canonical targets for compounds in 10 of these 21 pairs converged on either the cAMP/PKA signaling pathway, the cGMP/PKG signaling pathway, or both (**Appendix** Error! Reference source not found. and **Table 3.1**). At random, we would expect to find only three synergistic pairs on these pathways (enrichment factor of 3.73, see Methods). No other pathway

contained targets of more than four synergistic pairs, as was seen in both calcium signaling and gap junction pathways.

Table 3.1 Synergistic compound pairs that coverge on PKA/PKG signaling

Combination	Compound 1 MOA	Compound 2 MOA
Sodium Nitroprusside Loxapine succinate	Nitric oxide synthase	D2/3 receptor antagonist
Sodium Nitroprusside Domperidone	Nitric oxide synthase	D2/3 receptor antagonist
Sodium Nitroprusside Mianserin HCl	Nitric oxide synthase	5HT and a-2C adrenergic antagonist
Isoetarine mesylate Loxapine succinate	E-1/2 adrenergic receptor agonist	D2/3 receptor antagonist
Isoetarine mesylate Papaverine HCl	E-1/2 adrenergic receptor agonist	PDE10A inhibitor
Domperidone Papaverine HCl	D2/3 receptor antagonist	PDE10A inhibitor
Isoetarine mesylate Mianserin HCl	E-1/2 adrenergic receptor agonist	5HT and a-2C adrenergic antagonist
Sodium Nitroprusside Isoetarine mesylate	Nitric oxide synthase	E-1/2 adrenergic receptor agonist
Benztropine mesylate Isoetarine mesylate	M1 receptor antagonist	E-1/2 adrenergic receptor agonist
Domperidone Isoetarine mesylate	D2/3 receptor antagonist	E-1/2 adrenergic receptor agonist

We hypothesized that synergistic neuronal cell protection could arise in pairs of compounds that had the same effect on cAMP or cGMP signaling, but through distinct complementary mechanisms. For example, isoetarine is an agonist of the β1 adrenergic receptor (β1AR) (Isoetarine, 2016), which couples to Gs and stimulates conversion of ATP to cAMP by adenylate cyclase (AC). Benztropine is an antagonist of the M1 muscarinic receptor (Benztropine, 2016), blocking the Gi-coupled inhibition of AC activity. Thus, both compounds have the potential to increase PKA activity, but through different mechanisms: isoetarine stimulates AC, and benztropine antagonizes an AC inhibitor. Another example is the synergistic combination of domperidone and papaverine. Similar to benztropine, domperidone can elevate cAMP levels by antagonizing D2R (Barone, 1999). Papaverine inhibits the phosphodiesterases PDE4B and 10A

(Pöch and Kukovetz, 1971), reducing the hydrolysis of cAMP into AMP. The net effect of this combination is to increase cAMP levels and PKA activity through two complementary mechanisms. Thus, increasing cAMP levels and correspondingly activated PKA levels or by analogy cGMP/PKG levels may lead to cytoprotection. Multiple compounds targeting the same pathway is distinct from multiple compounds interacting with the same target. Whereas in the latter compounds may compete for the same target site and thus do not lead to enhanced modulation of the target, modulating different points on a pathway can result in synergy enabling more control in regulating the output of the pathway.

Because cAMP/PKA signaling is a key pathway involved in cell survival and has been implicated in the pathophysiology of HD (Lin, et al., 2013), we tested whether these synergistic compounds may be working through augmenting cAMP and activating PKA. We assessed the ability of benztropine, domperidone, isoetarine, loxapine, mianserin, papaverine, and sodium nitroprusside to modulate cAMP levels in the *STHdh*^{Q111} cells. cAMP levels were measured 15, 30, and 120 minutes after initial compound treatment in the presence of serum, which paralleled the pre-treatment stage of the PI assay, as it was anticipated that cAMP induction would be a relatively rapid response. All compounds, except for mianserin, showed at least a 2-fold increase in cAMP over the DMSO control at 15 minutes, which returned to control levels within 2 hours (**Figure** 3.8). Though only isoetarine showed a statistically significant increase in cAMP levels at 15 and 30 minutes, the overall profile of increased levels at 15 mins and the gradual decrease over time for all of the compounds suggested that a transient induction of cAMP did occur shortly after initial compound treatment. Sodium nitroprusside, which primarily acts through stimulating cGMP, also produced an increase in cAMP. This 2-fold increase in cAMP by the protective compounds contrasted the 250-fold increase in cAMP levels induced by forskolin. Interestingly,

forskolin did not show up as a hit in the LOPAC screen, nor did it show any protective effects when subsequently tested as a control in the PI assay run in parallel with the cAMP analysis (data not shown).

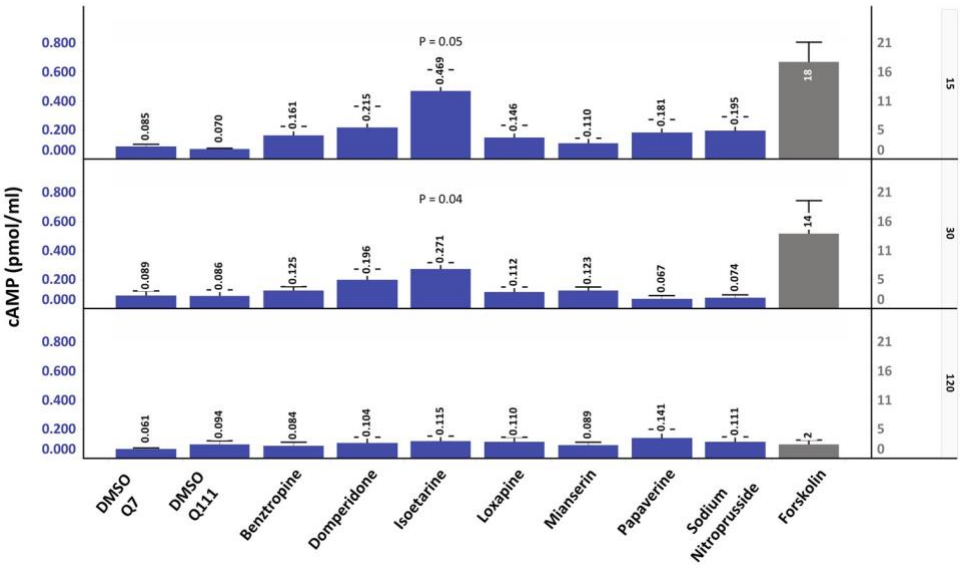


Figure 3.8 Protective compounds can induce cAMP

cAMP levels were determined in *STHdh*^{Q111} cells after incubation with benztropine (25 µM), domperidone (6 µM), isoetarine (50 µM), loxapine (6 µM), mianserin (25 µM), papaverine (25 µM), and sodium nitroprusside (66 µM) for 15, 30, and 120 minutes. Though isoetarine was the only compound to show a statistically significant change at 15 and 30 minutes, except for mianserin, the other compounds showed at least a two-fold increase in cAMP levels at 15 mins. Over time the induced levels of cAMP decreased back to the control levels. Forskolin significantly induced cAMP levels at 15 and 30 minutes with the highest levels seen at 15 minutes. The values are the average from three independent experiments (+/−S.E.) except papaverine where n = 2. All compounds except forskolin are plotted on the blue scale on the left, while forskolin is plotted on the grey scale on the right. The three panel rows are 15, 30, and 120 minutes. T-test was used to assess changes in cAMP levels relative to the *STHdh*^{Q111} cells treated with DMSO.

To determine if PKA may be involved in the protective effect of these compounds, we incubated the *STHdh*^{Q111} cells with benztropine, domperidone, isoetarine, loxapine, mianserin, papaverine, and sodium nitroprusside in the presence the PKA inhibitor H89 under the standard

PI protection assay conditions. H89 has been used extensively in the literature as a selective and potent inhibitor of PKA (Chijiwa, et al., 1990; Davies, et al., 2000). If the protection from cell death by these compounds involved activation of PKA, then the addition of an inhibitor of PKA would be expected to reverse the protective effects of the compounds. Co-incubation of 10 μM H89 with the Gi-coupled GPCR antagonists domperidone, loxapine, and mianserin resulted in 56, 52, and 35 percent reduction, respectively, in the level of protection, while the Gs-coupled agonist isoetarine resulted in a 34 percent reduction, the PDE inhibitor papaverine a 55 percent reduction, and the s-GC agonist sodium nitroprusside a 17 percent reduction compared to compound alone (**Figure** 3.9**a**). Since the primary canonical mechanism of sodium nitroprusside is activation of PKG, and given that H89 is ~10-fold selective for PKA over PKG, the absence of a marked effect with sodium nitroprusside is not unexpected. The relatively lower effect of H89 on the PKG activator sodium nitroprusside compared to the PKA activators is consistent with the canonical mechanisms of these compounds. To confirm inhibition of PKA activity by H89 under the conditions of the PI assay we measured the levels of nuclear pCREB using high content analysis (**Figure** 3.10). Consistent with the heterogeneity seen in the response of the *STHdh*^{Q111} cells to protection by the compounds, a heterogeneous distribution of pCREB levels was also detected (**Figure** 3.11). The levels of pCREB were decreased in the presence of 10 μM H89 in all cases indicating inhibition of PKA activity (**Figure** 3.9**b**). While H89 has been used extensively as a selective and potent inhibitor of PKA to understand the biology of PKA signal transduction, it has been reported that H89 has other effects as well(Murray, 2008). To address this, we also tested the effects of PKI, a reportedly more selective PKA inhibitor, on the activity of these compounds, however, PKI by itself was toxic to the *STHdh*^{Q111} cells which overshadowed any potential effect in inhibiting protection (data not shown).

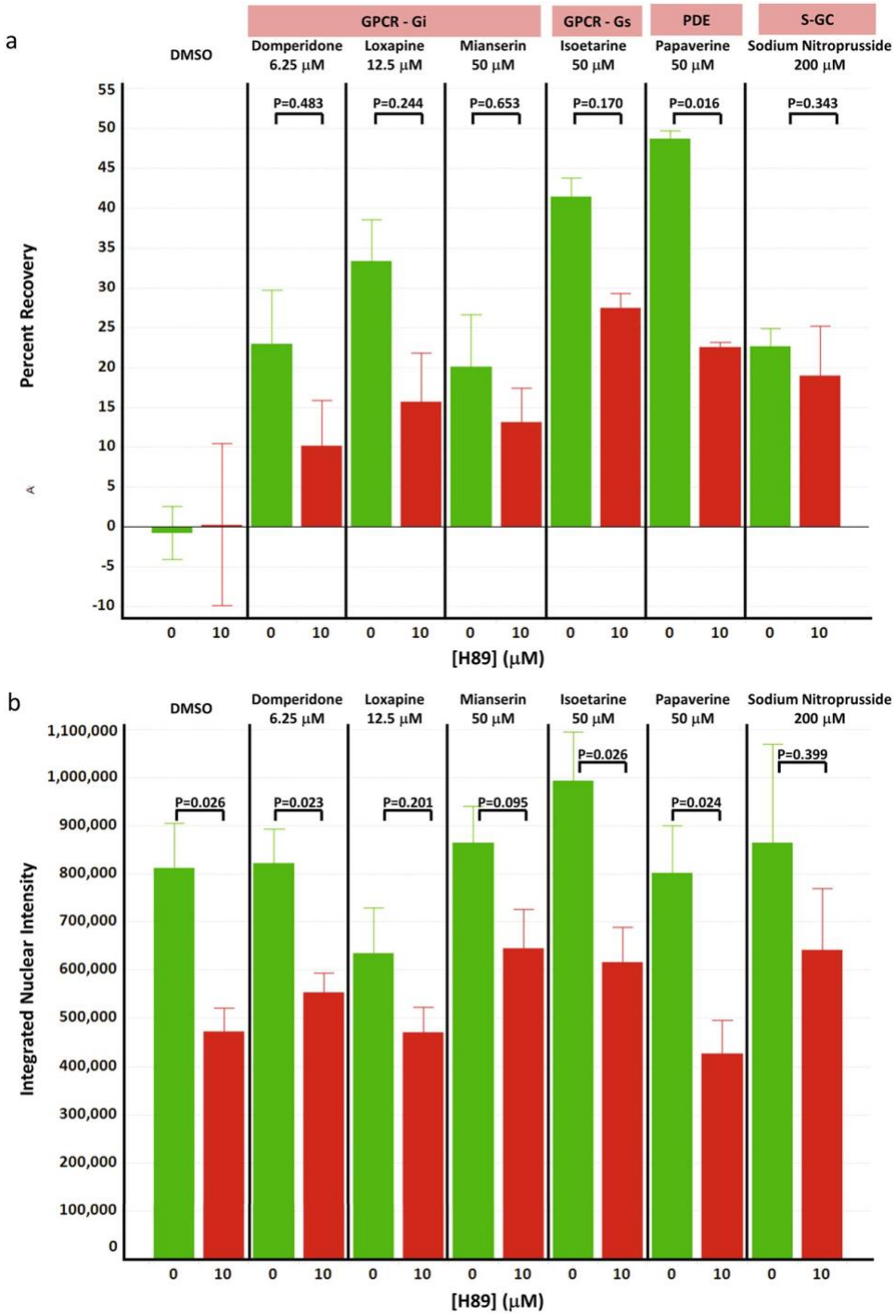


Figure 3.9 PKA inhibitor H89 inhibits the protective effects of several probes

(a) The protection of STH*dh*^{Q111} cells from mHTT induced cell death by domperidone (6 μM), isoetarine (50 μM), loxapine (12.5 μM), mianserin (50 μM), papaverine (50 μM), and sodium nitroprusside (200 μM) co-incubated with the PKA inhibitor H89 (10 μM) was assessed in the 384-well PI assay. Benztropine (50 μM) was also tested, however, combination with H89 resulted in increased toxicity over the cell death seen in the DMSO control. The concentrations

used were chosen to be on plateau of their respective activity curves (see Fig. 2). DMSO is H89 alone which showed no significant protection or toxicity. Analysis is from triplicate samples run in four independent experiments (Error bars are +/−SE). T-test was used to assess changes in the percent recovery levels relative to the *STHdh*^{Q111} cells treated with compound without H89. While only papaverine showed a statistically significant decrease, the other compounds showed a trend for H89 inhibition of the protective effects. (b) The integrated intensity of the pCREB signal was measured in the nucleus of the *STHdh*^{Q111} cells treated as above. CREB is a substrate for PKA and is used here as a surrogate marker for PKA activity to demonstrate inhibition of PKA activity by H89. Analysis is from triplicate samples run in four independent experiments (Error bars are +/−SE). T-test was used to assess changes in the pCREB intensity relative to the *STHdh*^{Q111} cells treated with compound without H89.

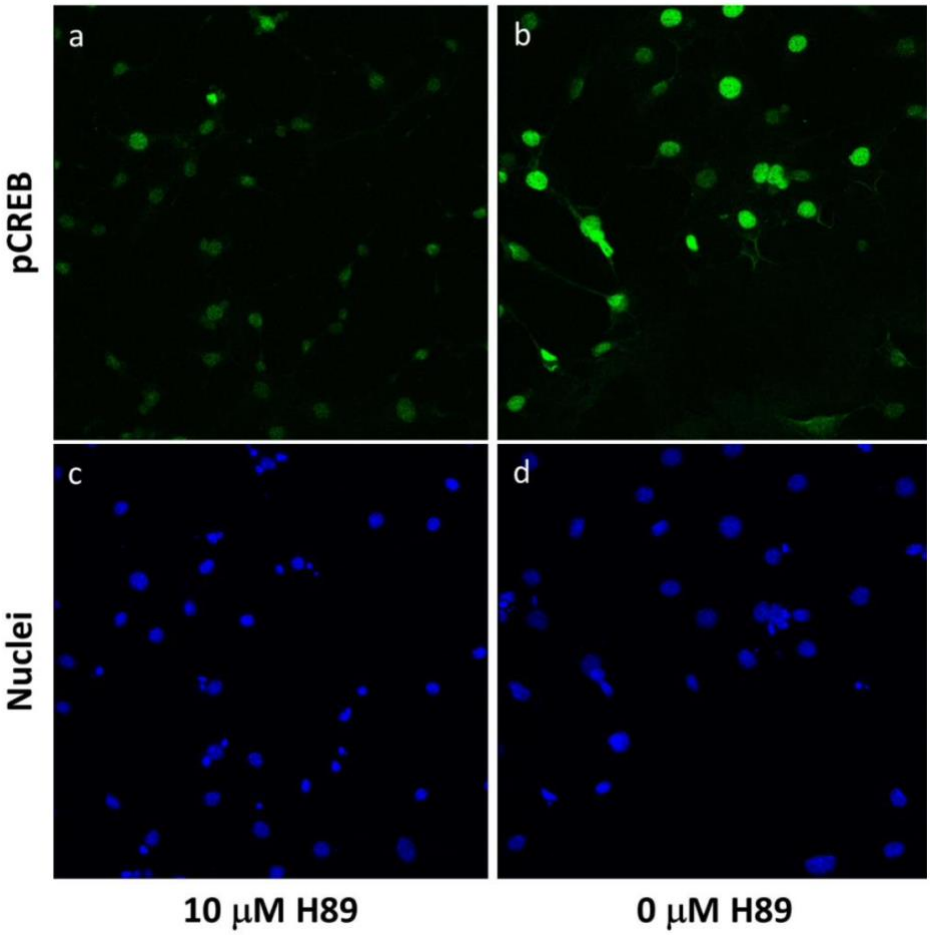


Figure 3.10 H89 inhibits phosphorylation of CREB at Ser 133

STHdh^{Q111} cells were treated either 10 μM H89 (a and c) in DMSO or DMSO alone (b and d) under the standard serum-free stress conditions at 37ºC for 24 before being fixed and labeled with anti-pCREB (Ser133). The nuclei were stained with Hoechst 33342. Images were acquired with a 40x objective. The pCREB images were scaled to 248 – 7903 gray levels and the nuclei were scaled to 562 – 14649 gray levels. The dimmer intensity of the pCREB in the presence of H89 indicates inhibition of PKA.

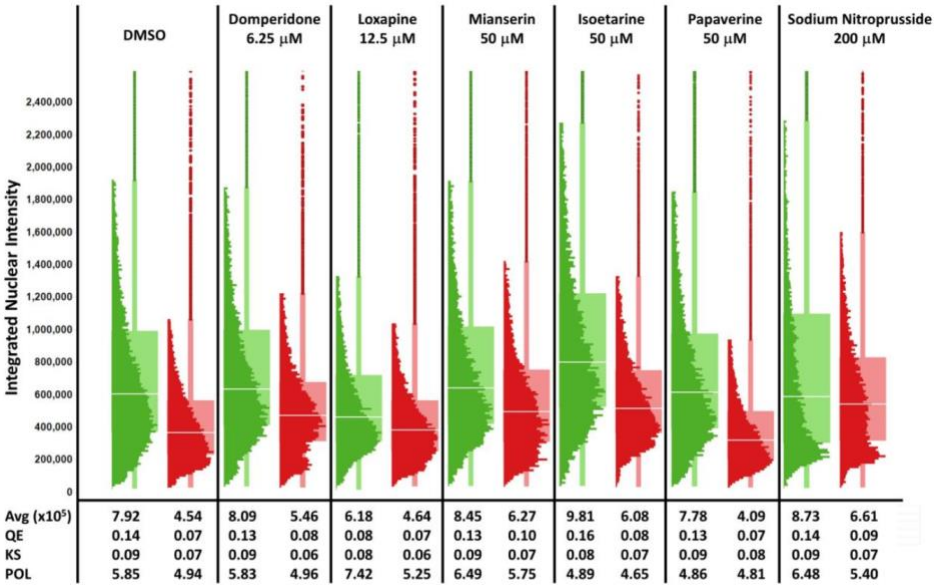


Figure 3.11 PKA activation is heterogeneous within the *STHdh*^{Q111} cell population

STHdh^{Q111} cells were treated with DMSO or protective compounds in the absence (green) or presence (red) of the PKA inhibitor H89 (10 μM) under the standard serum-free stress conditions at 37ºC for 24 before being fixed and labeled with anti-pCREB (Ser133). Images were acquired at 10x magnification and analyzed as described in Methods. HistoBox plots14 were generated showing the distribution of nuclear pCREB levels in the cell population. The plots represent the cumulative data from triplicate replicates run in each of four experiments. The average levels in the populations were calculated 14 along with the PHIs as described by Gough et al. . Quadratic entropy (QE, diversity measure) values > 0.03, Kolmogorov-Smirnov (KS, non-normality measure) values > 0.05, and percent outliers (POL) > 4.5 indicate non-normal, heterogeneous populations. These data show a high degree of heterogeneity in PKA activity as measured by its substrate CREB. Addition of H89 to the cells lowered the degree of heterogeneity in all cases, but did not completely normalize the populations.

To further assess PKA activation, we quantified the levels of PKA phosphorylated at threonine 197 (pPKA) in the catalytic subunit using high-content analysis. We examined the pPKA levels at 24 hours after serum free conditions since this was the condition where we measured the protection of the compounds. The levels of cytoplasmic pPKA were lower in the *STHdh*^{Q111}cells

relative to the *STHdh*^{Q7} (**Figure** 3.12), consistent with the hypothesis that elevated pPKA was associated with neuronal cell survival. Benztropine, isoetarine, loxapine, mianserin, and sodium nitroprusside exhibited a concentration-dependent increase in cytoplasmic pPKA approaching the levels of the wild type *STHdh*^{Q7} cells. The concentration response for domperidone, papaverine and forskolin was less pronounced. In contrast to the cytoplasm, the nuclear pPKA levels in the *STHdh*^{Q111} cells were higher than in the *STHdh*^{Q7}cells (**Figure** 3.12). None of the compounds showed a marked concentration-dependent decrease in the nuclear levels. The increase in cytosolic pPKA correlated with the percent recovery for these compounds (**Figure** 3.13); however, the concentration response curves between the compounds were distinct from each other. If pPKA were the only factor responsible for the protective effects of these compounds, then the concentration response curves for the pPKA effect on recovery would be expected to be the same. The fact that they were different suggests additional mechanisms were involved in the protection phenotype for these compounds.

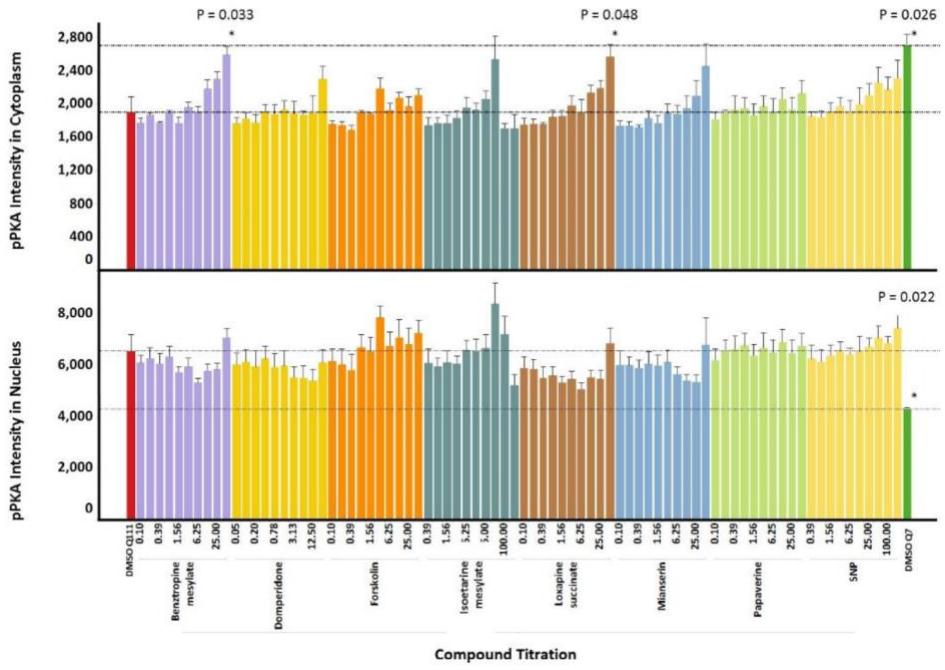


Figure 3.12 Protective compounds can activate PKA

Cytoplasmic and nuclear pPKA levels were measured in *STHdh*^{Q111} cells after incubation with benztropine, domperidone, isoetarine, loxapine, mianserin, papaverine, and Sodium Nitroprusside for 24h under serum free conditions following the protocol used for the PI protection assay. For cytoplasm levels the upper and lower dotted lines are the average level of *STHdh*^{Q7} and *STHdh*^{Q111} cells, respectively. For nuclear levels the upper and lower dotted lines are the average level of *STHdh*^{Q111} and *STHdh*^{Q7} cells, respectively. Data are the average from three independent experiments (+/- S.E.). T-test was used to assess changes in pPKA levels relative to the *STHdh*^{Q111} cells treated with DMSO. SNP = Sodium Nitroprusside.

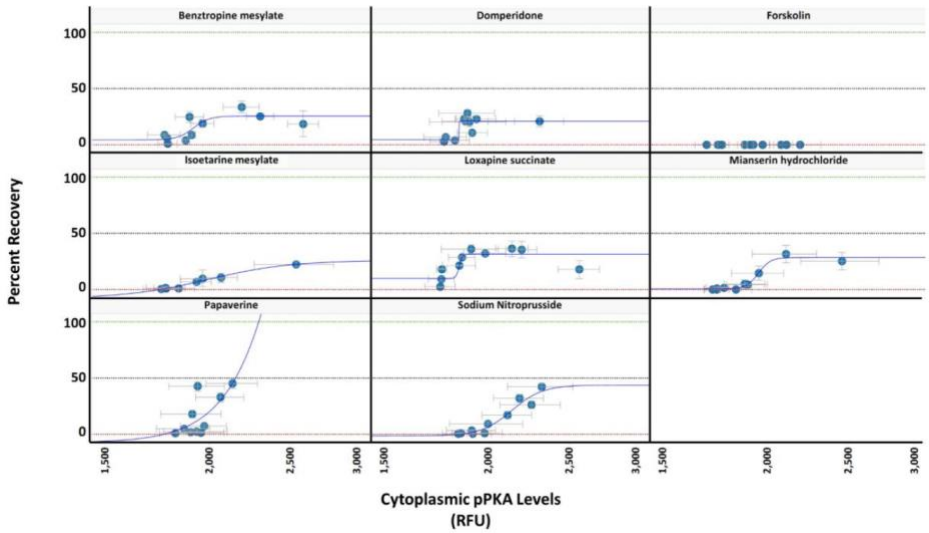


Figure 3.13 Correlation between Percent Recovery from mHTT toxicity and pPKA levels

The Percent Recovery assessed in the PI assay is plotted against the relative levels of pPKA induced by the compounds measured in the High content assay. All compounds increased pPKA though some were more effective and showed a robust concentration response (see Figure 3.9). Different response curves were observed among the protective compounds. Forskolin was not protective, but did show pPKA levels in the range where protection was seen for the other compounds. The Percent Recovery analysis is from triplicate samples run in least two independent runs, and the pPKA analysis is from triplicate samples run in three independent runs (+/- S.E.).

3.1.2.4 Some compounds may be protecting by non-canonical mechanisms

Our pathway analysis was based on using canonical mechanisms of action for the identified compounds; however, we hypothesized that the protective activity of some of the compounds might be through alternative mechanisms, as well. Several structurally distinct carbonic anhydrase inhibitors were present in the library of compounds, but only one of them, ethoxzolamide, showed protective activity in the PI assay (**Figure** 3.14). To determine if ethoxzolamide was acting through its canonical carbonic anhydrase inhibition mechanism, we synthesized its methyl sulfonyl analog in which the amine group that is critical for the carbonic anhydrase inhibition by this drug class (Supuran, et al., 2003) was replaced by an isosteric methyl group. We demonstrated that the methyl

sulfonyl analog of ethoxzolamide was approximately 7-times more potent than ethoxzolamide itself and equally efficacious (**Figure** 3.15). Though the methyl sulfonyl analog for inhibition of carbonic anhydrase was not tested directly, the activity of the methyl sulfonyl analog suggests that the protective activity observed with ethoxzolamide may be due to a distinct mechanism and not due to its canonical carbonic anhydrase inhibition.

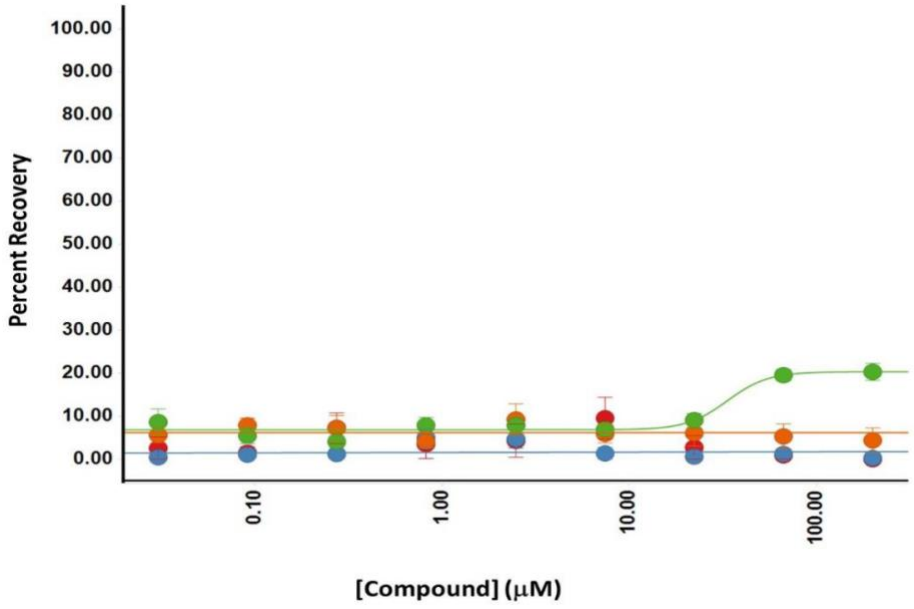


Figure 3.14 Comparison of carbonic anhydrase inhibitors in protecting *STHdh^{Q111}* cells

Inhibitors include ethoxzolamide (*green*), acetazolamide (*red*), dorzolamide (*orange*), and brinzolamide (*blue*), only ethoxzolamide demonstrated increased protection.

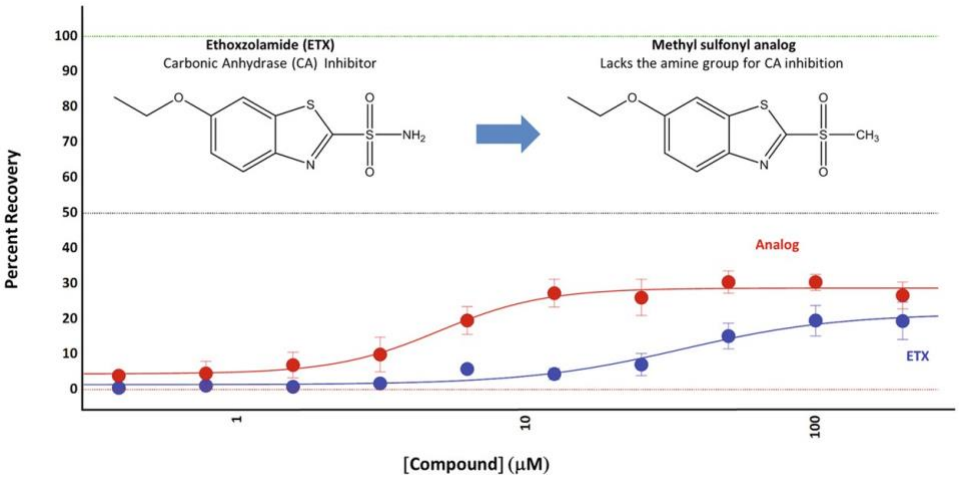


Figure 3.15 Ethoxzolamide may not work through the canonical carbonic anhydrase mechanism

The methyl sulfonyl analog of ETX does not contain the sulfonamide group of ETX and it is not expected to inhibit carbonic anhydrase (Supuran, et al., 2003), though we did not test this directly. This analog is 7-fold more potent than ETX in protecting *STHdh*^{Q111} cells from stress induced cell death in the propidium iodide assay suggesting that the mechanism of protection of ETX is not through carbonic anhydrase inhibition. Acetazolamide, brinzolamide and dorzolamide, all reported carbonic anhydride inhibitors, did not protect *STHdh*^{Q111} cells (see Figure 3.14) further supporting the idea that inhibition of carbonic anhydrase is not a protective mechanism. Interestingly, the methyl sulfonyl analog only protected ~50% of the *STHdh*^{Q111} cells consistent with the existence of distinct protection mechanisms in different subpopulations of cells.

3.1.3 Discussion

Despite major technological advances in genome editing, differentiation of patient-derived iPSCs, and recapitulation of complex disease phenotypes in human microphysiological models (i.e., organs-on-a-chip), our knowledge of disease mechanism is often the limiting factor for optimizing therapeutic strategies for patient cohorts. QSP has emerged as an approach to address this void (Perez-Nueno, 2015; Stern, et al., 2016). Commensurate with advances in the development of clinically relevant models, and complementary to systematic genetic approaches

(Martz, et al., 2014), we anticipate an increased use of mechanistically diverse and well annotated chemical libraries, especially those containing FDA approved drugs, to probe disease mechanism. This small molecule approach has the potential to lead directly to drug repurposing and optimal drug combination strategies that maximize efficacy and minimize toxicity, as well as to serve as a starting point for selecting targeted libraries for additional discovery efforts. Thus, we expect that this approach will play an increasingly important role in mechanistic studies and drug development efforts to address many of the 7,000 rare diseases that exist worldwide. In the case of HD, screening identified several drugs having well-defined canonical modes of action that partially protected against mutant HTT-induced neuronal cell death. The fact that only the mutant cell line shows cell death under the stress conditions demonstrates that this phenotype is disease dependent, and the fact that the compounds are protective in the mutant cell line indicates that they are active in reversing the disease dependent phenotype. Many combinations exhibited significant synergy, suggesting a functional network association among them involving PKA (PKG) signaling.

The analysis reported here suggested that cAMP/PKA signaling was involved in the protection of neuronal cells from mHTT-induced toxicity in the STHdh^{Q111} model. Several lines of evidence from the literature suggest that altered activity of the PKA (PKG) signaling is directly pathogenic and does not simply represent a beneficial compensatory mechanism for averting mHTT-induced cell death. Single cell analysis employing an optical pulse-chase method (Tsvetkov, et al., 2013) has demonstrated that neuron-to-neuron variation in protein homeostasis capacity (i.e., proteasome activity) contributes substantially to a given cell's susceptibility to the effects of misfolded proteins (Tsvetkov, et al., 2013). Specifically pertinent to HD, striatal neurons were, on average, more vulnerable to disease-causing misfolded mHTT and cleared a corresponding mHTT reporter more slowly than cortical and cerebellar neurons. Statistical

modeling linked intrinsic protein homeostasis capacity in striatal, cortical, and cerebellar neurons to their vulnerability to mHTT-induced degeneration. Furthermore, animal models of HD show that mHTT stress-induced impairment of the proteasomal capacity in the striatum is associated with lowered PKA activity (Lin, et al., 2013). This reduced PKA activity is caused by the accumulation of negative regulatory PKA subunits that are normally controlled by proteasomal degradation. Since it has also been shown that full proteasomal activity depends upon PKA phosphorylation, a feed-forward loop of diminished PKA and proteasomal activity has been suggested as an important component of HD pathogenesis. Consistent with the results presented here, pharmacologic intervention corroborated this hypothesis, as agents that increase cAMP and activate PKA restored proteasomal activity and ameliorated motor impairment (Tsvetkov, et al., 2013). By analogy, very recent results indicate a similar feed-forward loop operative in other tauopathies (Myeku, et al., 2016). Our results showing the inhibition of the protective activity of the compounds by a PKA inhibitor, a lower level of cytosolic pPKA in the mHTT cells relative to the wt cells under stress conditions, and the association of increasing pPKA with increasing recovery from cell death are consistent with the observations in the literature. The lack of a marked increase in pPKA by domperidone or papaverine does not necessarily contradict the observation that the PKA inhibitor H89 prevented protection by these compounds. The spatiotemporal activation and regulation of cAMP and PKA is complex (Allen and Zhang, 2006; Baillie, et al., 2005; DiPilato, et al., 2004; Rinaldi, et al., 2015; Sample, et al., 2012) and the 24 hour time point may not have been optimal to capture activation by all of the mechanisms. However, the fact that the PKA recovery curves were different among the compounds suggests that factors in addition to activation of PKA per se may also contribute to neuronal cell protection.

Forskolin also increased pPKA to levels that were associated with protection by the other compounds yet itself was not protective, further suggesting that additional factors are important for protection. Since forskolin was unable to induce protection from cell death in *STHdh*^{Q111} cells, it appears that regulatory nuances beyond simply a global and robust stimulation of cAMP downstream of specific GPCR machinery are necessary to elicit a protective response. While the inability of forskolin to protect could result from its well-known off-target effects (e.g., glucose transporter (Morris, et al., 1991)), strong nonselective stimulation of cAMP could result in antagonistic combinatorial effects consistent with our results showing that the majority of combinations of partially protective compounds were indeed antagonistic (or toxic). On the other hand, an intrinsic characteristic of cAMP/PKA signaling is compartmentalization, and subcellular localized generation of cAMP is tightly coupled to activation of PKA(Allen and Zhang, 2006; DiPilato, et al., 2004; Rinaldi, et al., 2015; Sample, et al., 2012). Therefore, it is tempting to speculate that crosstalk between two cAMP/PKA compartments could provide the basis for the observed synergy between two compounds acting along the cAMP/PKA signaling axis and result in the necessary spatial and temporal modulation of cAMP/PKA signaling to elicit a protective response. We expect that extension of the imaging analysis initiated in this study in conjunction with additional cAMP/PKA signaling biosensors will enable the role of signaling compartmentalization in the protection from the pleiotropic effects of mutant HTT to be determined and perhaps offer insights into the mechanistic underpinnings of the pathogenic dysregulation.

We found that the canonical targets of a number of compounds converge on a plausible mechanism for neuroprotection from mHTT toxicity, and that the literature supports the role of this mechanism in HD. However, this mechanism alone neither explains all of our results nor

provides a clear path to an HD therapeutic. Given the pleiotropic nature of mHTT, and evidenced by our synergistic results that do not involve cAMP/PKA signaling, we anticipate that other protective mechanisms exist. Further, although the present work focuses on within-pathway convergence as the mechanism of synergy, it is also possible that the synergistic effects that we see result from mechanistic heterogeneity within the cellular population. Addressing this possibility could provide insight into the basis for distinct vulnerabilities among subpopulations of mHTT-expressing cells and the relationships among the different pathways regulating their susceptibility to stress-induced cell death. In the next iteration of the QSP cycle, we are broadening the analysis to obtain a more complete picture of pathways and networks involved in the protection of the Q111 cells based on the canonical mechanisms of active probes. In addition, the canonical mechanisms may not be the only mechanisms through which compounds protect from mHTT toxicity, as exemplified by the activity of the ethoxzolamide analog. Although mHTT is pleiotropic, small molecule compounds can also interact with multiple targets; it has been estimated that most drugs bind to on average 6 targets (Kell, et al., 2013). Indeed, modulation of non-canonical targets in addition to activation of the PKA pathway by the seven probes could help explain why we see only partial inhibition of recovery by H89. Exploring non-canonical mechanisms has the potential to lead to the identification of novel pathways for neuronal cell protection and emphasizes the value of assembling chemical libraries containing structurally distinct probes that have the same canonical mechanism. Thus, in subsequent iterations of the QSP analysis we are applying various approaches (Schenone, et al., 2013) including chemical proteomics (Rix and Superti-Furga, 2009; Wright and Sieber, 2016) to identify the targets to which the protective compounds are binding, and computationally expanding the potential targets and pathways to predict non-canonical interactions of the protective compounds. We are also actively

expanding the scope of potential mechanisms by analyzing additional synergistic neuronal cell protective pairs and screening larger mechanistically annotated libraries (e.g. NCATS Pharmacologically Active Chemical Toolbox library). Key to this whole approach is the systems level analysis that ensures a mechanistically unbiased assessment of the biology, which will enable more efficient and novel approaches to therapeutic design in the long run.

The work presented here represents the first two iterations of the QSP platform approach developed at the University of Pittsburgh (Stern, et al., 2016) starting with mechanism-annotated probe compounds and a clinically relevant phenotypic assay, and leading to the identification of disease-relevant pathways. We show that an integrated chemogenomic strategy using information about probes that modulate a clinical phenotype can lead to testable hypotheses and provide insights to targetable biological mechanisms for disease treatment. To our knowledge, this is the first report of such an approach applied to HD, which can be broadly applied to any discovery strategy involving small molecule modulation of a disease phenotype.

3.1.4 Materials and Methods

3.1.4.1 Computational predictions of drug-target binding

We identified 83 compounds as potentially neuroprotective using a latent factor model (LFM) combined with structural similarity. Our LFM approach, Balestra (Cobanoglu, et al., 2013; Cobanoglu, et al., 2015), is based on probabilistic factorization of the incomplete drug-target interaction matrix. Given a binary matrix, R, of interactions between N drugs and M targets, Balestra decomposes it into the product of two matrices, U and V, that express the drugs and targets in terms of D latent variables:

$$\boldsymbol{R}_{N\times M} = \boldsymbol{U}^{T}_{N\times D}\boldsymbol{V}_{D\times M} \qquad (3.1)$$

This decomposition assigns values – loosely comparable to interaction probabilities – to the previously undetermined elements of $\boldsymbol{R}$. Our LFM was trained on chemical-target interaction data from DrugBank (version 4.0.0, approved drug subset) and STITCH (version 3.0, experimental data only) databases. We identified from the same DBs all canonical targets of 15 hit compounds from an earlier mitochondrial screen (Wang, et al., 2008) and 9 compounds that are in clinical trials for neuroprotection in HD. Compounds that the LFM predicted to have interaction values greater than 0.9 were selected as potentially neuroprotective. In addition to the LFM, the ROCS module in OpenEye software (Hawkins, et al., 2007) was used to predict neuroprotective compounds based on 3D structural similarity. A separate query was built based on the 3D shape and heavy atom properties of each of the 15 compounds from the mitochondrial screen. Each query was used to search compounds in DrugBank, and the top ranked compounds were selected based on the OpenEye ComboScore measure of shape and atom properties. The final set of predicted neuroprotective compounds was generated by merging the results from LFM prediction and 3D structural similarity search.

3.1.4.2 Pathway analysis

All canonical targets for the probes that showed cell protection were identified in DrugBank (version 4.5.0, approved drug subset) and STITCH ligand-protein interaction database (version 4.0, human subset with an experimental confidence score greater than 0.7), as well as data mining from the literature. The 32 probes were mapped to 75 targets and detailed drug-target interaction mapping was shown in detail in **Appendix Appendix A.2**. Each target, and each probe by association, was then mapped to one or more pathways in the KEGG pathway database

(http://www.kegg.jp, version 07, 2016, homo sapiens), ending up with 34 pathways as shown in **Appendix Appendix A.3**. We identified for further analysis all synergistic pairs of compounds in which the two compounds had different targets on the same pathway.

Over-representation of pathways among synergistic pairs in our screen is quantified using the enrichment factor

$$EF_i = \frac{\frac{N_{pairs_i}}{N_{pairs}}}{\left(\frac{N_{compounds_i}}{N_{compounds}}\right)^2} \qquad (3.2)$$

where N_{pairsi} is the number of synergistic pairs mapped into pathway i, N_{pairs}=61 is the total number of synergistic pairs identified in our combination screen, $Ncompoundsi$ is the number of compounds from DrugBank and STITCH that mapped into pathway i, and $Ncompounds$ is the total number of compounds we used from DrugBank and STITCH. The enrichment factor of a pathway is its propensity to be targeted by synergistic compound pairs in our screen.

3.1.5 Acknowledgement

The presented work is published as Pei, Fen, Hongchun Li, Mark J. Henderson, Steven A. Titus, Ajit Jadhav, Anton Simeonov, Murat Can Cobanoglu et al. "Connecting neuronal cell protective pathways and drug combinations in a Huntington's disease model through the application of quantitative systems pharmacology." *Scientific reports* 7, no. 1 (2017): 1-16. Drs. Mark E. Schurdak, Robert M. Friedlander, Andrew M. Stern and Timothy R. Lezon contributed to the design of the project, Dr. Murat Can Cobanoglu performed the initial computational screening of the compounds. Dr. Tongying Shun performed the analysis of drug combination

results. Drs. Mark J. Henderson, Steven A. Titus, Ajit Jadhav, Anton Simeonov, Seyed H. Mousavi, Lee McDermott, Prema Iyer, Michael Fioravanti, Diane Carlisle designed the performed the experiment assays. I performed the identification of drug-target interactions and pathway analysis, contributed to the analysis of drug combination. Drs. Ivet Bahar and D. Lansing Taylor supervised the project.

3.2 Drug Abuse: Addiction Progression Mechanism and the Effector Role of mTORC1

3.2.1 Introduction

Drug addiction is a chronic relapsing disorder characterized by compulsive, excessive, and self-damaging use of drugs of abuse. It is a debilitating condition that potentially leads to serious physiological injury, mental disorder and death, resulting in major health and social economic impacts worldwide (Lee, et al., 2016; Nestler, 2013). Substances with diverse chemical structures and mechanisms of action are known to cause addiction. Except for alcohol and tobacco, substances of abuse are commonly classified into six groups based on their primary targets or effects: cannabinoids (e.g. cannabis), opioids (e.g. morphine, heroin, fentanyl), CNS depressants (e.g. pentobarbital, diazepam), CNS stimulants (e.g. cocaine, amphetamine), hallucinogens (e.g. ketamine, lysergic acid diethylamide) and anabolic steroids (e.g. nandrolone, oxymetholone).

The primary actions of drugs of abuse have been well studied. In spite of the pleiotropy and heterogeneity of drugs of abuse, they share similar phenotypes: from acute intoxication to chronic dependence (Taylor, et al., 2013), the reinforcement shift from positive to negative through a three-stage cycle involving binge/intoxication, withdrawal/negative effect, and

preoccupation/anticipation (Koob and Volkow, 2016). Notably, virtually all drugs of abuse augment dopaminergic transmission in the reward system (Wise and Koob, 2014). However, the detailed cellular pathways of addiction processes are still far from known. For example, cocaine acts primarily as an inhibitor of dopamine (DA) transporter (DAT) and results in DA accumulation in the synapses of DA neurons (Shimada, et al., 1991; Volkow, et al., 1997). However, it has been shown that DA accumulation per se is not sufficient to account for the rewarding process associated with cocaine addiction; serotonin (5-HT) and noradrenaline/norepinephrine (NE) also play important roles (Rocha, et al., 1998; Sora, et al., 1998). Another example is ketamine, a nonselective antagonist for N-methyl-d-aspartate (NMDA) receptor (NMDAR), notably most effective in the amygdala and hippocampal regions of neurons (Collingridge, et al., 1983). In addition to its primary action, ketamine affects a number of other neurotransmitter receptors, including sigma-1 (Mendelsohn, et al., 1985) substance P (Okamoto, et al., 2003), opioid (Hustveit, et al., 1995), muscarinic acetylcholine (mACh) (Hirota, et al., 2002), nicotinic acetylcholine (nACh) (Coates and Flood, 2001), serotonin (Kapur and Seeman, 2002), and γ-aminobutyric acid (GABA) receptors (Hevers, et al., 2008). The promiscuity of drugs of abuse brings an additional layer of complexity, which prevents the development of efficient treatment against drug addiction.

In recent years there has been significant progress in the characterization of drug/target/pathway relations driven by the accumulation of drug-target interactions and pathways data, as well as the development of ML, in silico genomics, chemogenomics and QSP tools. Several innovative studies started to provide valuable information on substance abuse targets and pathways. For example, Li et al. curated 396 drug abuse related genes from the literature and identified five common pathways underlying the reward and addiction actions of cocaine, alcohol,

opioids and nicotine (Li, et al., 2008). Hu et al. analyzed the genes related to nicotine addiction via a pathway and network-based approach (Hu, et al., 2018). Biernacka et al. performed genome-wide analysis on 1165 alcohol-dependence cases and identified two pathways associated with alcohol dependence (Biernacka, et al., 2013). Xie et al. generated chemogenomics knowledgebases focused on G-protein coupled receptors (GPCRs) related to drugs of abuse in general (Xie, et al., 2014), and cannabinoids in particular (Xie, et al., 2016). Notably, these studies have shed light on selected categories or subgroups of drugs. There is a need to understand the intricate couplings between multiple pathways implicated in the cellular response to drugs of abuse, identify mechanisms common to various categories of drugs while distinguishing those unique to selected categories.

We undertake here such a systems-level approach using a dataset composed of six different categories of drugs of abuse. Following a QSP approach proposed earlier (Stern, et al., 2016), we provide a comprehensive, unbiased glimpse of the complex mechanisms implicated in addiction. Specifically, as shown in **Figure** 3.16, a set of 50 drugs of abuse with a diversity in chemical structures and pharmacological actions were collected as probes, and the known targets of these drugs as well as the targets predicted using our PMF method (Cobanoglu, et al., 2013) were analyzed to infer biological pathways associated with drug addiction. Our analysis yielded 142 known and 48 predicted targets and 173 pathways permitting us to identify both generic mechanisms regulating the responses to drug abuse as well as specific mechanisms associated with selected categories, which could both facilitate the development of auxiliary agents for treatment of addiction.

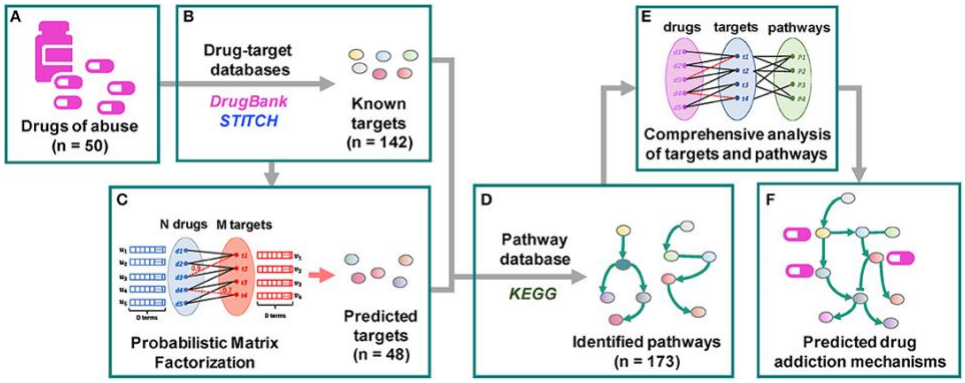


Figure 3.16 Workflow of the QSP analysis

(A) 50 drugs of abuse with a diversity of chemical structures and pharmacological actions were collected as probes. (B) 142 known targets of these drugs were identified through drug-target interaction database DrugBank and chemical-protein interaction database STITCH. (C) 48 predicted targets were predicted using our PMF method (Cobanoglu, et al., 2013). (D) 173 human pathways were inferred from the KEGG pathways database by mapping the known and predicted targets. (E,F) The pathways were grouped into 5 clusters. The functioning of identified targets and pathways and their involvement in drug addiction were comprehensively examined.

A key step in our approach is to identify the targets for drugs of abuse. There exists various drug-target interaction DBs, web servers and computational models, as summarized recently (Chen, et al., 2016). The drug-target interaction DBs utilized in this work are DrugBank (Wishart, et al., 2017) and STITCH (Szklarczyk, et al., 2016). DrugBank is a bioinformatics and cheminformatics resource that combines drug data with comprehensive target information. It is frequently updated, with the current version containing 10,562 drugs, 4,493 targets and corresponding 16,959 interactions. Since most of drugs of abuse are approved or withdrawn drugs, DrugBank is a good source for obtaining information on their interactions. STITCH, on the other hand, is much more extensive. It integrates chemical-protein interactions from experiments, other DBs, literature and predictions, resulting in data on 430,000 chemicals and 9,643,763 proteins

across 2,031 genomes. We have used the subset of human protein-chemicals data supported by experimental evidence. The method of approach adopted here is an important advance over our original PMF-based ML methodology for predicting drug-target interactions (Cobanoglu, et al., 2013). First, the approach originally developed for mining DrugBank has been extended to analyzing the STITCH DB, the content of which is 2-3 orders of magnitude larger than DrugBank (based on the respective numbers of interactions). Second, the information on predicted drug-target associations is complemented by pathway data on humans inferred from the KEGG pathway DB (December 2017 version) (Kanehisa, et al., 2017) upon pathway enrichment analysis of known and predicted targets. Third, the outputs are subjected to extensive analyses to detect recurrent patterns and formulate new hypotheses for preventive or therapeutic strategies against drug abuse.

3.2.2 Results

3.2.2.1 Functional similarity of drugs of abuse does not imply structural similarity, consistent with the multiplicity of their actions

Figure 3.17 presents a quantitative analysis of the functional and structural diversity of the examined n = 50 drugs of abuse, and the similarities among the m = 142 known targets of these addictive drugs. The n × n maps in **Figure** 3.17**A,B** display the drug-drug pairwise distances/dissimilarities based on their 2D fingerprints (**Figure** 3.17**A**), and their interaction patterns with their targets. **Figure** 3.17**C-D** display the corresponding dendrograms. The drugs are indexed and color-coded as in **Appendix Appendix B.1** and **Appendix Appendix B.2**. As expected, drugs belonging to the same functional category (same color) exhibit more similar interaction patterns (**Figure** 3.17**D**). However, we also note outliers, such as cocaine lying among opioids, as opposed to its categorization as a CNS stimulant, or promethazine, a CNS depressant,

lying among hallucinogens (shown by arrows). The peculiar behavior of cocaine is consistent with its high promiscuity (see **Figure** 3.18**A** for the number of targets associated with each examined drug). This type of promiscuity becomes even more apparent when the drugs are organized based on their structure (or 2D fingerprints; see section Materials and Methods) as may be seen in **Figure** 3.17**A**. For example, opioids (*cyan* labels/arc; clustered together in **Figure** 3.17**B,D** based on their interactions) are now distributed in two or more branches of the structure-based dendrogram in **Figure** 3.17**C**; likewise, CNS depressants (*blue*) and cannabinoids (*light brown*), grouped each as a single cluster in target-based dendrograms in **Figure** 3.17**D**, are now distributed into two or more clusters in **Figure** 3.17**C**.

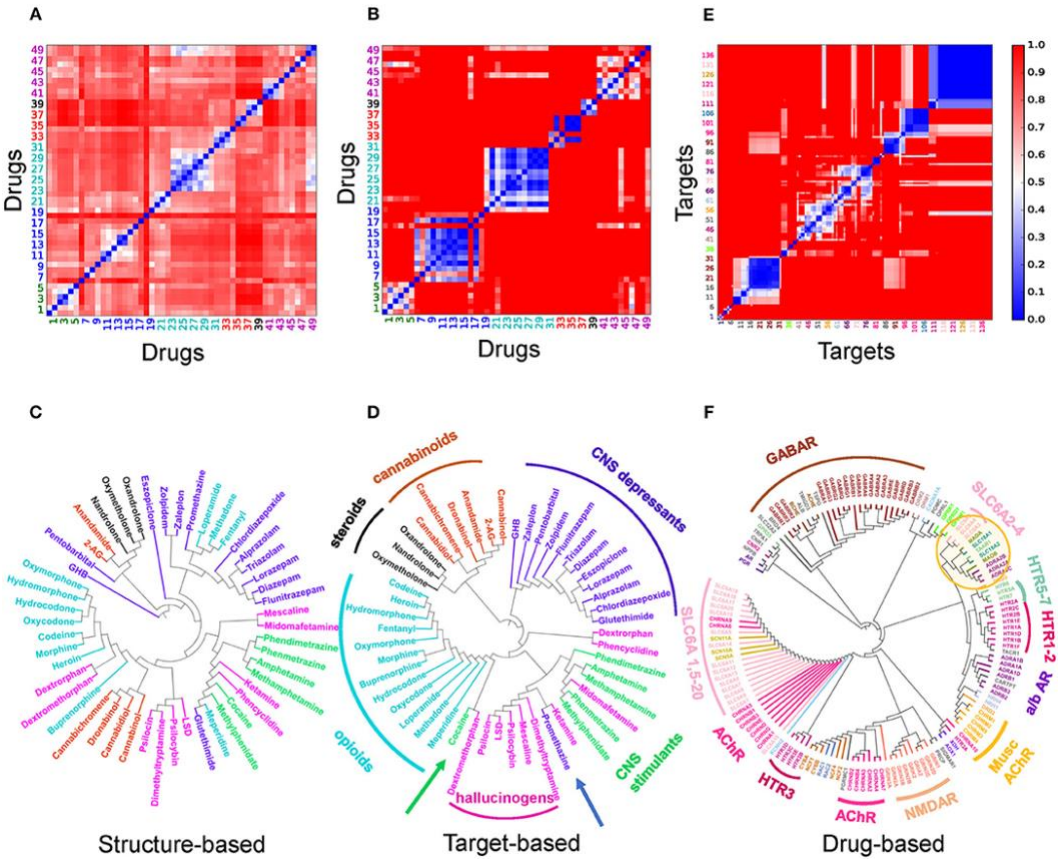


Figure 3.17 Dataset of 50 drugs of abuse: structure and interaction similarities, and classification of their targets

(A–D) Drug-drug distance maps for the studied 50 addictive drugs based on (A) 2D structure fingerprints and (B) interaction patterns with targets using the correlation cosines between their target vectors (see Materials and Methods), and corresponding dendrograms (C,D). The indices of drugs of abuse in (A,B) follow the same order as those used in Appedix Appendix B.1. The drug labels in (C,D) are color-coded based on their categories: CNS stimulants (*green*), CNS depressants (*blue*), opioids (*cyan*), cannabinoids (*light brown*), anabolic steroids (*black*) and hallucinogens (*magenta*). Note that the drugs of abuse in the same category do not necessarily show structural similarities nor similar interaction pattern with targets. (E) Pairwise distance map for the 142 known targets based on their interaction patterns with the 50 drugs. The indices in (E) follows the same order as those listed clockwise in the dendrogram (F). The tree maps in (C,D,F) are generated based on the respective distances values in the (A,B,E).

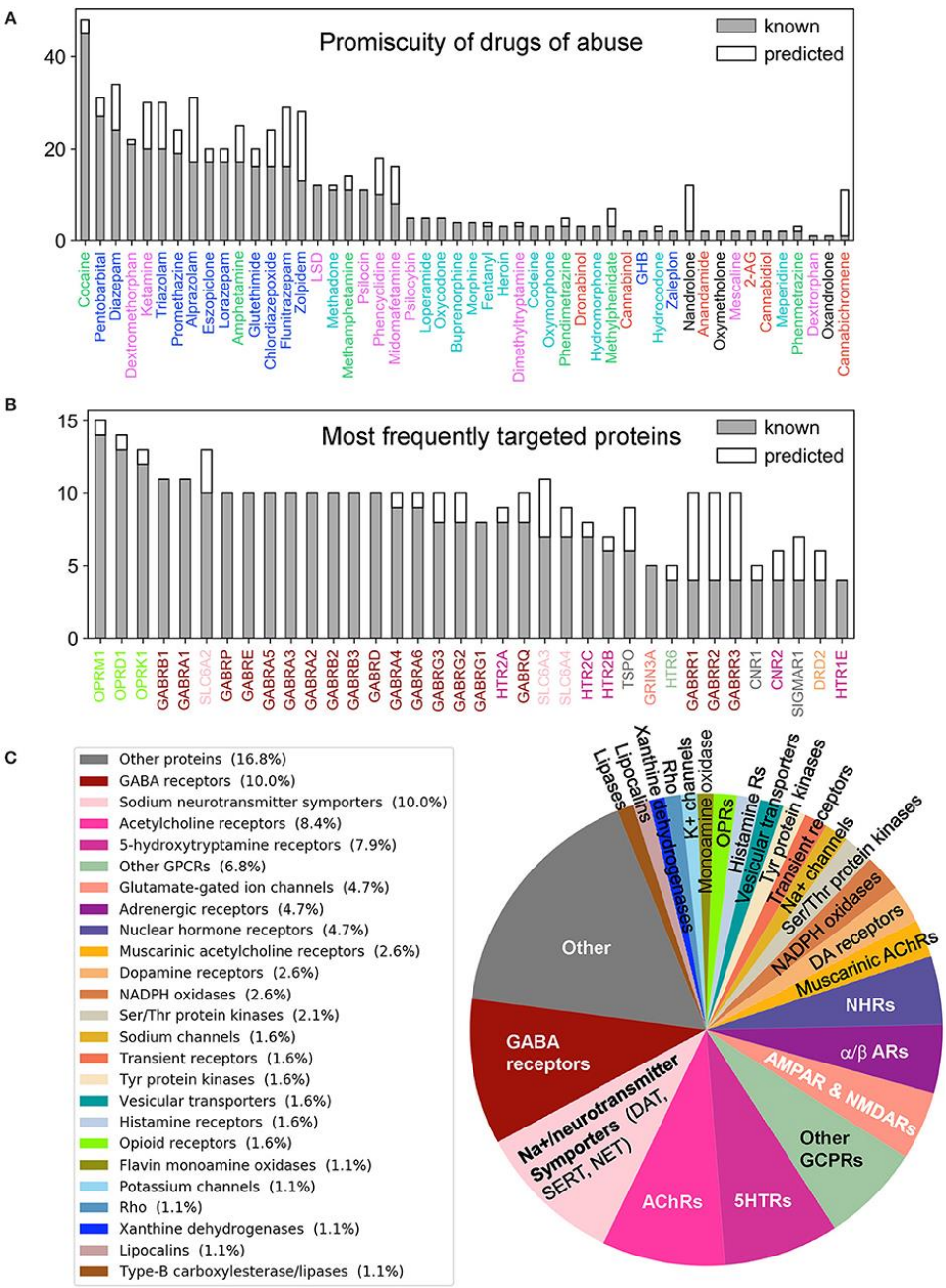


Figure 3.18 Promiscuity of drugs of abuse and their targets, and major families of proteins targeted by drugs of abuse

Number of known (*gray*) and predicted (*white*) interactions are shown by bars for (A) Drugs of abuse and (B) their targets. The examined set consists of 50 drugs of abuse and a total of 142 known and 48 predicted targets, involved in 445 (known) and 161 (predicted) interactions. (A) displays the number of interactions known or predicted for all 50 drugs. (B) Displays the results for the targets that interact with at least 4 known drugs (36 targets). The colors used

for names of drugs and targets are same as those used in Figure 3.17. (C) Displays the distribution of families of proteins targeted by drugs of abuse.

Overall these results suggest that the functional categorization of the drugs does not necessarily comply with their structural characteristics. The similar functionality presumably originates from targeting similar pathways, but the difference in the structure suggests that either their targets, or the binding sites on the same target, are different; or the binding is not selective enough such that multiple drugs can bind the same site. Consequently, a diversity of pathways or a multiplicity of cellular responses are triggered by the use and abuse of these drugs.

3.2.2.2 The Selected Drugs and Identified Targets Are Highly Diverse and Promiscuous

We evaluated the similarities between proteins targeted by drugs of abuse, based on their interaction patterns with the studied drugs of abuse. **Figure** 3.17**E-F** display the respective target-target distances, and corresponding dendrogram. We discern several groups of targets clustered together in consistency with their biological functions. For example, practically all GABA receptor subtypes (*brown*) are clustered together. This large cluster also includes the riboflavin transporter 2A (SLC52A2), which may be required for GABA release (Tritsch, et al., 2012). On the other hand, the different subtypes of serotonin (or 5-hydroxytryptamine, 5-HT) receptors (5HTRs) participate in distinct clusters pointing to the specificity of different subtypes vis-à-vis different drugs of abuse (labeled in **Figure** 3.17**F**).

The large majority of neurotransmitter transporters, such as Na+/Cl−-dependent GABA transporters (SLC6A1) and glycine transporter (SLC6A9) are in the same cluster (*pink*, labeled). Acetylcholine receptors also lie close to (or are even interspersed among) Na+/Cl−-dependent

neurotransmitter transporters, presumably due to shared drugs such as cocaine. However, the three transporters playing a crucial role in developing drug addiction, DAT, NE transporter (NET) and serotonin transporter (SERT) (labeled SLC6A2: NET, SLC6A3: DAT, SLC6A4: SERT) are distinguished by from all other neurotransmitter transporters as a completely disjoint group. The corresponding branch of the dendrogram (highlighted by the yellow circle) also includes vesicular amino acid transporters and trace amine-associated receptor 1 (TAAR1) known to interact with these transporters (Miller, 2011). We also note in the same branch two seemingly unrelated targets: flavin monoamine oxidase which draws attention to the role of oxidative events; and α2-adrenergic receptor subtypes A-C, which uses NE as a chemical messenger for mediating stimulant effects such as sensitization and reinstatement of drug seeking, and adenylate cyclase as another messenger to regulate cAMP levels (Sofuoglu and Sewell, 2009).

We identified 445 known interactions between these 50 drugs and 142 targets. We observe an average of 8.9 interactions per drug and 3.1 interactions per target. There are 23 promiscuous drugs that target at least 10 proteins as shown in **Figure** 3.18**A**. Cocaine, the most promiscuous psychostimulant, interacts with 45 known, and 3 predicted targets. It is known that cocaine binds DAT to lock it in the outward-facing state (OFS) and block the reuptake of DA. It similarly antagonizes SERT and NET (Heikkila, et al., 1975; Sora, et al., 1998), and also affects muscarinic acetylcholine receptors (mAChRs) M1 and M2 (Williams and Adinoff, 2008). Our PMF model also predicted a potential interaction between cocaine and M5. While this interaction is not listed in current DBs, there is experimental evidence suggesting that muscarinic AChR M5 plays an important role in reinforcing the effects of cocaine (Fink-Jensen, et al., 2003), in support of the PMF model prediction.

The PMF model enables us to predict novel targets. For example, anabolic steroid nandrolone has only two known interactions, and cannabinoid cannabichromene has one. However, 10 new targets were predicted with high confidence scores for each of them (**Figure** 3.19**A**). This is due to the data available in STITCH DB, which offers a large training dataset that enhances the performance of our ML approach. Overall, 89 new interactions were predicted for known targets, and 42 novel targets were predicted with 72 interactions. **Figure** 3.18**C** displays the distribution of all targets among different protein families. As will be further elaborated below, among the newly identified drug-target pairs, nandrolone-MAPK14 (mitogen-activated protein kinase 14, also known as p38α) and canabichromene-IKBKB (inhibitor of NFκ-B kinase subunit β) play a role in regulating mTORC1 signaling, which will be shown to be a potential effector of drug addiction.

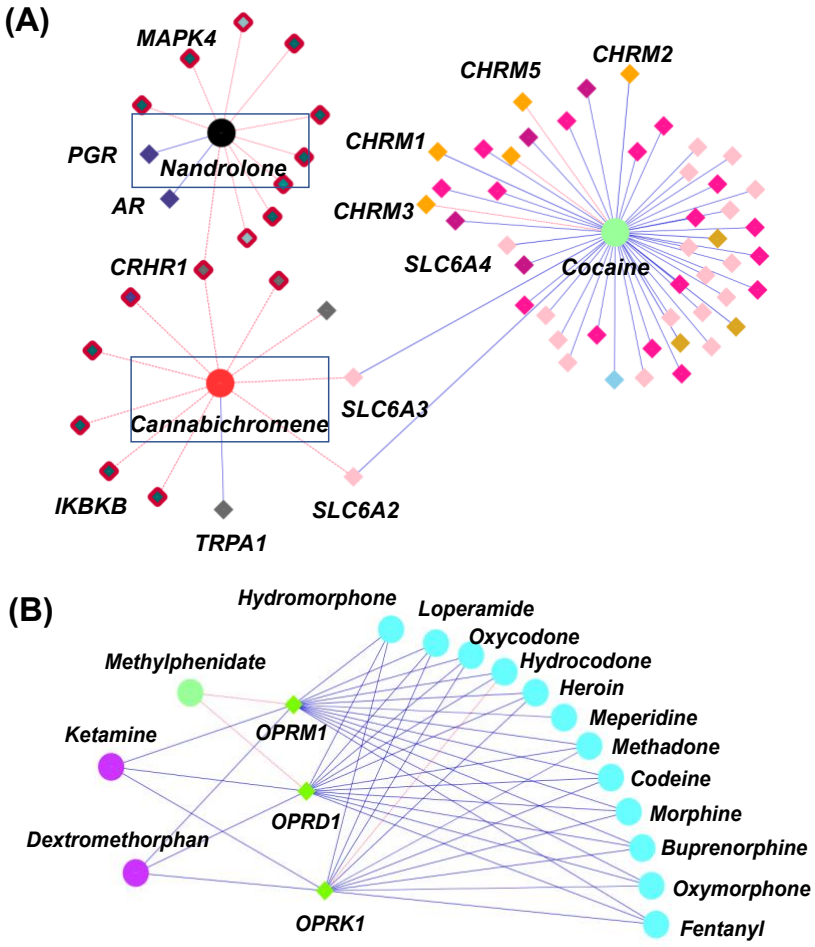


Figure 3.19 Prediction of new targets for known drugs of abuse

(A) Drug-target interactions are shown for three drugs of abuse, cocaine, cannabichromene and nandrolone. The colors of the drug nodes are consistent with the label colors in Figure 3.17C; the diamond nodes represent targets, the color of the target nodes is consistent with the label colors in Figure 3.17F; diamonds nodes with red borders represent predicted targets, diamond nodes without borders are the known targets; blue edges are known interactions and red edges are predicted interactions. Note that the norephinephrine transporter (NET/SLC6A2) and dopamine transporter (DAT/SLC6A3) are shared between cocaine and cannabichromene. Corticotropin-releasing factor receptor 1 (CRHR1) was predicted to be a new target shared between cannabichromene and nandrolone. (B) Drug-target interactions are shown for opioid receptors. Green diamonds are opioid receptors (OPRs): OPRM1, OPRD1 and OPRK1; 12 opioids interact with three OPRs either by existing evidence or prediction; hallucinogens ketamine and

dextromethorphan also interact with three OPRs; a novel interaction between OPRM1 and the CNS stimulant methylphenidate was predicted.

Turning to targets, three opioid receptors (OPRM1, OPRD1, and OPRL1) exhibit the highest level of promiscuity (**Figure** 3.19**B**). The μ-type opioid receptor (OPRM1) interacts with 14 known drugs including all opioids as well as ketamine and dextromethorphan. We also predicted a novel interaction between OPRM1 and the CNS stimulant methylphenidate. This is consistent with experimental observations that methylphenidate upregulates OPRM1's activity in the reward circuitry in a mouse model (Zhu, et al., 2011). Furthermore, tissue-based transcriptome analysis (Uhlen, et al., 2015) shows that 69% of our 190 targets are expressed in the brain, and 49 of them show elevated expression levels in the brain compared to other tissue types (**Table** 3.2). Among all the targets, NMDA receptor 1 (GRIN1) shows the highest elevated expression. It is also one of the top 5 enriched genes overall in the brain (Uhlen, et al., 2015).

Table 3.2 Enrichment of the 190 targets of addictive drugs in the brain and others

Category[a]	Count	Ref[b]	ER[c]	Targets[d]
Elevated in brain	49	1460	**3.4%**	**HTR5A; GABRB1; GRIA2; GABRG2; GABRG1; CHRNB2; GRIN2B; HTR2A; HTR2C; SLC6A17; GABRA5; GABRA4; GRIN1; GABRD; GABRA1; GABRB2; GABRA3; GABRA2; HRH3; P2RY12**; SLC6A1; SLC6A7; OPRL1; CNR1; CACNA1A; GRIN3A; SLC6A11; SLC6A15; CHRM5; CHRM4; CHRM3; CHRM1; CHRNA4; OPRK1; GABRB3; ADRA1B; GRIK2; GABRQ; GRIN2C; GRIN2A; HTR3B; OPRD1; GRIN2D; HTR1A; CCKBR; GLP1R; DRD5; CRHR1; DRD1
Moderately expressed in brain	82	13058	0.6%	SLC52A2; RAC2; PGRMC1; RAC1; CHRNB1; GABRP; CYBA; BRD4; PRCP; TSPO; SIGMAR1; ERBB2; NR3C1; HDAC6; S1PR1; EPHX2; MAPK14; PPARD; HMGCR; CTSS; CDK2; NR1H2; DPP7; NCF2; NCF1; SLC6A6; NCF4; GABRG3; SLC6A9; SLC6A8; ADRB2; BCHE; ADRB1; TMIGD3; ADRA2A; ADRA2C; HRH1; SLC6A13; SLC6A12; SLC6A16; CHRM2; CYBB; HTR7; AR; CARTPT; CHRNA2; POMC; ACHE; TACR1; ADRA1A; SLC6A20; ADRA1D; KCNH2; ALB; MAOB; MAOA; CHRNA7; HTR1B; GABRE; HTR1E; HTR1F; CHRNA5; PTGS2; IGF1R; PTGDR2; CRHR2; CALCRL; DHFR; PIK3CA; EGFR; THRB; IKBKB; PPARG; DHFR2; PTAFR; TYMS; SRD5A1; TRPV2; TRPV1; F10; P2RX7; CHEK1
Not detected in brain	59	5095	1.2%	CHRNE; SLC6A2; SLC6A5; SLC6A4; GABRR1; CHRNA10; AOX1; GPR55; TRPA1; GABRA6; CNR2; CHRNB3; CHRNB4; ADRA2B; XDH; SLC6A19; SLC6A18; PGR; HTR2B; NPPB; SLC6A14; SCN11A; HTR3E; GRIN3B; OPRM1; CHRNG; CHRNA1; CHRNA3; CHRNA9; TAAR1; ORM1; ORM2; DRD2; DRD3; ADRB3; HTR3C; CHRND; HTR3A; HTR1D; SLC18A2; SLC18A1; SCN5A; VDR; ESR2; CCKAR; GCGR; TRPV4; CALCA; SLC6A3; GABRR2; GABRR3; HRH4; SCN10A; HTR6; HTR3D; CHRNA6; SLC18A3; GLRA1; DRD4
Total	190	19613		

[a]Categories of targets (genes) are defined based on the mRNA expression levels of genes with the unit of Transcript Per Million (TPM). Elevated in brain: > 5-fold higher than other tissues (targets in bold are the most enriched in the brain); Moderately expressed in brain: at least 1 TPM in brain and other tissues; Not detected in brain: less than 1 TPM in brain.
[b]Number of genes in human proteome belong to each category.
[c]Enrichment ratio (ER) is the ratio of between numbers in Count and Ref columns. The ER in category of "Elevated in brain" is ~5-fold higher than that in "Moderately expressed in brain".
[d]Predicted targets are colored in *red*.

Taken together, the 50 selected drugs of abuse and the 142 known and 48 novel targets we identified cover a diversity of biological functions, are involved in many cellular pathways, and are generally promiscuous. In order to reveal the common mechanisms that underlie the development and escalation of drug addiction and also distinguish the effects specific to selected drugs, we proceed now to a detailed pathway analysis, presented next.

3.2.2.3 Pathway enrichment analysis reveals the major pathways implicated in various stages of addiction development

Our QSP analysis yielded a total of 173 pathways, including 114 associated with the known targets of the examined dataset of drugs of abuse, and 59 associated with the predicted targets. These pathways can be grouped in five categories (**Figure** 3.20**; Figure** 3.21**, Figure** 3.22):

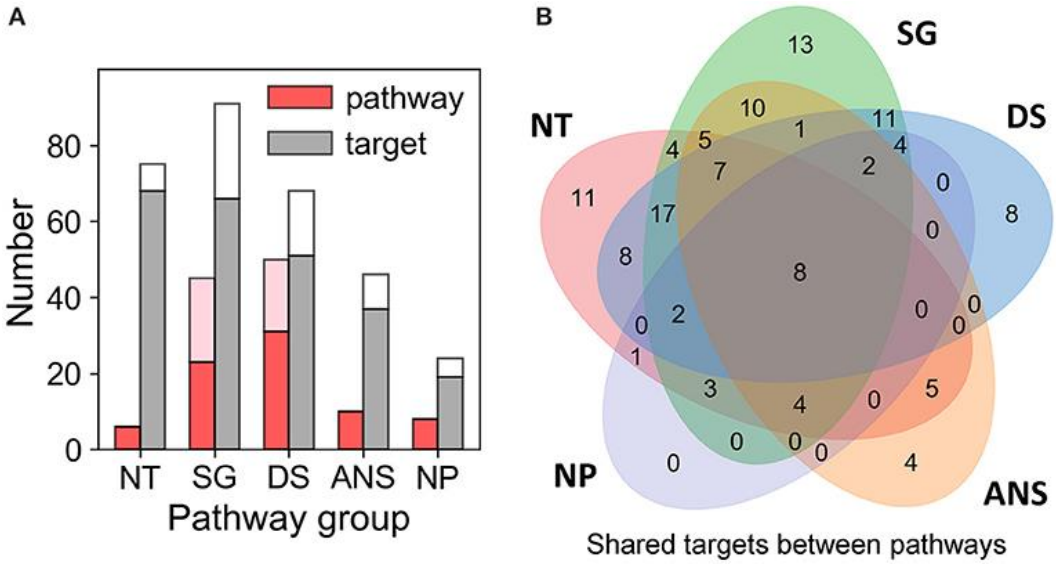


Figure 3.20 Results from pathway and target enrichments analysis

Five broad categories of pathways are distinguished among those involving the targets of drug abuse: NT, synaptic neurotransmission pathways; SG, signal transduction pathways; DS, disease-associated pathways; ANS, autonomic nervous system-innervation pathways; and NP, neuroplasticity related pathways. (A) Numbers of pathways (*red* bars) and targets (*gray* bars) of drug abuse lying in the five categories, based on data available in DrugBank and STITCH. The pink and white stacked bars are the corresponding numbers for pathways and targets additionally predicted by PMF. (B) Overlaps between the target content of the five pathway categories. Note that all targets belonging to the NP category pathways are represented in the other four categories.

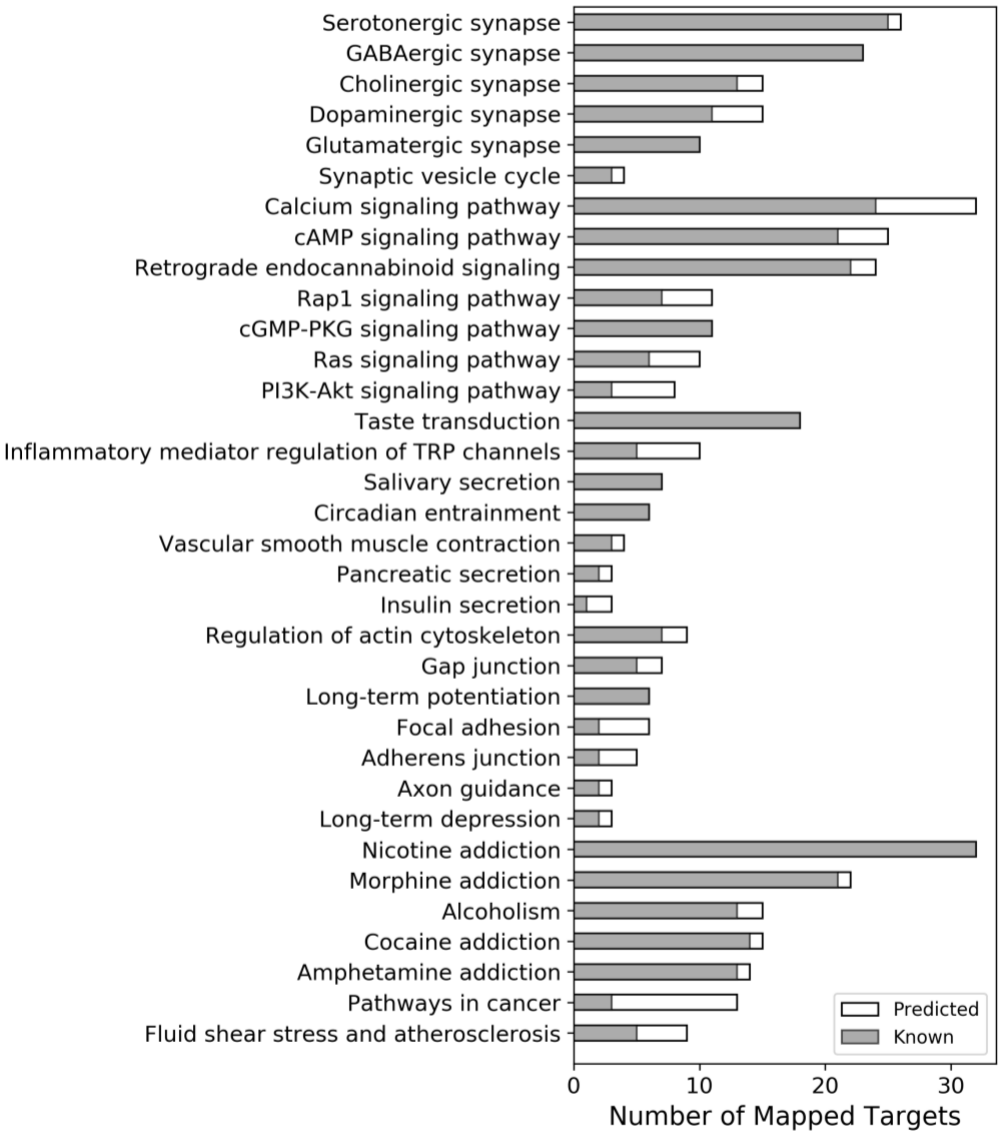


Figure 3.21 Pathways distinguished by the high propensity of targets of abused drugs

The bars here represent the number of known (*gray*) and predicted (*white*) proteins targeted by drugs of abuse in each pathway. Up to 7 pathways have been included in each case.

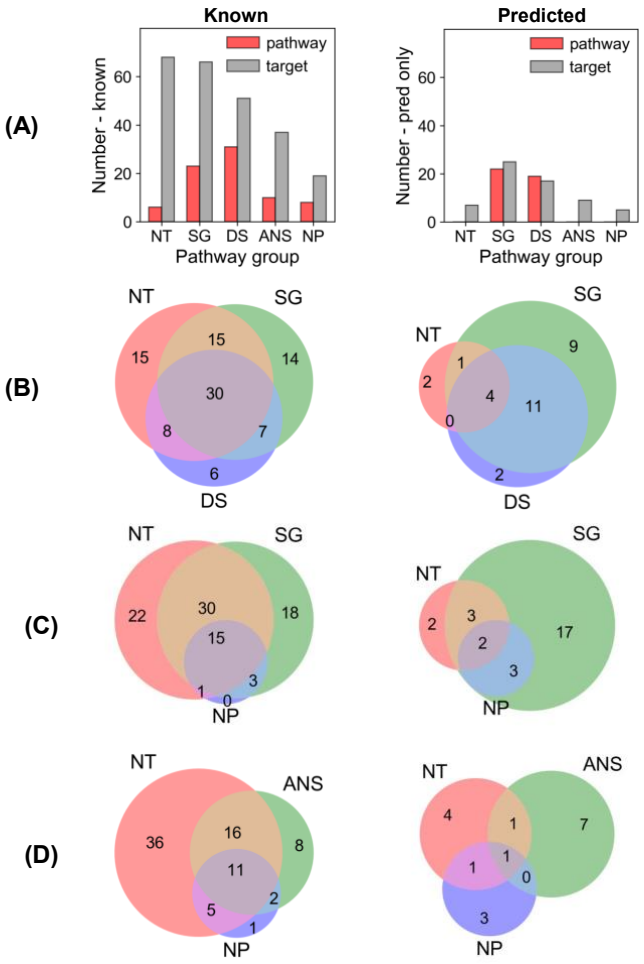


Figure 3.22 Pathway and target enrichments in five functional categories and the overlap of targets in different categories

(A) Numbers of drug addiction relative pathways (*red*) and targets (*gray*) of the five pathway categories (NT: neurotransmission related pathways; SG: signal transduction pathways; DS: disease pathways; ANS: ANS-innervation related pathways and NP: neuroplasticity related pathways) identified from known (*left*) targets and predicted targets (*right*) exclude those pathways and targets involved in known drug-target interactions, respectively. (B) Numbers of overlapped known (*left*) and predicted (*right*) targets between NT, DS and SG pathway categories. (C) Overlap results as (B) between NT, SG and NP pathway categories. (D) Overlap results as (B) between NT, ANS and NP pathway categories.

Synaptic Neurotransmission (NT). Six significantly enriched (with adjusted p-value < 0.05) pathways are associated with synaptic neurotransmission: dopaminergic, serotonergic, glutamatergic, synaptic vesicle cycle, cholinergic, and GABAergic synapses pathways. Sixty-eight known targets and 7 predicted targets are involved in these pathways. This is consistent with the fact that neurotransmission plays a dominant role in the rewarding system and is key to drug addiction (Volkow, et al., 2003).

Signal Transduction (SG). Forty-six intracellular signaling pathways were mapped by 92 targets comprised of 66 known and 25 predicted targets. Notably, many of these pathways have been reported to play a role in mediating the effects of drugs of abuse. These include the top five [calcium signaling (Li, et al., 2008), retrograde endocannabinoid signaling (Mechoulam and Parker, 2013), cGMP-PKG signaling (Shen, et al., 2016), cAMP signaling (Philibin, et al., 2011), and Rap1 signaling (Cahill, et al., 2016)] as well as some pathways with relatively low enrichment score (i.e., 0.2 < adjusted p-value), such as TNF signaling (Zhu, et al., 2018), MAPK signaling (Sun, et al., 2016), PI3K-Akt signaling (Neasta, et al., 2011), NF-κB signaling (Nennig and Schank, 2017), and mTOR signaling (Neasta, et al., 2014). We note that many receptors targeted by drugs of abuse take part in the KEGG neuroactive ligand-receptor interaction pathway. In the interest of focusing on intracellular signaling effects, we have not included these in the SG category; they are listed in the "Other Pathways".

Autonomic Nervous System-Innervation (ANS). We also identified 10 pathways regulating ANS-innervated systems such as endocrine secretion, taste transduction, and circadian entrainment. Recent evidences suggested drugs of abuse such as morphine (Al-Hasani and Bruchas, 2011) and cocaine (Moeller, et al., 1997; Prosser, et al., 2014) can influence ANS-

innervated systems and may contribute to the withdrawn symptoms associated with drug addiction. Thirty-seven known and 9 predicted targets take part in these pathways.

Neuroplasticity (NP). Eight enriched pathways with potential to alter the morphology of neurons, were found to be related to drug addiction. Among them, long-term potentiation (LTP) and long-term depression (LTD) are key to reward-related learning and addiction by modifying the fine tuning of dopaminergic firing (Jones and Bonci, 2005). Axon guidance pathway regulates the growth direction of neuron cells (Bahi and Dreyer, 2005). Regulation of actin cytoskeleton plays important role in morphological development and structural changes of neurons (Luo, 2002). Gap junctions connect neighboring neurons via intercellular channels that allow direct electrical communication (Belousov and Fontes, 2013) and regulate the efficiency of communication between electrical synapses (Belousov and Fontes, 2013). Nineteen known targets and 5 predicted targets are involved in these pathways. Insulin-like growth factor 1 receptor (IGF1R) is predicted as a target of drug triazolam. IGF1R is involved in LTP, adherens junction and focal adhesion pathways. It functions via canonical signaling pathways noted above in the SG category, such as the PI3K-Akt-mTOR and Ras-Raf-MAPK pathways (Lee, et al., 2016) and it plays important role in neuroplasticity (Lee, et al., 2016). We note that the NP group involves many pathways directly relevant to drug addiction (Bahi and Dreyer, 2005; Kalivas and Volkow, 2011; Moradi, et al., 2013; Rothenfluh and Cowan, 2013). There is no target unique to this particular group of pathways (Figure 3.20B). However, the fact that the targets belonging to the NP group are also shared by other groups consolidates the significance of these targets.

Disease-Associated Pathways (DS). Fifty enriched pathways mapped by 51 known and 17 predicted targets are associated with diverse diseases in different organs such as brain, liver, and lung. They also cover various drug addiction mechanisms including: nicotine addiction,

morphine addiction, cocaine addiction, amphetamine addiction, and alcoholism. Additionally, there are "other pathways" such as those involved in cell migration, differentiation, immune responses, and metabolic events.

Taken together, the enrichment analysis reveals five major categories of pathways that regulate the three stages of drug addiction cycle: (1) binge and intoxication, (2) withdrawal and negative affect, and (3) preoccupation and anticipation (or craving) (Koob and Volkow, 2010). Drugs of abuse directly affect neurotransmission pathways: they increase the accumulation of DA and other neurotransmitters in the synaptic and extrasynaptic regions, which in turn results in the hedonic feeling (stage 1) and triggers the DA reward system. Dysregulation of ANS-innervation pathways may cause negative effects and feelings (stage 2) and feedback to the CNS. Addictive drugs impair executive processes by disrupting the reward system (neurotransmission pathways) and imparting morphological changes via neuroplasticity pathways (e.g., LTD and LTP), which then result in craving (stage 3). Below, we present an in-depth analysis of the role of these pathways or their shared targets in drug addiction.

3.2.2.4 Selected targets shared by dominant pathways emerge as common mediators of drug addiction

We next analyzed the overlapping targets between the pathways in different functional categories.

First, we note that eight pleiotropic proteins are shared by all five categories (at the intersection of the five Venn diagrams in Figure 3.20B): AMPA receptor (subtype GluA2; GRIA2), NMDA receptors 1 and 2A-D (designated as GRIN1, GRIN2A, GRIN2B, GRIN2C, and GRIN2D) and voltage-dependent calcium channel Cav2.1 (or CACNA1A) as well as the predicted target phosphatidylinositol 3-kinase class 1A catalytic subunit α (PIK3CA).

Second, 15 proteins are distinguished as targets of four of these major pathways: Serotonin receptors 5HTR2-A, -B and -C), GABAA receptors 1-6 (GABRA1- GABRA6), β-1 adrenergic receptor 1 (ADRB1), Ras-related C3 botulinum toxin substrate 1 (RAC1; member of Rho family of GTPases), mAChR M3 (CHRM3) and DA receptor D2 (DRD2), and two predicted targets - p38α (MAPK14) and DA receptor D1 (DRD1).

AMPA receptor plays a crucial role in LTP and LTD, which are vital to neuroplasticity, memory and learning (Volkow, et al., 2016). Serotonin receptors, expressed in both the CNS and the peripheral nervous system (e.g., gastrointestinal tract), are responsible for anxiety, impulsivity, memory, mood, sleep, thermoregulation, blood pressure, gastrointestinal motility, and nausea (Pytliak, et al., 2011). They have been proposed to be therapeutic targets for treating cocaine use disorder (Howell and Cunningham, 2015). RAC1 is involved in five neuroplasticity pathways, including axon guidance, adherens junction and tight junction pathways, and 13 intracellular signal transduction pathways. It regulates neuroplasticity, as well as apoptosis and autophagy (Natsvlishvili, et al., 2015). DA receptor D2 is a target of 28 drugs of abuse (out of 50 examined here) and is involved in cAMP signaling, and gap junction pathways, in addition to dopaminergic signaling. It is implicated in reward mechanisms in the brain (Blum, et al., 1996) and the regulation of drug-seeking behaviors (Edwards, et al., 2007). Finally, PI3K turns out to be the most pleiotropic target among those targeted by drugs of abuse, being involved in 61 pathways identified here, including neuroplasticity pathways such as axon guidance, and several downstream signaling pathways such as PI3K-Akt, mTOR, Ras and Jak-STAT pathways.

Overall, the above listed 23 proteins shared by at least four different groups of pathways are distinguished here as highly pleiotropic proteins involved in the large majority of pathway categories implicated in drug abuse. Most of them are ligand- or voltage-gated ion channels or

neurotransmitter receptors, mainly AMPAR, NMDAR, Cav2.1, mAChR, and serotonin and DA receptors. However, it is interesting to note the targets PI3K and p38α, not currently reported in DrugBank and STITCH, emerge as highly pleiotropic targets of the drugs of abuse. These are suggested by the current analysis to directly or indirectly affect addiction development and await future experimental validation. Finally, a number of proteins take part in specific drug-abuse-related pathways and might serve as targets for selective treatments.

3.2.2.5 Pathway Enrichment Highlights the Interference of Drugs of Abuse with Synaptic Neurotransmission

It is broadly known that neurotransmitters such as DA, 5-HT, NE, endogenous opioids, ACh, endogenous cannabinoids, Glu, and GABA are implicated in drug addiction (Benarroch, 2012; Everitt and Robbins, 2005; Parolaro and Rubino, 2008; Tomkins and Sellers, 2001). Our analysis also showed that the serotonergic synapse (adjusted p-value p^*_i = 2.01E-18), GABAergic synapse (p^*_i = 1.19E-17), cholinergic synapse (p^*_i = 2.36E-07), dopaminergic synapse (p^*_i = 1.66E-06) and glutamatergic synapse (p^*_i = 1.86E-03) pathways were significantly enriched. A total number of 34 drugs (across six different groups) target at least one of these pathways. However, the identification of a pathway does not necessarily mean that the drug directly affects that particular neurotransmitter transport/signaling. There may be indirect effects due to the crosstalk between synaptic signaling pathways. For example, the ionotropic glutamate receptors NMDAR and AMPAR are also the downstream mediators in the dopaminergic synapse pathway. Likewise, GABARs are downstream mediators in the serotonergic synapse pathway.

In Figure 3.23, we highlight five major neurotransmission events that directly mediate addiction, and illustrate how eight drugs of abuse interfere with them. Despite the promiscuity of

the drugs of abuse, some selectively map onto a single synaptic neurotransmission pathway. For example, psilocin [a hallucinogen whose structure is similar to 5HT (Diaz, 1997)] interacts with several types of 5HTRs, regulating serotonergic synapse exclusively (see Figure 3.22). In contract, loperamide (not shown) affects all neurotransmission pathways by interacting with the voltage-dependent P/Q-type calcium channel (VGCC), regulating calcium flux on synapses. Cocaine targets four of these synaptic neurotransmission events (serotonergic, GABAergic, cholinergic, and dopaminergic synapses), through its interactions with 5-HT3R, sodium- and chloride-dependent GABA transporter (GAT), muscarinic (M1 and M2) and nicotinic AChRs, and DAT, respectively. Methadone affects three synaptic neurotransmissions, including serotonergic synapse, dopaminergic synapse, and glutamatergic synapse through the interactions with SERT, DAT, and glutamate receptors (NMDAR), respectively.

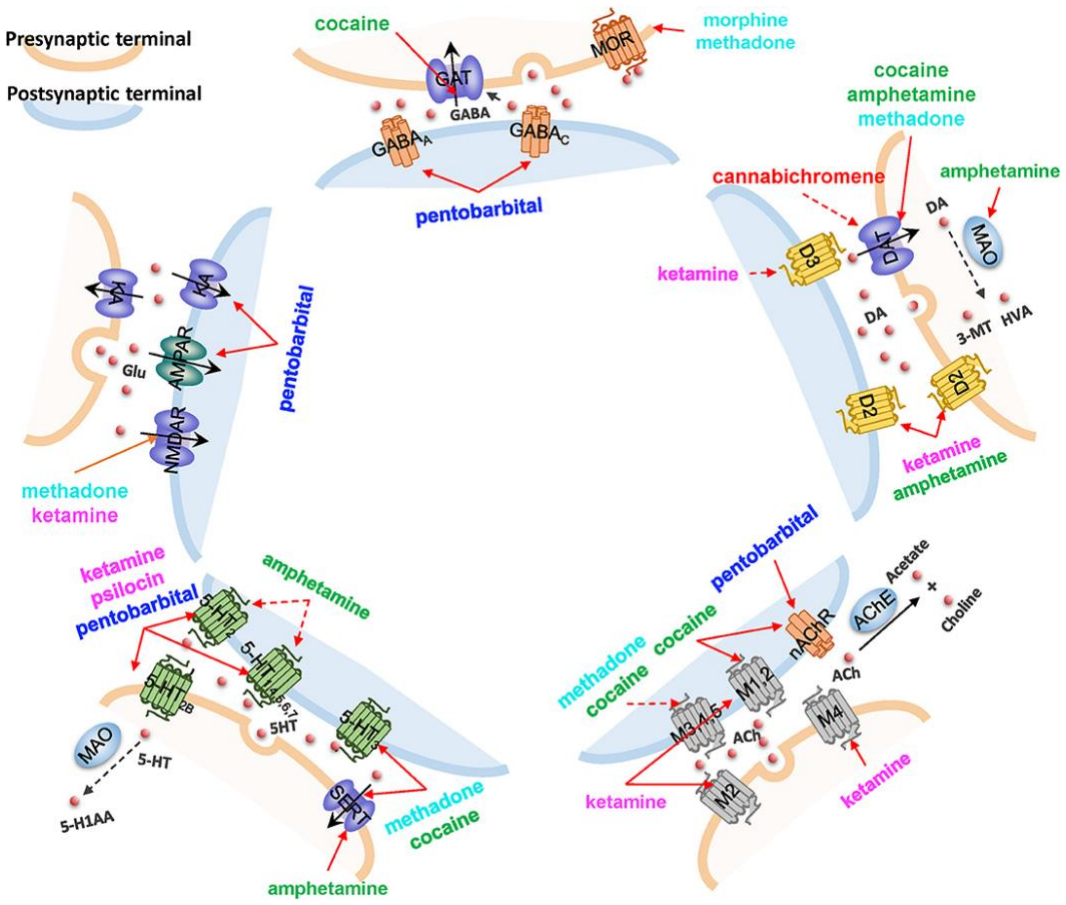


Figure 3.23 The impact of drugs of abuse on synaptic neurotransmission

Five major neurotransmission events are highlighted, mediated by (counterclockwise, starting from top): GABA receptors and transporters, ionotropic glutamate receptors (NMDAR and AMPAR) and cation channels, serotonin (5HT) receptors (5-HTR) and transporters (SERT), muscarinic or nicotinic AChRs, and dopamine (DA) receptors and transporters. Vesicular monoamine transporters (VMAT) that translocate DA are also shown. Drugs affecting the different pathways are listed, color coded with their categories, as presented in Figure 3.17. Solid red arrows indicate a known drug-target interaction, dashed red arrows indicate predicted drug-target interactions. Other molecules shown in the diagram are: KA, kainate receptor; MAO, monoamine oxidase; HVA, homovanillate; 3-MT, 3-methoxytyramine; MOR, mu-type opioid receptor; AChE, acetylcholinesterase; and 5-H1AA, 5-hydroxyindoleacetate.

It is worth noting that the current analysis helps us generate new hypotheses, yet to be experimentally validated, on the ways drugs of abuse affect neurotransmission. In addition to the

new role of the muscarinic AChR M5 suggested by the current analysis in section the selected drugs and identified targets are highly diverse and promiscuous, our PMF model suggested that cannabichromene, a cannabinoid whose primary target is the transient receptor (TRPA1), could interact with DAT and thus regulate dopaminergic transmission, which will require further examination.

The above synaptic neurotransmission events act as upstream signaling modules that "sense" the early effects of drug abuse. In the next section, we focus on the downstream signaling events elicited by drug abuse.

3.2.2.6 mTORC1 emerges as a potential downstream-effector activated by drugs abuse

The calcium-, cAMP-, Rap1-, Ras-, AMPK-, ErbB-, MAPK-, and PI3K-Akt-signaling pathways in the SG category crosstalk with each other and form a unified signaling network. As shown in **Figure** 3.24, ligand-binding to GPCRs modulates the production of cAMP, which leads to the activation of Rap1. Activated Rap1 modules the Ca2+ signaling by inducing the production of inositol triphosphate (IP3) and also activates the PI3K-Akt signaling cascade. Stimulations of ErbB family of receptor tyrosine kinases (related to epidermal growth factor receptor EGFR) as well as insulin-like growth factor receptor IGF1R trigger both PI3K-Akt and MAPK signaling cascades (proteins colored blue in Figure 3.24). Notably all these pathways merge and regulate a group of downstream proteins (shown in *dark yellow* in **Figure** 3.24); and at the center of this cluster lies the mammalian target of rapamycin (mTOR) complex 1 (mTORC1) which is likely to be synergistically regulated by all these merging pathways.

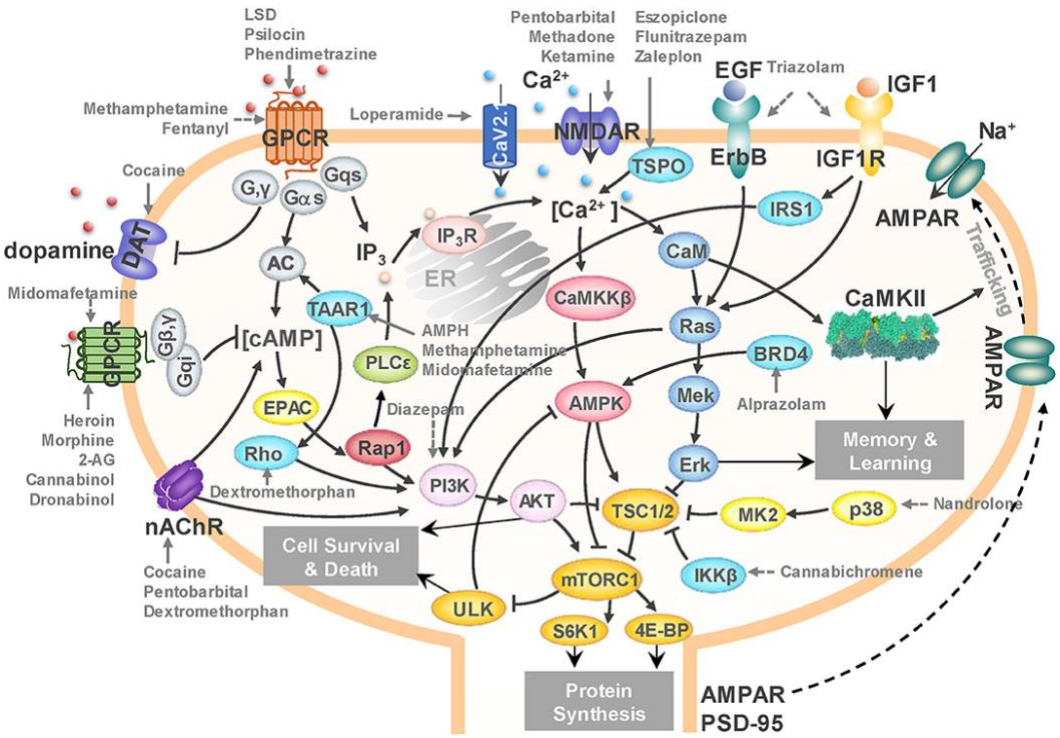


Figure 3.24 A unified signaling network mediates the effects of drugs of abuse

Black arrows represent the activation, inhibition, and translocation events during signal transduction. Solid gray arrows represent the known drug-target interactions. Dashed gray arrows represent predicted drug-target interactions. The diagram illustrates the targets of several drugs of abuse belonging to different categories: loperamide, fentanyl, heroin, morphine, and methadone from opioids; midomafetamine, ketamine, dextromethorphan, LSD, and psilocin from hallucinogens; triazolam, diazepam, alprazolam, pentobarbital, eszopiclone, flunitrazepam, and zaleplon from CNS depressants; cannabichromene, 2-AG, cannabinol, and dronabinol from cannabinoids; methamphetamine, cocaine, AMPH, and phendimetrazine from CNS stimulants; and nandrolone from anabolic steroids. mTORC1 emerges as a hub where the effects on several targets of addictive drugs appear to be consolidated to lead to cell death and/or protein synthesis in the CNS, and in particular, to AMPAR/PSD95 synthesis that induces morphological changes in the dendrites.

mTORC1 is not only a master regulator of autophagy (Rabanal-Ruiz, et al., 2017), but also controls protein synthesis and transcription (Ma and Blenis, 2009). It has been reported to promote neuroadaptation following exposure to drugs of abuse including cocaine, alcohol, morphine and

Δ9-tetrahydrocannabinol (THC) (Neasta, et al., 2014). Our results lead to the hypothesis that mTORC1 may act as a universal effector of the cellular response to drug abuse at an advanced (preoccupation and anticipation, or craving) stage, controlling the synthesis of selected proteins and ensuing cell growth, which may result in persistent alterations in the dendritic morphology and neuronal circuitry.

In **Figure** 3.24, selected interactions between drugs from different substance groups and their targets are highlighted using gray arrows. The figure illustrates that not only many known drug-target interactions, but also predicted ones involved in the unified signaling network. For example, our PMF model predicted that diazepam would interact with PI3K to influence mTORC1 signaling (dashed gray arrows denote predictions). It has been reported that Ro5-4864, a benzodiazepine derivative of diazepam suppresses activation of PI3K (Yousefi, et al., 2013), which corroborates our prediction. We further predicted that cannabichromene may interact with IκB kinase β (IKKβ) to regulate mTORC1 by inhibiting TSC1/2. Interestingly, another cannabinoid, arachidonoyl ethanolamine, is known to directly inhibits IKKβ (Sancho, et al., 2003). Taken together, our results suggest a unified network that underlies the development of drugs addiction, in which mTORC1 appears to play a key effector role.

3.2.3 Discussion

In the present study we focused on the targets and pathways affected by drugs of abuse, toward gaining a systems-level understanding of key players and dominant interactions that control the response to drug abuse and the development of drug addiction. Using ML methods, we focused on 50 drugs of abuse that form a chemically and functionally diverse set, and analyzed their 142 targets as well as the corresponding cellular pathways and their crosstalk. Our analysis identified:

(i) 48 additional proteins targeted by drugs of abuse, including PIK3CA, IKBKB, EGFR, and IGF1R, are shown to be key mediators of downstream effects of drug abuse.

(ii) 161 new interactions between the drugs of abuse and the known and predicted targets, including those between cocaine and M5, methylphenidate and OPRM1, and diazepam and PI3K, not reported in existing DBs, but supported by prior experiments, and others (e.g., the interactions of cannabichromene with IKBKB and DAT) that await experimental validation.

(iii) A dataset of 70 pathways, composed of 6 neurotransmission pathways, 46 signal transduction pathways, 8 neuroplasticity pathways and 10 autonomic nervous system innervation pathways which are proposed to govern different stages of the molecular, cellular and tissue level responses to drug abuse and in addiction development.

Overall, our comprehensive analysis led to new hypotheses on drug-target interactions and signaling and regulation mechanism elicited by drugs of abuse in general, along with those on selected targets and pathways for specific drugs. Below we elaborate on the biological and biomedical implications of these findings.

3.2.3.1 Persistent restructuring in neuronal systems as a feature underlying drug addiction

Enriched pathways in the neuroplasticity category include gap junction, LTP, LDP, adherens junction, regulation of actin cytoskeleton, focal adhesion, axon guidance, and tight junction. These are responsible for the changes in the morphology of dendrites. For instance, DA regulates excitatory synaptic plasticity by modulating the strength and size of synapses through LTP and LTD (De Roo, et al., 2008; Volkow and Morales, 2015). The restructuring of dendritic spines involves the rearrangements of cytoskeleton and actin-myosin (Volkow and Morales, 2015). The axon guidance molecules guide the direction of neuronal growth.

Drugs of abuse can induce the changes in CNS through these pathways. For example, chronic exposure to cocaine increases dendritic spine density in medium spiny neurons (Russo, et al., 2010). The disruption in axon guidance pathway and alteration in synaptic geometry can result in drug-related plasticity (Bahi and Dreyer, 2005). The persistent restructuring in the CNS caused by drugs of abuse is responsible for long-term behavioral plasticity driving addiction (Russo, et al., 2010; Volkow, et al., 2003; Volkow and Morales, 2015). As will be further discussed below, mTORC1 plays a central role in the synthesis of new proteins (e.g., AMPARs) and thereby neuronal (dendrites) growth, alteration of the synaptic geometry and therefore rewiring of the neuronal circuitry.

3.2.3.2 ANS may mediate the negative-reinforcement of drug addiction

The current study further points to pathways regulating the ANS-innervated systems. As the NP pathways influence the neuroplasticity in the ANS, we hypothesize that drugs of abuse might induce a persistent restructuring in the ANS as well. The drug-related plasticity in ANS may lead to the dysregulation of ANS-innervated systems and cause negative effects and feelings during the second stage of drug addiction. Drug addiction is well known as a brain disease (Volkow and Morales, 2015). However, many drugs of abuse can disrupt the activity of ANS and cause disorders in ANS-innervated systems (Al-Hasani and Bruchas, 2011; Huang, 2017). For example, opioids (e.g., morphine) alter neuronal excitability and neurotransmission in the ANS (Wood and Galligan, 2004), and induce disorders in gastrointestinal system, smooth muscle, skin, cardiovascular, and immune system (Al-Hasani and Bruchas, 2011). Cannabinoids (e.g., THC) modulate the exocytotic NE release in ANS-innervated organs through presynaptic cannabinoid receptors (Ishac, et al., 1996).

The pathways we identified in the ANS category regulate insulin secretion, gastric acid secretion, vascular smooth muscle contraction, pancreatic secretion, salivary secretion, and renin secretion. Their dysfunction may be associated with the autonomic withdrawal syndrome, such as thermoregulatory disorder (chills and sweats) and gastrointestinal upset (abdominal cramps and diarrhea), which has been observed in drug/substance users (Wise and Koob, 2014). In addition, the stress and depression caused by these negative effects may be part of the negative reinforcement of drug addiction (Koob and Le Moal, 2001; Self and Nestler, 1995). In other words, the drug induced ANS disorders can feedback to CNS and mediate the negative reinforcement. Compared to the structural changes in CNS, the disorder and persistent restructuring in ANS is less studied and it could be a future direction in the study of development of drug addiction and related diseases.

3.2.3.3 mTORC1 appears as a key mediator of cellular morphological changes elicited in response to continued drug abuse

The functioning and regulation of mTOR signaling has been elucidated over the past two decades. It became clear that mTORC1 plays a crucial role in regulating diverse cellular processes including protein synthesis, autophagy, lipid metabolism, and mitochondrial biogenesis (Saxton and Sabatini, 2017). In the brain, mTORC1 coordinates neural development, circuit formation, synaptic plasticity, and long-term memory (Lipton and Sahin, 2014). The dysregulation of mTORC1 pathway is associated with many neurodevelopmental and neurodegenerative diseases such as Parkinson's disease and Alzheimer's disease. mTORC1 has been noted to be an important mediator of the development of drug addiction and relapse vulnerability (Dayas, et al., 2012). Accumulating evidences show that pharmacological inhibition of mTORC1 (often through rapamycin treatment) can prevent sensitization of methamphetamine-induced place preference

(Narita, et al., 2005), reduce craving in heroin addicts (Shi, et al., 2009), attenuate the expression of alcohol-induced locomotor sensitization (Neasta, et al., 2010), suppress the expression of cocaine-induced place preference (Bailey, et al., 2012), protect against the expression of drug-seeking and relapse by reducing AMPAR (GluA1) and CaMKII levels (James, et al., 2014), and inhibit reconsolidation of morphine-associated memories (Lin, et al., 2014).

Our unbiased computational analysis based on a diverse set of 50 drugs of abuse supports the hypothesis that mTORC1 may act as a universal effector or controller of neuroadaptations induced by drugs of abuse (Neasta, et al., 2014). The major signal transduction pathways we identified that involve targets of drugs of abuse interconnect and converge to the mTORC1 signaling cascade (**Figure** 3.24). Most drugs of abuse in our list target upstream regulators of mTORC1, including membrane receptors (e.g., GPCRs, RTKs and NMDAR), kinases (e.g., PI3K, p38α, and IKKβ), and ion channels (e.g., CaV2.1 and TRPV2). Notably, the impact of some of these known or predicted targets has been experimentally confirmed. For example, blockade of the known target NMDAR using MK801 reduces the amnesic-like effects of cannabinoid THC (Puighermanal, et al., 2009). Likewise, inhibition of PI3K (a predicted target) by LY294002 suppresses morphine-induced place preference in rats (Cui, et al., 2010) and the expression of cocaine-sensitization (Izzo, et al., 2002). Our results thus provide a pool of candidate targets implicated in cellular responses to addictive drugs, which await to be consolidated by further tests.

The downstream effectors of mTORC1, which specifically mediate drug behavioral plasticity is far from known. mTORC1 can mediate the activation of S6Ks and 4E-BPs, which leads to increased production of proteins required for synaptic plasticity including AMPAR and PSD-95 (Dayas, et al., 2012). EM reconstruction of hippocampal neuropil showed the variability in the size and shape of dendrites depending on synaptic activity (Bartol, et al., 2015), which in

turn correlates with information storage. Recently studies have revealed that Atg5- and Atg7-dependent autophagy in dopaminergic neurons regulates cellular and behavioral responses to morphine (Su, et al., 2017). Cocaine exposure results in ER stress-induced and mTORC1-dependent autophagy (Guo, et al., 2015). Fentanyl induces autophagy via activation of ROS/MAPK pathway (Yao, et al., 2016). Methamphetamine induces autophagy through the κ-opioid receptor (Ma, et al., 2014). These observations are consistent with the currently inferred role of mTORC1 as a downstream effector of cellular responses to drug addiction.

3.2.3.4 Drug repurposing opportunities for combating drug addiction

Autophagy modulating drugs have been shown to have therapeutic effects against liver and lung diseases. The signaling network presented in **Figure** 3.24 involves many targets of such drugs. For instance, carbamazepine affects IP3 production and enhances autophagy via calcium-AMPK-mTORC1 pathway (Hidvegi, et al., 2010). It has been identified as a potential drug for treating α1-antitrypsin deficiency, hepatic fibrosis, and lung proteinopathy (Hidvegi, et al., 2010; Hidvegi, et al., 2015). Rapamycin is a potential drug for lung disease such as fibrosis (Abdulrahman, et al., 2011; Patel, et al., 2012). Other liver and lung drugs which facilitate the removal of aggregates by promoting autophagy may also affect drug-related neurodegenerative disorders. **Table** 3.3 summarizes 15 autophagy-modulating drugs for liver and lung diseases. Target identification and pathway analysis of this subset of drugs using the same protocol as those adopted for the 50 drugs of abuse indeed confirmed that drugs of abuse and liver/lung drugs share many common pathways (**Figure** 3.25). Notably, among those pathways, neuroactive ligand-receptor interactions, calcium signaling, and serotonergic synapse pathways are among the top 10 enriched pathways of both drugs of abuse and liver/lung drugs. Amphetamine addiction and alcoholism are also enriched by targets of liver/lung drugs. Thus, an interesting future direction is

to examine whether autophagy modulating drugs for liver and lung diseases could be repurposed, if necessary, by suitable refinements to increase their selectivity, for treating drug addiction.

Table 3.3 Fifteen drugs for liver and lung diseases

No.	Drug name	DrugBank ID	Pubchem ID	Disease	Reference
1	Carbamazepine	DB00564	2554	α1-antitrypsin deficiency; hepatic fibrosis; lung Proteinopathy	(Hidvegi, et al., 2010; Hidvegi, et al., 2015)
2	Fluphenazine	DB00623	3372	α1-antitrypsin deficiency; lung Proteinopathy	(Hidvegi, et al., 2015; Li, et al., 2014)
3	Cantharidin	NA	5944	α1-antitrypsin deficiency	(Krichevsky, et al., 2010)
4	Pimozide	DB01100	16362	α1-antitrypsin deficiency	(Park, et al., 2010)
5	Tamoxifen	DB00675	2733525	α1-antitrypsin deficiency	(de Mol, et al., 2010)
6	Phenylbutyric Acid	NA	4775	α1-antitrypsin deficiency	(Burrows, et al., 2000)
7	Vorinostat	NA	5311	α1-antitrypsin deficiency	(Bouchecareilh, et al., 2012)
8	Glycerol	DB09462	753	α1-antitrypsin deficiency	(Burrows, et al., 2000)
9	Fluspirilene	DB04842	3396	α1-antitrypsin deficiency	(O'Reilly, et al., 2014)
10	Ezetimibe	NA	150311	α1-antitrypsin deficiency	(Yamamura, et al., 2014)
11	Gemfibrozil	DB01241	3463	COPD-emphysema	(Bodas, et al., 2017)
12	Fisetin	NA	5281614	COPD-emphysema	(Bodas, et al., 2017)
13	Cysteamine	NA	6058	COPD-emphysema	(Bodas, et al., 2016; Shivalingappa, et al., 2016)
14	S-Nitrosoglutathione	NA	104858	COPD-emphysema	(Bodas, et al., 2017)
15	Rapamycin	NA	5284616	lung fibrosis; cystic fibrosis	(Abdulrahman, et al., 2011; Patel, et al., 2012) (Kouvelas, et al., 2008)

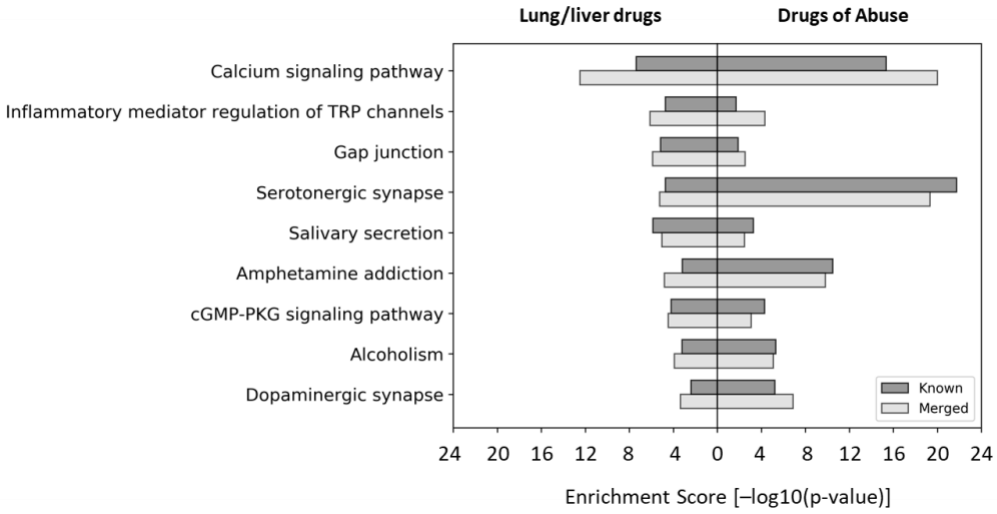


Figure 3.25 The enriched overlapping pathways for lung/liver drugs and drugs of abuse

Enrichment p-values based on known (*gray*) targets or merged (*lightgray*) targets (including both known and predicted targets) are calculated separately for each pathway, the pathways are ranked by the p-value calculated by merged targets of lung/liver drugs. There are 10 overlapping pathways with enrichment score (-log10(p-value)) over 2, between the pathways enriched by the examined sets of lung/liver drugs and drugs of abuse. Note that the neuroactive ligand-receptor interaction pathway, which has the highest enrichment score (known: 16.8, merged: 6.5 for lung/liver drugs; known: 104.6, merged: 93.05 for drugs of abuse) is not shown for visualization purpose.

3.2.4 Materials and Methods

3.2.4.1 Selection of drugs of abuse and their known targets

We selected as input 50 drugs commonly known as drugs of abuse using two basic criteria: (i) diversity in terms of structure and mode of action, and (ii) availability of information on at least one human target protein in DrugBank v5 or STITCH v5. The selected drugs represent six different categories: CNS stimulants, CNS depressants, opioids, cannabinoids, anabolic steroids, and hallucinogens (see **Appendix Appendix B.1** and **Appendix Appendix B.2**).

A dataset of 142 known targets were retrieved from DrugBank and STITCH DBs for these 50 drugs. The list includes all targets reported for these drugs in DrugBank, and those with high confidence score, based on experiments, reported in STITCH. Each chemical-target interaction is annotated with five confidence scores in STITCH: experimental, DB, text-mining, prediction, and a combination score of the previous four, each ranging from 0 to 1. We selected the human protein targets with experimental confidence scores of 0.4 or higher, ending up with 142 targets and 445 drug-target interactions.

Structure-based and interaction-pattern-based similarities between pairs of drugs were evaluated using two different criteria. The former was based on structure-based distance calculated as the Tanimoto distance between their 2D structure fingerprints. Tanimoto distances were evaluated using Python RDKit suite (RDKit: Open-Source Cheminformatics Software. https://www.rdkit.org/). Similarities based on their interactions patterns with known targets were evaluated by evaluating target-based distances. To this aim, we represented each drug i by a 142-dimensional "target vector" di, the entries of which represent the known targets and are assigned values of 0 or 1, depending on the existence/observation of an interaction between the corresponding target and drug i. Interaction-pattern similarities between drug pairs i and j were evaluated by calculating the correlation cosine $\cos(d_i \cdot d_j) = (d_i \cdot d_j)/(|d_i|\,|d_j|)$ between these vectors, and the corresponding cosine distance is $[1-\cos(d_i \cdot d_j)]$. Likewise, ligand-based distances between target pairs i and j were evaluated as the cosine distance between the 50-dimensional vectors t_i and t_j corresponding to the two targets, the entries of which are 0 or 1 depending on absence or existence of an interaction between the target and the corresponding drug of abuse.

3.2.4.2 Probabilistic matrix factorization (PMF) based drug-target interaction prediction

Using this PMF based ML approach (Cobanoglu, et al., 2013; Cobanoglu, et al., 2015), we trained two PMF models, one based on 11,681 drug-target interactions between 6,640 drugs and 2,255 targets from DrugBank v5, and the other based on 8,579,843 chemical-target interactions for 311,507 chemicals and 9,457 targets from STITCH v5 human experimentally confirmed subset, respectively. We evaluated the confidence scores in the range [0, 1] for each predicted drug-target interaction, in both cases. We selected the interactions with confidence scores higher than 0.7 within the top 10 predicted targets for each input drug. This led to 161 novel interactions identified between 27 out of the 50 input drugs and 89 targets (composed of 41 known and 48 novel targets).

3.2.4.3 Pathway Enrichment Analysis

We mapped the 50 drugs with 142 known and 48 predicted targets to the KEGG pathways (version December 2017, homo sapiens). 114 and 173 pathways were mapped by 142 known targets and all targets (both known and predicted) respectively. In order to prioritize enriched pathways, we calculated the hypergeometric p-values based on the targets as the enrichment score as described in Chapter 1.5.6.

The source code used for generating the results reported in this study is available at https://github.com/Fengithub/DA.

3.2.5 Acknowledgments

The presented work is published as Pei, Fen, Hongchun Li, Bing Liu, and Ivet Bahar. "Quantitative systems pharmacological analysis of drugs of abuse reveals the pleiotropy of their

targets and the effector role of mTORC1." *Frontiers in pharmacology* 10 (2019): 191. Dr. Hongchun Li and I contributed to the implementation and analysis of the research. Dr. Bing Liu contributed to the analysis of the results, the project was designed and supervised by Dr. Ivet Bahar.

3.3 Non-Alcoholic Fatty Liver Disease (NAFLD): Identification of Repurposable Drugs

3.3.1 Introduction

Non-alcoholic fatty liver disease (NAFLD) comprises a spectrum of progressive disease stages from simple steatosis (fatty liver) termed NAFL to a more serious condition, nonalcoholic steatohepatitis (NASH), involving inflammation, hepatocyte damage (i.e., ballooning) and most often pericellular fibrosis (Brunt, et al., 2015; Satapathy and Sanyal, 2015). NASH itself is a risk factor for cirrhosis and end-stage liver disease requiring liver transplantation and for hepatocellular carcinoma (HCC) that insidiously can progress asymptomatically before cirrhosis is diagnosed (Loomba and Sanyal, 2013; Mikolasevic, et al., 2018). The prevalence of NAFLD is approximately 25% across adult populations world-wide with the proportion of those with NASH predicted to increase over the next decade (Demir, et al., 2015). Despite the major public health problem NAFLD presents and the economic burden it exacts, no single drug has yet been specifically approved for NAFLD (Polyzos, et al., 2020). The challenges facing this unmet need appear to be rooted in the complexity and intrinsic heterogeneity of NAFLD that has variable rates of progression and clinical manifestations across individual patients, with most patients progressing to advanced fibrosis over decades in contrast to approximately 20% who progress

much more rapidly (McPherson, et al., 2015; Sanyal, 2019).This heterogeneity appears to reflect the complex pathogenesis of NAFLD involving diverse but convergent signaling cues from the environment, the microbiome, metabolism, comorbidities, and genetic risk factors (Friedman, et al., 2018).

Therefore to more accurately predict disease progression and response to emerging therapies for NAFLD, the research community has adopted systems-based approaches such as QSP that can comprehensively and unbiasedly integrate molecular, cell, and clinical data to generate predictive models of disease progression (Mardinoglu, et al., 2018). These can then be iteratively tested in experimental models to identify emergent disease-specific networks and predictive biomarkers mechanistically linked to NAFLD pathogenesis (Taylor, et al., 2019; Wooden, et al., 2017). An overarching goal of implementing a QSP approach for addressing NAFLD heterogeneity is to identify NAFLD subtypes having distinguishable mechanisms of disease progression. It is hypothesized that this disease subclassification that has remained elusive thus far, will enable precision medicine, leading to therapeutic advances optimized for individual patients (Stern, et al., 2016). The integration of molecular, cell, and clinical data has begun to generate molecular signatures for NAFLD progression (Middleton, et al., 2018) but the experimental testing of predicted mechanistic hypotheses and therapeutic strategies has been limited by the availability of preclinical models that recapitulate critical aspects of the human disease (Mann, et al., 2016). For example, whereas steatosis can be recapitulated in murine models, fibrosis, a key clinical biomarker of NASH progression, is not generally observed (Hebbard and George, 2011). Furthermore, even if significant fibrosis was observed in animal models, it is unlikely that they would mimic the disease heterogeneity observed in the clinic. To meet the need for developing preclinical patient-specific NAFLD models we and others have developed MPS

that recapitulate critical aspects of normal acinus multicellular architecture and function (Edmondson, et al., 2014; Taylor, et al., 2019). When these systems are perfused with non-esterified fatty acids, glucose, insulin, and inflammatory cytokines mimicking a metabolic syndrome milieu that promotes hepatic insulin resistance, clinically relevant NASH-like changes were observed (Feaver, et al., 2016; Kostrzewski, et al., 2020). These changes include increases in *de novo* lipogenesis, gluconeogenesis, oxidative and ER stress, production of inflammatory and fibrogenic cytokines accompanied by hepatocyte injury and enhanced stellate cell activation. Overall, human liver, biomimetic MPS appear to mirror key aspects of NAFLD progression and provide a model consistent with the conceptual framework that NAFLD represents the hepatic expression of the metabolic syndrome in the majority of patients (Lee-Montiel, et al., 2017; Li, et al., 2018; Vernetti, et al., 2017; Vernetti, et al., 2016).

Herein, we describe the implementation of a QSP-based platform (**Figure** 3.26) that starts with the computational modeling of individual patient-derived hepatic RNAseq data encompassing a full spectrum of NAFLD disease states from simple steatosis, NASH, to cirrhosis (Gerhard, et al., 2018). Gene signatures specifically associated with NAFLD progression are derived and then approved and investigational drugs that are predicted to normalize these gene signatures are identified. These drugs, prioritized by independent but convergent criteria and serving as mechanistic probes with the potential to be repurposed for a NAFLD indication, are tested in the liver MPS to determine their predicted effects on gene expression and corresponding ability to halt or reverse NAFLD progression.

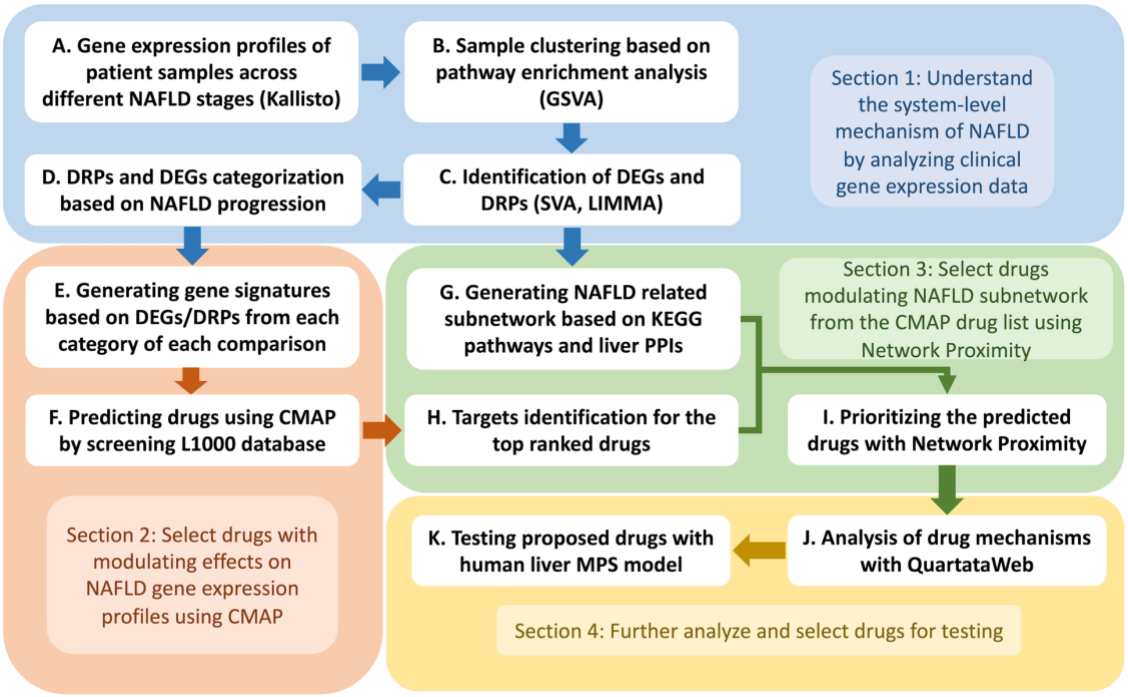


Figure 3.26 Overview of workflow used to predict drugs for NAFLD via clinical gene expression profiles

Section 1 shows the steps (A-D) that were used to understand the system-level (genes, pathways) mechanism of NAFLD based on the clinical gene expression data of NAFLD cohort. (A) The approach starts with RNA-seq data derived from NAFLD patient liver samples from a representative cohort of the NAFLD spectrum. Raw gene expression data was first pseudo-aligned to the Ensembl v94 (Zerbino, et al., 2018) human transcriptome via Kallisto (Bray, et al., 2016). (B) The gene expression levels for each patient were mapped to MSigDB v7.0 C2 KEGG (Liberzon, et al., 2011) pathways using gene set variation analysis (GSVA). The resulting patient x pathway matrix was clustered using hierarchal clustering, and three clusters were created by cutting the dendrogram at the 3rd level (see **Figure** 3.26 **and Table** 3.4). (C) Then, differentially expressed genes (DEGs) and differentially regulated pathways (DRPs) within each cluster comparison were identified. (D) These DRPs were categorized into different categories based on domain knowledge of KEGG pathways involved in NAFLD progression, termed NAFLD categories, DEGs mapped on each DRP were annotated with the corresponding category of that pathway, see **Appendix Appendix C.1**. In section 2, we used the results from section 1 to create representative gene signatures composed of genes which are responsible for driving disease progression, and predicted drugs with modulating effects on these gene signatures using CMap. (E) Using the results from step D, gene-sets of the up and down regulated DEGs

from four NAFLD related categories of three comparisons were generated as gene signatures, respectively, ending up with 12 gene signatures. (F) These gene signatures were used as inputs to perform CMap using the L1000 dataset respectively, top ranked drugs were considered as NAFLD gene expression potential modulators and were selected for further analysis (**Appendix Appendix C.3 and Appendix Appendix C.4**). In section 3, we used to rank and filter drugs modulating NAFLD subnetwork from the CMap prediction list (section 2) using an independent Network Proximity method, where the NAFLD network was created by using the data from section 1. (G) Liver PPI network were used as the background network, DEGs mapped on 11 pathways closely related to NAFLD were selected as the NAFLD sub-network. (H) targets of the drugs predicted by CMap were identified using drug-target interaction database, DrugBank. (I) Network Proximity were calculated to measure the relationship between targets of each drug in step H and the NAFLD subnetwork within the background PPI network in step G (**Appendix Appendix C.5**). The top ranked drugs were selected as potential drugs modulating NAFLD gene expression profiles, and closely targeting NAFLD subnetwork as well. Section 4, Steps J-K were used to further analyse the mechanisms of each predicted drug and select a final set of drugs for testing. J) The mechanisms of the predicted drugs and the drugs in clinical trials were analysed through a drug-target-pathway analysis tool, QuartataWeb, targets and pathways related to NAFLD were highlighted (**Appendix** Error! Reference source not found. and **Figure** 3.30). (K) The overall analysis leads to a proposed list of drugs that are potentially effective for NAFLD, and are planned to test with human liver MPS model.

3.3.2 Results

3.3.2.1 Patient clustering based on pathway variation is consistent with NAFLD clinical stages

Figure 3.27 shows the results of unsupervised clustering of KEGG pathway enrichment scores from the 182 patient samples across different stages of NAFLD including 36 normal, 46 steatosis, 50 lob inflammation and 50 fibrosis. The dendrogram was cut at the 3rd level, this resulted in 3 clusters that were each significantly enriched in one of the stages (**Figure** 3.27). The first cluster is composed of 43.3% normal patients and 48.1% patients with simple steatosis, termed

Normal & Steatosis (N&S), highlighting the challenge of distinguishing these two cohorts by gene expression analysis alone when inflammation is not present. The second cluster is predominated by patients with lobular inflammation but little or no fibrosis (70.3%), termed Predominately Lobular Inflammation (PLI). The third cluster is comprised of patients with advanced disease having fibrosis, termed Predominately Fibrosis (PF). The sample clustering is significantly associated (Pearson's Chi-squared Test) with NAFLD stage ($p < 2.2e-16$) and type 2 diabetes (T2D) status ($p = 0.01$). Details of the sample composition of each cluster could be found in **Table** 3.4. The clustering in **Figure** 3.27 also shows that the distribution of sex, body mass index (BMI) and age are similar across different clusters, while the occurrence of T2D in cluster PF (55%) is relatively higher than that of cluster N&S (32%) and PLI (32%).

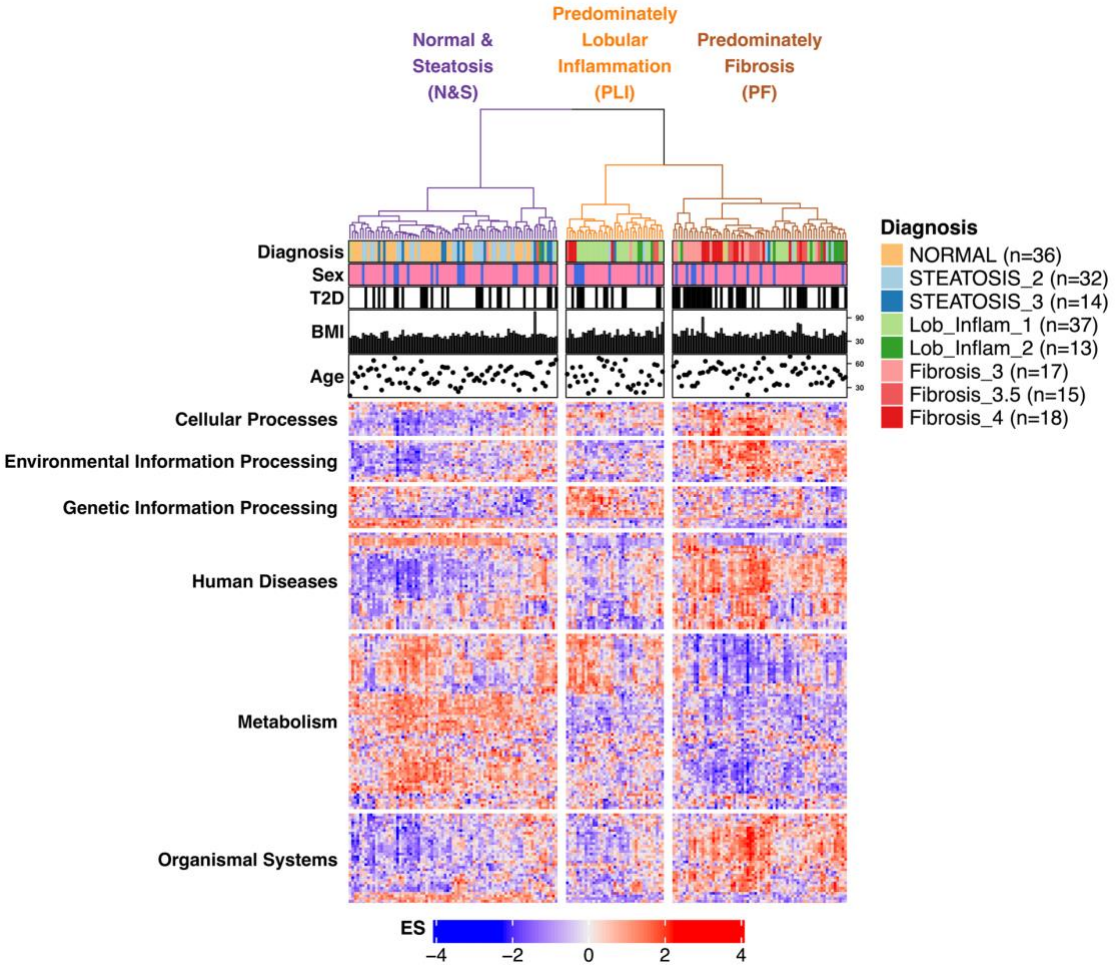


Figure 3.27 Unsupervised clustering of individual patients based on KEGG pathway enrichment scores gene set variation analysis (GSVA) broadly follows disease stage

The heatmap shows hierarchical clustering of individual samples based on the enrichment of MSigDB v7.0 C2 KEGG pathways, using the standard GSVA pipeline, except that the batch variables predicted by SVA were first removed from the gene expression matrix before the matrix was used as input for GSVA. The columns are individual samples, rows are KEGG pathways which are grouped according to the KEGG pathway groups. Both rows and columns have been clustered. The plots above the heatmap show the patient metadata: the top 2 bar indicates the color-coded diagnosis and patient sex, the third indicates (with a black tick mark) if the patient is diagnosed with T2D, and the last 2 plots show the BMI and age of the patient. Each entry of the heatmap represents the enrichment score for a particular pathway based on the gene expression profile of an individual patient. The 3 column clusters are named according to the predominate patient classification in each cluster: the first is almost entirely normal & steatosis (N&S) patients, the second is predominately lobular inflammation (PLI), and the third is predominately Fibrosis (PF), details of sample

proportion in each cluster is shown in Table 3.4. Comparing PLI vs. N&S, PF vs. N&S, and PF vs. PLI, yields a total of 59, 125, 50 DRPs (adjusted p-value < 0.001), respectively.

Table 3.4 Distribution of 182 patient samples in each patient cluster based on GSVA clustering

Stage	Normal	Steatosis		Lobular Inflammation		Fibrosis			T2D
Cluster		Stage 2	Stage 3	Stage 1	Stage 2	Stage 3	Stage 3.5	Stage 4	
N&S	35 (44.3%)	26 (32.9%)	12 (15.2%)	3 (3.8%)	1 (1.3%)	1 (1.3%)	1 (1.3%)	0 (0.0%)	25 (31.6%)
PLI	0 (0.0%)	2 (5.4%)	1 (2.7%)	23 (62.2%)	3 (8.1%)	1 (2.7%)	3 (8.1%)	4 (10.8%)	12 (32.4%)
PF	1 (1.5%)	4 (6.1%)	1 (1.5%)	11 (16.7%)	9 (13.6%)	15 (22.7%)	11 (16.7%)	14 (21.2%)	36 (54.5%)
total	36 (19.8%)	32 (17.6%)	14 (7.7%)	37 (20.3%)	13 (7.1%)	17 (9.3%)	15 (8.2%)	18 (9.9%)	73 (40.1%)

Table showing the percentage of each stage and counts for each of the NAFLD patient categories shown in **Figure** 3.27. The last column shows the number of patients with T2D.

3.3.2.2 The differentially regulated pathways (DRPs) identified among different patient clusters reveals the major pathways implicated in NAFLD progression

The comparisons of PLI vs. N&S, PF vs. N&S and PF vs. PLI yielded a total of 139 DRPs (FDR < 0.001), including 45 (32%) metabolism pathways, 31 (22%) human disease pathways, 28 (20%) organismal systems pathways, 14 (10%) environmental information processing pathways, 11 (8%) genetic information pathways and 10 (7%) cellular process pathways, as shown in **Figure** 3.28**A**. Overall, this set of pathways is consistent with the intrinsic heterogeneity of NAFLD that reflects the diverse but convergent impacts of the environment, metabolism, comorbidities, and genetic risk factors (Sanyal, 2019). More specifically, many of these DRPs can be grouped into at least one of four categories that comprise our current conceptual framework of NAFLD progression (**Figure** 3.28**B-C**): C1. Insulin resistance and oxidative stress, C2. Cell stress,

apoptosis and lipotoxicity, C3. Inflammation, C4. Fibrosis, as well as C5. Disease related pathways, C6. Other associated pathways that relate to comorbidities such as cardiovascular disease and cancer. Finally, a seventh category (C7) is comprised of three DRPs with no clear association to NAFLD or the metabolic syndrome. The detailed pathway description and categorization can be found in **Appendix Appendix C.1**.

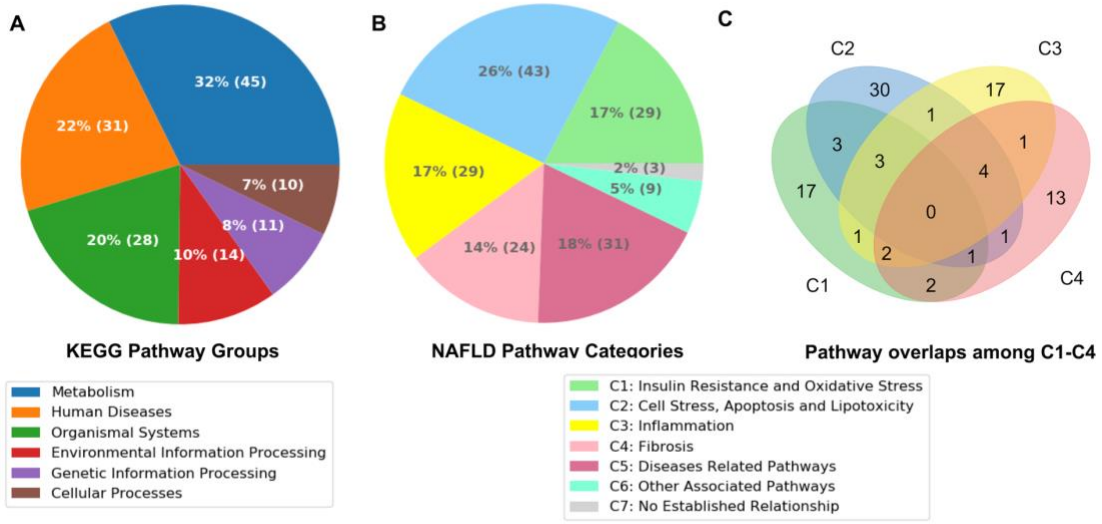


Figure 3.28 Distribution of the enriched KEGG pathways in KEGG pathway groups and NAFLD pathway categories

(A) Number of enriched pathways (FDR < 0.001) identified between the PLI vs N&S and PF vs N&S in 6 KEGG pathway groups. The details of the pathway groups and subgroups for each comparison can be found in Appendix Appendix C.1. (B) Number of enriched pathways categorized according to the NAFLD disease progression. Details of the pathways in each category are in Appendix Appendix C.2. (C) Number of pathways overlapped among categories C1-C4.

Insulin resistance and oxidative stress. Insulin resistance plays a central role in the pathophysiology of NAFLD, leads to the concept that NAFLD represents the hepatic expression

of the metabolic syndrome. As the development of NAFL and further into NASH, mitochondria injury, increased cycling of the cytochrome P450 system, and changes in peroxisomal function potentially drive the hepatic oxidative stress. A total number of 29 DRPs were categorized in this category. For example, regulation of actin cytoskeleton pathway is associated with oxidative stress through Keap1-Nrf2-ARE pathway, actin cytoskeleton helps localize the Nrf2 in the cytoplasm, while Nrf2 is activated by hepatic oxidants, protecting cell from oxidative stress (Chambel, et al., 2015).

Cell stress, apoptosis and lipotoxicity. In the patients with NASH, excess lipids increase and induce apoptosis, which leads to cell injury and death. Multiple cellular process and signaling pathways (e.g. apoptosis, cell cycle, MAPK signaling pathways), lipid metabolic pathways (e.g. fatty acid degradation), vitamins and xenobiotics metabolism pathways (e.g. one carbon pool by folate, metabolism of xenobiotics by cytochrome p450) are involved in this process. A total number of 43 DRPs were categorized in this category.

Inflammation. As NASH progresses, an activated innate immune system and increased inflammation response are observed. Immune pathways such as antigen processing and presentation, B cell receptor signaling pathways, Fc epsilon RI signaling pathway are involved in this process. Besides that, melanogenesis is hypothesized to abate oxidative and inflammation in adipose tissue (Page, et al., 2011). In the renin-angiotensin system, experimental and clinical findings show that Angiotensin-(1-7) by binding to Mas receptor opposes Angiotensin II actions mediated by AT1 receptors in liver tissue, by eliciting anti-inflammatory, anti-oxidative and anti-fibrotic effects (Simoes, et al., 2017). Experimental studies show the crosstalk between renin-angiotensin system and insulin signaling, resulting in the worsening of insulin resistance (Paschos and Tziomalos, 2012). A total number of 29 DRPs were categorized in this category.

Fibrosis. Fibrosis is widely recognized as the hallmark of disease progression in NASH. Driver pathways of fibrogenesis such as TGF-beta signaling (Feaver, et al., 2016), Hedgehog signaling (Syn, et al., 2009) are identified. Other cellular and signaling pathways including p53 signaling pathway, regulation of actin cytoskeleton and gap junction are demonstrated to be associated with fibrosis. Multiple amino acid metabolism (Gaggini, et al., 2018) and glycan metabolism pathways (Rostami and Parsian, 2013) are also identified and might play a role in the fibrosis process. A total number of 24 DRPs were categorized in this category.

Disease related pathways. There are 31 DRPs are associated with human diseases, including cancer, cardiovascular diseases, immune diseases, neurodegenerative disease, infectious diseases, endocrine and metabolic diseases. These diseases might be the complications of NAFLD, such as cardiovascular diseases or immune diseases.

Other associated pathways. There are 9 pathways in this category that might be associated with NAFLD, including pathways in circulatory systems, excretory system, nervous system, and sensory system. These pathways might contribute to the causal relationship between NAFLD and other systems, for example, cardiovascular risk, where the exact mechanisms are not fully elucidated.

No established relationship. As far as we know, there is no evidence showing the relationship between NAFLD and the remaining 3 pathways, oocyte melosis, dorso-ventral axis formation and progesterone-mediated oocyte maturation pathway.

Taken together, the pathway enrichment analysis reveals four major categories (C1-C4) with 96 pathways that regulate the different stages of NAFLD progression. We then demonstrated the links between the these DRPs and NAFLD progression by analyzing the top 10 enriched DRPs for each comparison.

As shown in **Figure** 3.29**A**, the top 10 most enriched DRPs for both the PF vs. N&S and the PLI vs. N&S comparisons are consistent with the metabolic underpinning, and resultant cellular stress and inflammatory response intrinsic to NAFLD pathogenesis. In these two comparisons, hepatic fructose uptake and metabolism would support *de novo* lipogenesis (DNL) that is largely unregulated providing a major source of excess free fatty acids in patients with NAFLD (Softic, et al., 2016). Mitochondrial and peroxisomal beta-oxidation of this excess of fatty acids leads to production of reactive oxygen species (ROS) (Aon, et al., 2014) that exceeds the regulatory capacity of the Keap1-Nrf-ARE pathway (Chambel, et al., 2015) resulting in oxidation of NADPH derived from the fructose dependent pentose phosphate pathway (Jin, et al., 2018). Excess free fatty acids also lead to ceramides through glycosphingolipid biosynthesis and these toxic lipids in conjunction with ROS directly down-modulate insulin receptor signaling to promote insulin resistance (Rostami and Parsian, 2013). A hallmark of the latter is enhanced gluconeogenesis evidenced by dysregulated glyoxylate and dicarboxylate metabolism (Kanehisa, et al., 2017). Excess lipids also promote endoplasmic reticulum (ER) stress and the unfolded protein response (UPR) in patients with NASH (Maiers and Malhi, 2019) consistent with the prominent DRP, ubiquitin-mediated proteolysis, for these two comparisons (Luo, et al., 2018). The UPR is now known to be a critical link between cell stress, inflammation, apoptosis (Maiers and Malhi, 2019) contributing to the perturbation of cell-cell communication involving gap junction dysregulation mediating NAFLD progression (Hernandez-Guerra, et al., 2019).

Not surprisingly, Complementary to the 10 most enriched DRPs in each of the PF vs. N&S and PLI vs. N&S comparisons, the comparison between PLI and PF is consistent with fibrosis being the widely recognized hallmark of disease progression in NASH (**Figure** 3.29**A**). Each of

the 10 DRPs in this latter comparison have been shown to have a role in NASH-mediated hepatic fibrosis with several involved in stellate cell activation.

Besides that, there are 4 pathways (**Figure** 3.29**B**) shared by all three comparisons, including gap junction, wnt signaling pathway, RNA degradation, amino sugar and nucleotide sugar metabolism, which also show up in the top 10 pathways list of at least one of the three comparisons (marked with a red star in **Figure** 3.29**A**). The dysfunction of gap junctions affects a wide variety of liver processes, such as differentiation, cell death, inflammation and fibrosis, and there exit drugs that modulate gap junction, it can be an attractive target pathway for NAFLD (Hernandez-Guerra, et al., 2019). Wnt signaling inhibits the adipocyte differentiation, its impaired function may trigger lipotoxicity (Gunaratnam, et al., 2014). It is also demonstrated to play the major role in liver fibrosis and inflammation in mice models (Wang, et al., 2015). Studies have shown that various regulatory non-coding RNAs play essential role in hepatic lipid regulation, inflammation and fibrosis (Sulaiman, et al., 2019). Amino sugar and nucleotide sugar metabolism links to many carbohydrate metabolisms including fructose, glucose, glycan et el. These are major mediators of insulin resistance, oxidative stress, lipotoxicity and inflammation in NAFLD process (Jensen, et al., 2018).

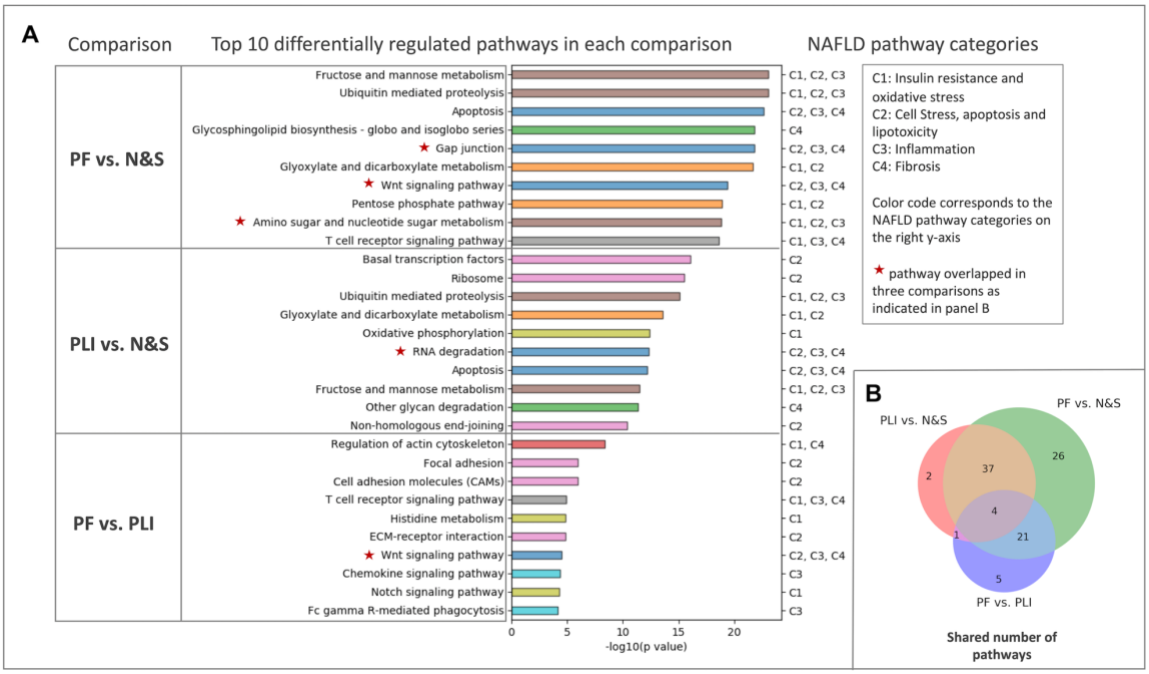


Figure 3.29 Top 10 differentially regulated pathways (A) and overlapping pathways (B) among three comparisons

(A). The top differentially regulated pathways ranked by the FDR adjusted p values through the linear modelling equivalent of a two sample, moderated t-test of each comparison (only pathways in NAFLD category 1-4 were included). X-axis represent the –log10(p value) of each pathway, y-axis on the left are pathway names, y-axis on the right are NAFLD pathway category labels of the corresponding pathway. The color of each bar codes the pathway categories of each pathway. (B). The Venn diagram of the overlapping differentially regulated pathways among three comparisons (only pathways in NAFLD category 1-4 are included).

Together, the analysis of this transcriptomic data set appears to have corroborated the clinical relevance of these DRPs in the context of NAFLD and the conceptual framework for its progression. Details of the full list of DRPs for each comparison can be found in **Appendix Appendix C.2**. Although each of these identified DRPs has the potential to be a drug target, their large number and diversity, the prospect of redundancy and the uncertainty regarding their individual contribution to NAFLD pathogenesis especially across a heterogeneous patient

population, all present challenges to translating the information into therapeutic strategies. To help meet this overarching challenge we hypothesize that differentially expressed gene signatures for each of the 4 NAFLD categories will reflect disease-specific networks for different stages of disease progression. These disease-specific networks will model how individual category-specific DRPs contribute and communicate to form emergent hubs that can be pharmacologically modulated. We have tested this hypothesis using the following approach.

3.3.2.3 Drug predictions via connectivity map (CMap)

In order to predict drugs that modulate individual components of NAFLD progression. We classified the DEGs that mapped on to the categorized DRPs identified in the three comparisons. For each of these comparisons, a DEG signature resulted for each NAFLD progression category C1. insulin resistance and oxidative stress, C2. cell stress, apoptosis and lipotoxicity, C3. inflammation, C4. Fibrosis, generating a total of 12 gene signatures. Each of these 12 gene signatures was then used as input to query CMap (Lamb, et al., 2006; Subramanian, et al., 2017). CMap connects the differentially expressed gene signature between two disease states to drugs and other pharmacologically active compounds that can normalize the gene signature. In the context of this study, the output of CMap enables the pharmacologic testing of the hypothesis that normalization of the gene signatures between two disease states will halt or even reverse disease progression in a NAFLD model (see below). The output connectivity score ranges from -0.91 to 0.90, representing respectively the inverse to the most similar gene signature produced by the corresponding pharmacologic agent in comparison to the input signature. Since our objective is to identify drugs that can be repurposed for preventing NAFLD progression, we focused on CMap outputs present in DrugBank that could promote the inversion of the disease-associated gene signature in each NAFLD category. The top 10 ranked drugs for each of the 12 queries were

selected, resulting in 70 unique predicted drugs, 38 of which appeared as an output in more than one query (see **Appendix Appendix C.3** and **Appendix Appendix C.4**).

For PLI vs. N&S, the top-ranking drug for C1 is yohimbine and respectively, diethylcarbamazine for C2, dorsomorphin for C3, and guanfacine for C4. Yohimbine has 12 annotated targets, that include adrenoceptor A 2A-C (ADRA2A-C), dopamine receptor D2 & 3, (DRD2-2), 5-hydroxytryptamine receptors 1 & 2 (HTR1A-B & D, HTR2A-C), and the potassium inwardly rectifying channel family member KCNJ1. ADRA2A has been associated with both alcoholic and nonalcoholic liver disease (Jia, et al., 2018), as well as liver fibrosis in animal models down-modulating hepatic stellate cell activation (Schwinghammer, et al., 2020). HTR2A has been shown to contribute to steatosis in mice (Choi, et al., 2018). Upregulation of DRD2 was previously shown in NAFLD patients (Mehta, et al., 2014). Both DRD2 and HTR2A-C are in the KEGG gap junction pathway. Diethylcarbamazine targets arachidonate 5-Lipoxygenase (ALOX5) and prostaglandin-endoperoxide synthase 1 (PTGS1). ALOX5 is in the KEGG Fc epsilon RI signaling pathway. Animal studies have shown that ALOX5 (Ma, et al., 2017; Martinez-Clemente, et al., 2010) plays a role in steatosis induced inflammation. PTGS1 has likewise been implicated in inflammation (Henkel, et al., 2018). Guanfacine targets ADRA2A-B, which is discussed above.

For **PF vs. N&S,** the top drug using the gene sets from C1-C3 is dorsomorphin. This drug targets activin receptor type-1 (ACVR1) and peptidyl-prolyl cis-trans isomerase (FKBP1A). Both of these genes have previously been implicated in NAFLD, through their roles in BMP signaling (Herrera, et al., 2017). Also, ACVR1 activation in Kupffer cells promotes a pro-inflammatory phenotype (Kiagiadaki, et al., 2018). ACVR1 is in the KEGG TGF-beta signaling pathway and cytokine-cytokine receptor interaction pathways. The top drug from C4 is amonafide, which

targets DNA Topoisomerase II A & Beta (TOP2A, TOP2B). A meta-analysis had found that TOP2A expression was positively correlated with NAFLD (Ryaboshapkina and Hammar, 2017).

For **PF vs. PLI**, the top drug using the gene sets from C1-C4 is gliquidone, which is used to treat T2D. It targets ATP binding cassette subfamily C member 8 (ABCC8) and potassium inwardly rectifying channel subfamily J member 8 (KCNJ8). ABCC8 is a member of ABC transporters KEGG pathway. Variants of these genes are associated with T2D (Gloyn, et al., 2003). They are also thought to play a role in NAFLD, specifically connecting the metabolic and liver disease phenotypes (Blackett and Sanghera, 2013).

Overall, the top predicted drugs tend to have targets that have associations with NAFLD. These are both mechanistic associations demonstrated through animal models, and observations from patients.

3.3.2.4 Predicted drug prioritization using network proximity analysis

To prioritize the list of 70 drugs from CMap we constructed a NAFLD subnetwork and used proximity to this network (Guney, et al., 2016) to enhance the specificity and relevance of the CMap analysis for NAFLD. In essence this algorithm connects NAFLD-associated gene signatures to drug-target profiles and maps the targets of a particular drug to the network protein nodes. Drugs with target profiles that most closely overlap with a subset of protein nodes in the NAFLD network are prioritized for further analysis (see below).

Construction of NAFLD associated protein-protein interaction network. The current conceptual framework of NAFLD involves diverse but convergent pathways. The KEGG pathway DB annotated a map showing the stage-dependent progression of NAFLD. In the first stage, lipid accumulation caused by insulin resistance and suppression of free fatty acids (FAAs) disposal. In addition, two transcription factors, SREBP-1c and PPAR-α, activate key enzymes of lipogenesis

and increase the synthesis of FAAs in liver. In the second stage, as a consequence of the progression to NASH, the production of reactive oxygen species (ROS) is enhanced due to oxidation. The lipid peroxidation can further cause the production of cytokines, promoting cell death, inflammation and fibrosis. The activation of JNK, which is induced by ER stress, TNF-α and FAAs, is also associated with NAFLD progression. Increased JNK promotes cytokine production and initiation of HCC. We used the NAFLD pathway as the main pathway, then expended to its 10 neighboring pathways that are known to crosstalk with the main NAFLD pathway, ending up with a NAFLD network comprising 11 associated pathways. The 10 neighboring pathways were TNF signaling pathway, insulin signaling pathway, type II diabetes mellitus, PI3K-Akt signaling pathway, adipocytokine signaling pathway, PPAR signaling pathway, fatty acid biosynthesis, protein processing in endoplasmic reticulum, oxidative phosphorylation and apoptosis. Among these pathways, apoptosis appeared in the top 10 pathways of both PF vs. N&S and PLI vs. N&S, oxidative phosphorylation appeared in the top 10 pathways of PLI vs. N&S (**Figure** 3.29**A**).

Given the total number of 2209 DEGs in our three comparisons PLI vs. N&S, PF vs. N&S and PF vs. PLI, 183 DEGs mapped to these 11 NAFLD associated pathways. We then mapped these 183 DEGs on the liver protein-protein interactome (Marinka Zitnik and Leskovec, 2018), resulting in a subnetwork with 104 protein nodes and 308 PPIs (see **Figure 4**). The degrees of the subnetwork nodes range from 1 to 17, with 5.92 neighbors on average. The top 10 hub proteins were mitogen-activated protein kinase 8 (MAPK8), NF-kappa-B essential modulator (IKBKG), mitogen-activated protein kinase 3 (MAPK3), protein kinase C α (PRKCA), caspase 8 (CASP8), signal transducer and activator of transcription 3 (STAT3), mitogen-activated protein kinase kinase kinase 7 (MAP3K7), 14-3-3 protein gamma (YWHAG) and protein kinase C zeta type

(PRKCZ). MAPK8 and IKBKG were two proteins with the most interaction partners (degree = 17) in our subnetwork. MAPK8 is a member of the MAP kinase and JNK family, acting as an integration point for multiple biochemicals signals, it involves in 7 of the 11 NAFLD associated pathways including NAFLD main pathway, TNF signaling pathway, insulin signaling pathway, Type II diabetes mellitus, adipocytokine signaling pathway, protein processing in endoplasmic reticulum and apoptosis. IKBKG is a regulatory subunit of the IKK core complex which phosphorylates inhibitors of NF-kappa-B thus leading to the dissociation of the inhibitor/NF-kappa-B complex and ultimately the degradation of the inhibitor. IKBKG is involved in 4 of the 11 NAFLD associated pathways including TNF signaling pathway, PI3K-Akt signaling pathway, adipocytokine signaling pathway and apoptosis.

Predicted drugs prioritized by network proximity. The NAFLD PPI subnetwork constructed in the previous section were considered as the disease module, we prioritized 49 of the 70 drugs predicted by CMap by evaluating the significance of the distance between their targets and our NAFLD disease module in the background liver PPI interactome, using the network proximity measure proposed by (Guney, et al., 2016). The remaining 21 drugs do not have annotated targets in the liver PPI interactome, therefore were not involved in the network proximity analysis. The network proximity measure for each drug is represented by a z-score ranging from -3.60 to 1.89, negative z-score means the targets of the drug is closer to the disease module than a random set of targets. Therefore, the lower the z-score of a predicted drug the more likely it is to modulate the signaling in our NAFLD disease module (see **Appendix Appendix C.5**).

3.3.2.5 Comparison of the predicted drugs and NAFLD clinical trial drugs

In order to further understand the mode of action of our predicted drugs and make a rational list of drugs for experimental testing, we examined the targets and pathways of all 49 predicted

drugs, as well as 84 drugs under clinical trials (detailed information of these drugs are listed in **Appendix Appendix C.7**) for NAFLD or NASH from DrugBank. Eight out of those 84 drugs including exenatide, emricasan, pradigastat, niacin, selonsertib, atorvastatin, pentoxifylline and simtuzumab, show positive result at a certain clinical trial phase. The results of the remaining 76 drugs are either not submitted or negative.

According to the drug-target-mapping using QuartataWeb, we mapped 23 predicted drugs and 20 clinical trial drugs that directly link to the 11 NAFLD associated pathways through their targets, detailed relationship among the drugs, targets and pathways are shown in **Appendix** Error! Reference source not found.. Eleven of the 20 clinical trial drugs affect the NAFLD main pathway, while only 5 of the 23 predicted drugs affect the NAFLD main pathway, suggesting the diversity of our predicted drugs compared with the clinical trial drugs. Emricasan affects the NAFLD main pathway, TNF signaling pathway and apoptosis by targeting two caspases CASP3 and CASP7.

3.3.2.6 Selection of the final list of drugs to test in Liver MPS

The focus of the present study has been the construction of a computational platform pipeline to anchor the prediction of drugs for potential repurposing. There are multiple approaches to ranking drugs out of this pipeline. Here, we propose the most promising drugs for the initial experimental testing to demonstrate a proof of concept. We selected 19 drugs from the 49 calculated drugs with network proximity Z-score below zero (statistically targeting the disease module), then excluded the ones that have not been approved or with serious hepatotoxicity side effects, ending up with 12 drugs. We plan to initially test a few drugs on one of the multiple phenotypic characteristics of NAFLD, steatosis. 8 drugs targeting steatosis-related pathways according to the KEGG pathway map were selected as the first experimental set. The steatosis-related targets and pathways of these drugs are shown in **Figure** 3.30. There are six steatosis-

related pathways and seven targets targeted by these 8 drugs. Specifically, everolimus is a mTOR inhibitor, which plays a role in PI3K-Akt signaling pathway, Type II diabetes mellitus, and insulin signaling pathway. Celecoxib interact with 3-Phosphoinositide dependent protein kinase 1 (PDPK1), which is involved in PI3K-Akt signaling pathway, insulin signaling pathway, TNF signaling pathway, PPAR signaling pathway. Promazine, dosulepin and ziprasidone interact with cholinergic receptor muscarinic 1 and 2 (CHRM1, CHRM2), which are involved in PI3K-Akt signaling pathway. Quinapril interacts with angiotensin converting enzyme (ACE), which plays a role in renin-angiotensin system. Isradipine blocks calcium voltage-gated channels (CACNA1C, CACNA1D), which are important components of the type II diabetes mellitus pathway. Curcumin interacts with peroxisome proliferator activated receptor gamma (PPARG), which modulates the PPAR signaling pathway. Taken together, the identification of these targets and pathways indicates the potential modulating effects of the 8 drugs on steatosis, however, these might not be the only mechanisms because other drug-target interactions might exist but not annotated in DrugBank yet, and potential side effects should also be taken into consideration. Therefore, we listed the indication, MOA, adverse effects and literature support of each drug in **Appendix Appendix C.6**, which help us to manually select the most promising drugs to test. On the other hand, the other drugs that don’t target the analyzed steatosis-related pathways might also be interesting to explore in future studies, they might work either through unknown drug-target interactions or novel steatosis mechanisms.

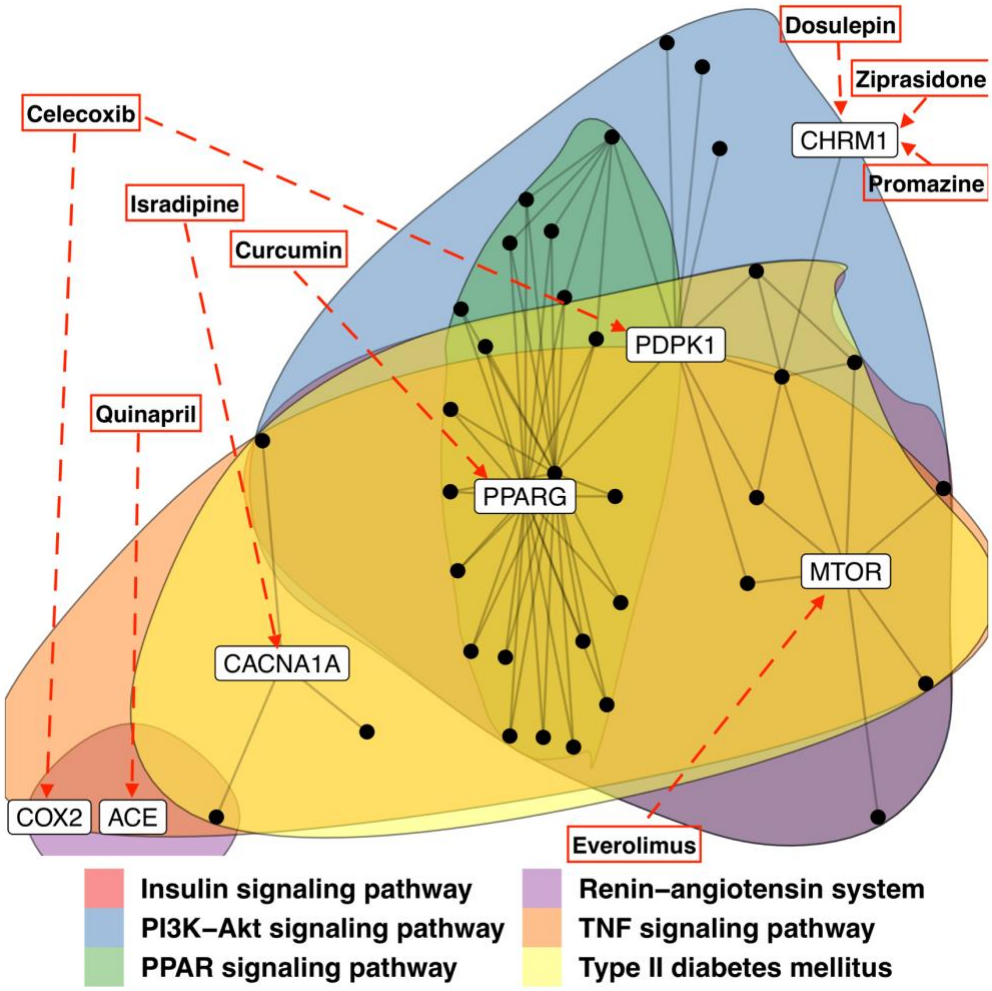


Figure 3.30 Subnetwork of the steatosis-related targets and pathways of the 8 prioritized drugs

The black nodes represent genes and the edges are associations between them as indicated in KEGG pathways. The shaded regions covering the nodes show the corresponding KEGG pathways. The labelled gene nodes represent the targets of the prioritized drugs, they are angiotensin I converting enzyme (ACE), cholinergic receptor muscarinic 1 (CHRM1), calcium voltage-gated channel subunit α1 A (CACNA1A), prostaglandin-endoperoxide synthase 2 (COX2), 3-phosphoinositide dependent protein kinase 1 (PDPK1), peroxisome proliferator activated receptor gamma (PPARG), and mechanistic target of rapamycin kinase (mTOR) (listed in red boxes).

3.3.3 Discussion

In this study, we carried out a data-driven, unbiased and efficient QSP approach that focused on identifying potential drugs that could be repurposed for a NAFLD indication. The overall analysis was driven by the RNA-seq data from a representative cohort of the NAFLD patients and a NAFLD associated PPI network derived from KEGG pathway map, toward gaining a system-level understanding of the key players involved in the steatosis, inflammation and fibrosis in NAFLD progression, leading to network-based drug repurposing. By constructing 12 gene signatures from 4 pre-defined NAFLD progression associated categories and 3 NAFLD indications, we obtained 70 drugs that have the potential to reverse gene signatures. By constructing a NAFLD associated PPI network and utilizing network proximity to further evaluated the drug effects of these drugs to the NAFLD disease module, leading to a rationale of selecting 2 potential steatosis drugs to test in the fully validated, human liver acinus MPS (LAMPS) model that can readily quantify steatosis and secrotome content. However, the computational pipeline developed here will be used in the future to explore a large range of drug selection approaches. The human vascularized liver microphysiology system (vLAMPS) has been developed to explore complex liver diseases such as NAFLD and the Metabolic Syndrome where many read-outs are required to test the impact of drugs on parameters such as steatosis, oxidative stress, secretome contents, immune cell infiltration, fibrosis, insulin resistance, etc. Future drug testing will be performed on the vLAMPS characterizing the impact of each drug on all of the parameters, not just steatosis.

3.3.3.1 QSP approach complemented the target-centric drug design for NAFLD

NAFLD is a highly prevalent disease and important unmet medical needs (Younossi, et al., 2018). Due to the intrinsic heterogeneity in NAFLD progression and the complexity of underlie molecular networks, the therapeutic benefits of many existing drugs under clinical trials for NAFLD remain to be proven (Polyzos, et al., 2020). A significant amount of research focus on designing target-centric drugs, including obeticholic acid (a farnesoid X receptor agonist) (Mudaliar, et al., 2013; Neuschwander-Tetri, et al., 2015; Pockros, et al., 2019; Younossi, et al., 2019), elafibronor (a peroxisome proliferator activated receptor [PPAR]-α/δ dual agonist) (Ratziu, et al., 2016), cenicriviroc (a CC chemokine receptor antagonist) (Friedman, et al., 2018), selonsertib (an apoptosis signal-regulating kinase-1 inhibitor) (Loomba, et al., 2018) and resmetirom (a thyroid hormone receptor agonist) (Harrison, et al., 2019). Among which, obeticholic acid has been rejected for fibrosis due to nonalcoholic steatohepatitis (NASH) by FDA recently. Therefore, there's an urgent need to seek new strategies and advance the treatment. On the other hand, as the accumulation of more and more multi-scale data and the development of computational and systems biology techniques, designing drugs that targeting the multiple key driver pathways or networks for NAFLD rather than a specific target becomes achievable and might be more efficient. In this work, we took the advantage of an unbiased, data-driven QSP approach for this aim.

The patient cohort we are using contains 182 patients distributed in normal, steatosis, lobular inflammation and fibrosis stages, is the largest and most diverse data sample analyzed in similar studies to our knowledge, which enable us to perform unbiased data analysis. Our analysis shows that the patients' clinical stages are consistent with their gene profile clusters to some extent, indicating the treatment for early stage might be different from later stages. Therefore, in the next

step, we identified DEGs and DRPs in three different comparisons, corresponding to different stage comparisons. Given the large amount of DEGs and DRPs identified in each comparison, we categorized them in different groups according to the current knowledge of NAFLD progression, in order to predict drugs focusing on a specific phenotype, as well as the ones work for multiple phenotypes. This categorization could be changed if more insights or new relationships between pathways and NAFLD gained in the future. Network proximity helps us to adjust the potential error caused by CMap prediction alone, and further filter drugs. The overall workflow is rationale to identify system-level drugs half or reverse the NAFLD progression, which complement the target-centric strategies. New discoveries of the NAFLD mechanisms and further experimental testing are required to adjust the process in the next QSP iteration.

3.3.3.2 Evidences support the potential effect of our proposed drugs

Due to the loss of experimental time during the COVID-19 shut-down, we have focused on the computational platform with just proof of concept on testing the modulating effects of our proposed drugs just for steatosis. Therefore, we selected the most promising predicted drugs for steatosis based on evidences collected from CMap prediction, network prioritizing, targets on steatosis-related pathways and literature support of its relationship with NAFLD.

The first one is Everolimus, which is an FDA approved drug used as an immunosuppresent to prevent rejection of organ transplants and in the treatment of renal cell cancer and other tumors. In our analysis, it is predicted by CMap C1 (rank 1st), C3 (rank 1st) and C4 (rank 2nd) with relatively high ranking (see **Appendix Appendix C.3**). The network proximity analysis also gives very promising result with a ranking of 5th (see **Appendix Appendix C.5**). Specifically, its target mTOR plays an essential role in three NAFLD steatosis-related pathways: PI3K-Akt signaling pathway, type II diabetes mellitus and insulin signaling pathway (see **Appendix Appendix C.6**).

Furthermore, it has been reported that Everolimus inhibited hepatic lipid accumulation and improved metabolic parameters in a fast food induced mice model of NASH, even though the inflammatory and fibrotic responses still exhibited despite the reduced hepatic steatosis (Love, et al., 2017). Therefore, it would be of great interest to test its effect on steatosis in our liver MPS model, and also investigate its potential in reducing the inflammatory and fibrotic responses in human models in future studies.

The second proposed drug is celecoxib, an FDA approved nonsteroidal anti-inflammatory drug indicated for pain relieve caused by osteoarthritis, rheumatoid arthritis and ankylosing spondylitis. In our analysis, it is predicted by CMap C2 (rank 5th) (see **Appendix Appendix C.3**) with a network proximity ranking of 19th (see **Appendix Appendix C.5**). Celecoxib targets COX-2 enzyme, evidence shows that it partially restores autophagic flux via downregulation of COX-2 and alleviates steatosis in vitro and in vivo (Liu, et al., 2018). In addition, a non-selective COX inhibitor aspirin was shown to be protective of NAFLD progression in a retrospective study (Simon, et al., 2019), further support that inhibiting COX-2 might help celecoxib to alleviate steatosis. Besides its primary target, we also identified that it interacts with PDPK1, which plays a role in four steatosis-related pathways insulin signaling pathway, PI3K-Akt signaling pathway, TNF signaling pathway and PPAR signaling pathway. This might help explain why celecoxib, as an anti-inflammatory drug, shows up in the top list of C2 instead of C3 in CMap prediction. Studies also show that celecoxib attenuates liver steatosis and inflammation in NAFLD in a rat model (Chen, et al., 2011). Taken together, it is worthwhile to test celecoxib in our liver MPS model.

The other 6 drugs passed the filter of our computational pipeline but with weaker literature evidence can also be found in **Appendix Appendix C.5**, which can serve as references for future analysis. We have to point out that without well designed mechanistic experimental verification,

the mechanisms of how these drugs work in NAFLD network remain unclear, the predictions and supporting evidences only help us to generate potential hypotheses and reduce the amount of test.

3.3.3.3 Drug combinations have the potential to advance the NAFLD treatment

Depending on the underlying mechanisms of action, certain drugs might be more effective on steatosis, while the others might be more effective on inflammation or fibrosis. In our analysis, the top list of drugs predicted by CMap category C1 or C2 in might be more effective on steatosis, while the drugs predicted by category C3 might be more effective on inflammation, C4 might be effective on fibrosis. Given the multifactorial pathogenesis of NAFLD, the drugs that predicted by multiple categories, or combination usage of drugs from complementary categories may prove to be more effective and suitable for NAFLD in the long term. For a specific NAFLD stage or condition, combination strategies targeting different key driver pathways would also be interesting to explore in future studies. For example, our predicted drug quinapril targets the renin-angiotensin system, different from other major steatosis-related pathways targeted by the two proposed drugs. It is also reported that quinapril helps attenuate the progression of metabolic syndrome (Khan, et al., 2004). Therefore, quinapril can be a good candidate to combine with one of the proposed drugs for potential beneficial effects.

3.3.4 Materials and Methods

3.3.4.1 Generation of gene expression profiles

The RNA-seq data are derived from patient samples of wedge biopsies taken from livers of patients undergoing bariatric surgery, as described in Gerhard et al., 2018 (Gerhard, et al., 2018). Patients were diagnosed according to the predominant liver histology finding as normal, steatosis,

lobular inflammation, or fibrosis (Gerhard, et al., 2018) . Patients were further grouped into disease stages according liver histology (Gerhard, et al., 2018). RNA was extracted from the tissue samples and paired end libraries were prepared from polyA-selected RNA (Gerhard, et al., 2018). Sequencing was performed using an illumina HiSeq2000 (Gerhard, et al., 2018).

Figure 3.26**A** represents the data pre-processing, paired fastq-files were pseudoaligned to the human Ensembl (Frankish, et al., 2017) v94 transcriptome using Kallisto (Bray, et al., 2016) following the recommended pipeline. Estimated transcript abundances were then summarized into gene-level estimates using Tximport (Soneson, et al., 2015) with the settings recommended for LIMMA-VOOM (Law, et al., 2014; Ritchie, et al., 2015). Subsequent data analysis followed the standard LIMMA linear modelling approach (Law, et al., 2016; Smyth, 2004), with exceptions of using quantile normalization and surrogate variable analysis (SVA) (Leek, et al., 2012; Leek and Storey, 2007) to identify batch effects.

3.3.4.2 KEGG pathway analysis and identification of differentially expressed genes (DEGs) and differentially regulated pathways (DRPs)

Pathway analysis was performed on the resulting gene expression matrix obtained from the previous step (**Figure** 3.26**B**). The gene expression levels were first pre-processed so that the batch effects predicted from SVA were removed from the data using the *removeBatchEffect* function from LIMMA. Next the gene expression values for each patient were mapped to MSigDB v7.0 C2 KEGG (Liberzon, et al., 2011) pathways using gene set variation analysis (GSVA) (Hanzelmann, et al., 2013). The resulting sample x pathway matrix was clustered using hierarchal clustering, and new groups were created by cutting the column dendrogram at the 3rd level. New groups were defined based on the samples within each cluster.

DEGs were identified by first row scaling the gene expression data and then applying the standard LIMMA-VOOM pipeline (**Figure** 3.26**C**) (Law, et al., 2014; Ritchie, et al., 2015; Smyth, 2004). The results are from pairwise contrasts of the 3 new groups identified in the clustering step. DRPs were identified in the same way except that the output GSVA data was used instead of gene expression data. Each DRP was assigned as an integral component to one or more of 4 distinctly annotated categories known to be involved in NAFLD progression as described in the Results. The up- and down-regulated genes mapping to the DRPs in each category were then used to generate gene signatures. For each of the three pairwise comparisons among the three newly generated clusters (see above) 4 category-specific gene signatures were generated resulting in 12 total gene signatures. These signatures were then used to query an extensively annotated connectivity map to identify drugs with the potential to normalize each of these gene signatures (**Figure** 3.26**D**).

3.3.4.3 Drug predictions via CMap

Drugs were then predicted (**Figure** 3.26**F**) using the LINCS L1000 level 5 (GSE92742) expression DB (Subramanian, et al., 2017) that was downloaded from the University of Pittsburgh Center for Research Computing's HTC cluster. The DB was filtered to keep only small molecule perturbation instances. This yielded a total of 205,034 unique instances which consists of 20,413 compounds, 71 cell types, 6 & 24 hr time-points, and a range of concentrations. This DB consists of perturbation signatures (PS), which is a vector of continuous gene expression values for each perturbagen. We used our 12 gene signatures generated in the previous step as input to query CMap, the tool calculated the similarity of the query signature with each perturbation signature in the L1000 DB, and assigned a connectivity score (z-score) for each compound perturbation based on the adjusted enrichment statistic of the similarities. The compounds were ranked by their z-scores in an ascending order, where lower z-score represents better reverse effect of the input

signature. Compounds not in DrugBank were removed, and the top 10 ranked compounds for each input signature were selected for further analysis.

3.3.4.4 Drug prioritization via network proximity

The resulting drugs were further selected by network proximity (**Figure** 3.26**G**). The basic idea of network proximity is to evaluate the significance of the network distance between a drug and a given disease module. The methodology has been developed by Guney E et el., (Guney, et al., 2016) and based on the assumption that a drug is effective to a disease by targeting proteins within or in the immediate vicinity of the corresponding disease module. In essence this approach provides an independent criterion for increasing the specificity of the CMap analysis to enable drug prioritization for experimental testing.

Construction of NAFLD associated protein-protein interaction network. We constructed a NAFLD associated protein-protein interaction network by identifying the DEGs that take part in 11 NAFLD related pathways and the protein-protein interactions (PPIs) among these DEGs translated proteins in liver interactome. In details, we select the KEGG pathway map of NAFLD, which illustrates a stage-dependent progression of NAFLD, and is closely connected with 10 other pathways, including TNF signaling pathway, Insulin signaling pathway, Type II diabetes mellitus, PI3K-Akt signaling pathway, Adipocytokine signaling pathway, PPAR signaling pathway, Fatty acid biosynthesis, Protein processing in endoplasmic reticulum, oxidative phosphorylation and Apoptosis. Among the total number of 2209 DEGs, 183 DEGs were mapped on these 11 NAFLD associated pathways. We then mapped these 183 DEGs on the liver protein-protein interactome (Marinka Zitnik and Leskovec, 2018), resulting in a subnetwork with 104 protein nodes and 308 PPIs (as shown in Figure 4). This subnetwork served as the NAFLD

associated PPI network and the 104 proteins served as the NAFLD disease proteins in the network proximity calculation in the following drug prioritizing step.

Construction of the drug-target interaction network. We constructed the drug-target interaction (DTI) network on the 70 CMap predicted drugs by acquiring DTIs from DrugBank (v5.1.5) (Wishart, et al., 2018). A total number of 192 DTIs were identified between 51 drugs and 115 protein targets, with no annotated targets for the remaining 19 drugs.

The constructed NAFLD associated PPI network (see above) serves as the disease module containing 104 NAFLD disease proteins (S). For each drug, given a set of targets (T) from the constructed DTI network (see above), the closest distance measured by the average shortest distance path between nodes s and the nearest disease protein t in the human liver PPI interactome was calculated as:

$$\boldsymbol{d}(\boldsymbol{S},\boldsymbol{T}) = \frac{\mathbf{1}}{\|\boldsymbol{T}\|}\sum_{\boldsymbol{t}\in\boldsymbol{T}} \mathbf{min}_{\mathbf{s}\in\mathbf{S}}\, \boldsymbol{d}(\boldsymbol{s},\boldsymbol{t}) \qquad (3.3)$$

Then a reference distance distribution was constructed, corresponding to the expected distance between two randomly selected groups of proteins of the same size and degree distribution as the original disease proteins and drug targets in the network. This procedure was repeated 1000 times, the mean and standard deviation of the reference distance distribution were used to calculate a z-score by converting an observed distance to a normalized distance. After the calculation, each drug was assigned a z-score to evaluate its effects on NAFLD disease module, where lower z-score represents that the targets of the drug is closer to the disease module, namely more effective of the drug.

3.3.4.5 Identification of targets and pathways for predicted and clinical trial drugs

The targets and pathways from both the clinical and predicted drugs (**Figure** 3.26**I**) are evaluated using our in-house drug-target-pathway mapping tool QuartataWeb (Li, et al., 2020). This was used to map both the clinical and predicted drugs to targets and pathways, with the updated data sources: DrugBank (v5.1.5) and KEGG (updated on October 23, 2019). Drugs were mapped to targets based on the DTI annotations and the corresponding targets were mapped to pathways based on gene-pathway associations, leading to the drug-target-pathway relationships. We focused on the 11 NAFLD associated pathways defined in previous section and reported 23 predicted drugs and 20 clinical drugs that mapped on these pathways.

3.3.5 Acknowledgments

The content of this subsection is included in a publication in preparation. Daniel Lefever (Ph.D. student) and I contributed to the design, implementation and analysis of the project presented here and wrote the first draft of the developing manuscript. Drs. D. Lansing Taylor, Andrew Michael Stern and Albert H Gough contributed to the design and discussion of the methods and results. The project was supervised by Drs. D. Lansing Taylor, Bert Gough and Ivet Bahar.

4.0 Future Directions

Over the past decade, the aim of drug discovery has shifted from designing selective ligands for a specific target to understanding how drugs modulate cellular networks, in order to predict drug targets and their role in human pathophysiology, leading to a new paradigm of QSP (Perez-Nueno, 2015). New methods and tools have been developed to efficiently learn from drug-target interactions, drug-drug interactions, protein-protein interactions and pathway-gene associations so as to generate information and accelerate translational science. However, we are still far from achieving the goals of QSP. Huge efforts are needed for precise modeling of biological networks and more comprehensive frameworks are needed to combine computational and experimental QSP approaches. In this work, I focused on developing computational QSP methods and tools to explore/enable their applications in understanding disease mechanisms and developing novel therapeutic strategies for complex diseases. Our study demonstrated the power of machine learning (ML) models in facilitating the prediction of drug-target interactions and protein-protein interactions, and novel applications of QSP in understanding and discovering drugs for complex diseases including HD, drug abuse, and NAFLD. Below, I will briefly recap the conclusions reached from the analyses described in each chapter and discuss future directions of improvements or applications.

4.1 Future Development of QuartataWeb

In Chapter 1.0, we developed an integrated chemical-target-pathway mapping tool QuartataWeb for chemogenomics and polypharmacology analysis. Using QuartataWeb, users can retrieve data and generate information on experimentally verified and computational predicted drug-target interactions from DrugBank or chemical-protein interaction from STITCH; the targets are linked to KEGG pathways and GO annotations, which enable understanding the drug effects via targets and cellular pathways. It allows users to query a list of chemicals, drug combinations, or multiple targets.

The current version of QuartataWeb depends on four publicly available DBs: DrugBank, STITCH, KEGG and GOA. The ML methodology (and model parameters) should be updated when there are significant changes in new releases of these DBs. Besides that, many other popular drug/chemical-target interaction and pathway DBs based can be included based on our framework. The following DBs can be taken into consideration: SuperTarget (Hecker, et al., 2012), ZINC (Irwin, et al., 2012), TTD (Li, et al., 2018) and ChEMBL(Mendez, et al., 2019). SuperTarget provides comprehensive data services including 332,828 interactions between 6,219 proteins and 195,770 compounds. ZINC is a free and curated large collection of commercially available compounds. ChEMBL is a large bioactivity DB including 15,207,914 biological activities about 2,275,906 small molecules and 12,091 targets based on publications from several core Medicinal Chemistry journals.

Even though PMF predicts drug-target interaction with high efficiency and accuracy, it has difficulty predicting interactions involving new drugs or targets for which there are no known interactions. These are usually referred to as “cold-start” problem in recommendation systems. Since the CTI datasets are usually located at or near low-dimensional nonlinear manifolds,

advanced matrix factorization methods can be adopted to solve this problem in the future such as the variations of Graph Regularized Matrix Factorization (Cui, et al., 2019; Ezzat, et al., 2017; Mongia and Majumdar, 2020).

In terms of the cellular effects of chemicals, QuartataWeb first links chemicals to targets, then map their targets to pathways and GO annotations, thus the corresponding pathways and GO annotations can be used to infer the cellular effects of these chemicals. However, many chemicals are promiscuous, and their targets are involved in multiple pathways and networks, such that the actual impact of chemicals on cellular networks may be very complex. Pathways and GO annotations only partially disclose the underlying biological processes. Other biological processes such as protein-protein interactions and gene regulations may also play an important role in disease development or therapy. In the future, additional analyses of protein-protein interaction networks and network-based metrics might help further evaluate the cellular effects of chemicals or targets.

4.2 Future Development and Application of PPI Prediction

In Chapter 2.0, we adapted a symLMF-based methodology to predict large-scale PPIs, purely based on the existing PPI network, without dependence of any protein sequence or structure information. We showed that the proposed method can be efficiently applied in completing the entire interactomes, or recommending the most promising interaction partners of a certain protein. Actually, symLMF can be broadly applied to many settings including non-binary input. Future analysis could focus on applying symLMF to other PPI prediction task, e.g. estimating binding affinities. In PPI binding affinity predictions, the protein complex interface properties are essential to determine the binding affinity, and the impact of mutations should also be considered.

Therefore, individual binding sites instead of proteins should be considered as items in the symLMF model. As more and more high-quality binding affinity data would be available in the future, it would be interesting to evaluate if symLMF is able to capture the latent factors of different binding sites, enable more precise protein-protein binding predictions.

In addition, considering our big picture of developing integrated QSP tools, it will be valuable to develop a large-scale PPI prediction web-server using symLMF algorithm, and integrate it with the QuartataWeb described in Chapter 1.0. The integration of PPI networks would complement the pathway and GOA analyses of the cellular effects of any input drugs or targets of interest. In these analyses, context-specific (tissues, disease conditions et al.) PPI networks can provide valuable insights into key research questions like identifying disease mechanisms or effective drugs. Therefore, applying symLMF on context-specific PPI networks is suggested in future web server development. Besides, studies confirmed that interacting proteins tend to be located within the same compartment, or in physically adjacent compartments (Gandhi, et al., 2006). In practice, it is suggested to remove PPIs between two proteins not sharing any subcellular localizations not only in training datasets, but also in predictions.

4.3 Future Work of Three Applications of QSP

In Chapter 3.0, we demonstrated the applications of QSP methodology to three complex diseases/disorders: HD, drug abuse and NAFLD. QSP application can start from any components and customize specific workflows based on the available data sources and tools. In the first study, we started from the phenotypic screening of active compound probes and drug combinations against a well-established HD model, and identified important neuronal cell protection related

pathways through an initial chemogenomics analysis. In the future, it would be of interest to explore medium spiny neurons derived from human iPSC in the context of human neuronal MPS (Pamies, et al., 2017) that recapitulate critical cell intrinsic and extrinsic microenvironments. Further development of a comprehensive computational model of disease progression through the integration of the chemogenomic analysis and transcriptomic profiles of HD in both mouse and human tissues would be helpful, which will enable refinement of testable hypotheses. Furthermore, additional iterations of experimentally testing hypotheses and refining models should lead to emergent properties of HD disease and therapeutic strategy design.

In the second study of Chapter 3.0, we selected 50 representative drugs of abuse from 6 different categories and carried out a comprehensive analysis of the targets and pathways of these drugs. Our study identified key pathways at different stages of drug addiction cycle, as well as the cell signaling and regulation events associated with drug abuse. The results invite attention to new targets of addictive drugs and pathways implicated in the development of addiction, as well as new therapeutic opportunities, beyond those usually investigated by previous studies. The validation of our predictions requires comprehensive wet-lab bioactivity assays in the future. In particular, the establishment of the proposed role of mTORC1 would require *in vitro* and *in vivo* longitudinal studies given that our current study points to the involvement of mTORC1 at later stages of drug addiction. A similar combined computational-experimental framework as described in Section 3.1 could be adopted to extend the current study and establish new strategies, which would provide insights into the pleiotropy of the targets of addictive drugs as well as the common signaling platforms that may serve as mediators of drug addiction. In addition, knowledge of pathways implicated in drug addiction may be used, as a next step, to construct kinetic models to quantitatively assess the orchestration of signals induced by pathway crosstalk.

Furthermore, both target-centric and network-centric drug repurposing strategies can be carried out based on the exploring of addiction mechanisms. Known or newly verified targets of abused drugs can be queried in QuartataWeb (Chapter 1.0) to obtain their interacting drugs. Input Type III of the tool helps to obtained target-centric repurposable drug candidates, both known and predicted. Input Type I helps to identify polypharmalogical drugs that targeting multiple targets, or suggest drug combinations that targeting complementary key components of a network. The network effects of the drug can be further quantitatively modeled in the abovementioned kinetic models.

In the third study of Chapter 3.0, we implemented a QSP approach that started with the pathway enrichment analysis of RNA-seq data from a full spectrum of NAFLD patients, then proposed two most promising drugs to halt or reverse NAFLD progression by integrating CMap and Network Proximity methods. As a next step, we would test the proposed drugs in the liver acinus MPS model (LAMPS), and perform another iteration of drug repurposing with adjusted criteria based on the test results. In particular, there are three steps that could be adjusted in future studies: 1. Use an alternative cutoff value (top 10 in our current study) to select drugs predicted by CMap. 2. Include non-approved (experimental) drugs for testing if the results of the selected approved drugs turn out to be negative, even though non-approved drugs have more safety concerns, drugs under investigations for other indications can be promising for NAFLD. 3. We plan to initially test a few drugs on one of the multiple phenotypic characteristics of NAFLD, steatosis, so drugs targeting steatosis-related pathways were prioritized. In future studies, the pathways and networks of other phenotypic characteristics of NAFLD could be utilized to prioritize other phenotypic modulating drugs. In addition, our current approach provides new hypotheses for the potential mechanisms of the proposed drugs by computationally identifying

their targets and pathways. Further mechanistic experiments should provide more insights into the MOAs of specific drugs of interest after the testing of more drugs in the vascularized liver acinus MPS (vLAMPS). This should also enhance our understanding of NAFLD progression and optimize our development of personalized therapeutic strategies.

Appendix A Supporting Materials for the Huntington's Disease Study

Appendix A.1 Relative solubility of protective compounds

Compounds were prepared in DMSO and diluted as describe in Methods. To demonstrate that the maximum concentrations used in the PI assay were below their aqueous solubility limit, 2 Pl of the top four DMSO concentrations were diluted in 38 Pl PBS (pH 7.4) and their optical absorbance was measured from 230 to 1000 nm. A linear plot of concentration vs peak absorbance (after subtraction of DMSO/PBS blank) indicated that the aqueous solubility limit had not been reach in that concentration range. A plateau in the curve suggested that the solubility limit was being reached for the compound.

Compound Name	Max Absorbance Wavelength (nM)	Relative Solubility (PM)	n
(Z)-Gugglesterone	250	> 50	2
3-tropanyl-indole-3-carboxylate hydrochloride	230	> 50	3
Beclomethasone	240	> 50	2
Benztropine mesylate	230	> 50	3
Betamethasone	240	> 50	2
Budesonide	250	25	2
Cyproheptadine hydrochloride	230	> 50	2
Domperidone	230	> 50	2
Ethoxzolamide	300	100 - 200	2
Flutamide	230	> 50	2
Hydrocortisone	250	> 50	2
Isoetarine mesylate	230	> 200	2
JWH-015	-	ND	
Lansoprazole	280	> 200	2
Lonidamine	230	> 200	2
Loxapine succinate	230	> 50	3
Meclizine	230	25	2

Mianserin hydrochloride	230	> 50	3
m-Iodobenzylguanidine hemisulfate	230	> 50	2
Papaverine hydrochloride	240	> 50	4
PD 168,077 maleate	230	> 50	3
Quipazine, N-methyl-,dimaleate	240	25	3
Ruthenium red	540	> 50	3
SB 203186	230	> 50	2
Sodium Nitroprusside	230	> 200	2
Tetradecylthioacetic acid	230	> 50	2
Triamcinolone	240	> 50	3
Triprolidine hydrochloride	230	> 50	3
U-83836 dihydrochloride	230	25	2
Vinpocetine	230	> 50	2

Appendix A.2 Targets from DrugBank and STITCH for 32 identified probes

Targets were ranked by the number of interacting probes, probes interact with each target were listed in the corresponding row.

Target ID	Uniprot ID	Target name	Probes Count	Probes
T1	P35367	Histamine H1 receptor	7	Meclizine, Domperidone, Benzatropine, Loxapine, Cyproheptadine, Mianserin, Triprolidine
T2	P28223	5-hydroxytryptamine receptor 2A	6	Loxapine, Benzatropine, Domperidone, Quipazine, N-methyl-,dimaleate, Mianserin, Cyproheptadine
T3	P04150	Glucocorticoid receptor	6	Triamcinolone, Budesonide, Betamethasone, Hydrocortisone, Beclomethasone, Prednisolone
T4	P18825	A-2C adrenergic receptor	5	Mianserin, Benzatropine, Loxapine, Domperidone, Cyproheptadine
T5	P35462	D(3) dopamine receptor	5	Domperidone, Cyproheptadine, Benzatropine, Loxapine, Mianserin
T6	P14416	D(2) dopamine receptor	5	Benzatropine, Mianserin, Domperidone, Loxapine, Cyproheptadine
T7	P28335	5-hydroxytryptamine receptor 2C	5	Mianserin, Loxapine, Benzatropine, Quipazine, N-methyl-,dimaleate, Cyproheptadine
T8	P08172	Muscarinic acetylcholine receptor M2	4	Benzatropine, Loxapine, Cyproheptadine, Mianserin
T9	P08173	Muscarinic acetylcholine receptor M4	4	Loxapine, Cyproheptadine, Benzatropine, Mianserin
T10	P11229	Muscarinic acetylcholine receptor M1	4	Mianserin, Loxapine, Cyproheptadine, Benzatropine
T11	P08913	A-2A adrenergic receptor	4	Benzatropine, Mianserin, Cyproheptadine, Loxapine
T12	P18089	A-2B adrenergic receptor	4	Benzatropine, Cyproheptadine, Loxapine, Mianserin
T13	P41595	5-hydroxytryptamine receptor 2B	4	Benzatropine, Quipazine, N-methyl-,dimaleate, Mianserin, Cyproheptadine
T14	P08912	Muscarinic acetylcholine receptor M5	4	Cyproheptadine, Loxapine, Mianserin, Benzatropine
T15	P20309	Muscarinic acetylcholine receptor M3	4	Cyproheptadine, Benzatropine, Loxapine, Mianserin
T16	P34969	5-hydroxytryptamine receptor 7	3	Loxapine, Cyproheptadine, Mianserin
T17	Q9H3N8	Histamine H4 receptor	3	Cyproheptadine, Mianserin, Loxapine
T18	Q01959	Sodium-dependent dopamine transporter	3	Loxapine, Benzatropine, Mianserin
T19	P21728	D(1A) dopamine receptor	3	Cyproheptadine, Loxapine, Mianserin

T20	P25100	A-1D adrenergic receptor	3	Cyproheptadine, Mianserin, Benzatropine
T21	P50406	5-hydroxytryptamine receptor 6	3	Cyproheptadine, Mianserin, Loxapine
T22	P08908	5-hydroxytryptamine receptor 1A	3	Mianserin, Loxapine, Cyproheptadine
T23	P31645	Sodium-dependent serotonin transporter	3	Quipazine, N-methyl-,dimaleate, Loxapine, Mianserin
T24	P28221	5-hydroxytryptamine receptor 1D	2	Mianserin, Loxapine
T25	P23975	Sodium-dependent noradrenaline transporter	2	Loxapine, Mianserin
T26	P35368	A-1B adrenergic receptor	2	Loxapine, Mianserin
T27	P21918	D(1B) dopamine receptor	2	Loxapine, Mianserin
T28	P35348	A-1A adrenergic receptor	2	Loxapine, Mianserin
T29	P25021	Histamine H2 receptor	2	Cyproheptadine, Loxapine
T30	P08588	Beta-1 adrenergic receptor	2	Loxapine, Isoetarine
T31	P98153	Integral membrane protein DGCR2/IDD	1	Ethoxzolamide
T32	Q9UBN7	Histone deacetylase 6	1	Vorinostat
T33	Q9ULX7	Carbonic anhydrase 14	1	Ethoxzolamide
T34	P04083	Annexin A1	1	Hydrocortisone
T35	P10636	Microtubule-associated protein tau	1	Lansoprazole
T36	P41145	Kappa-type opioid receptor	1	Mianserin
T37	Q9UKV0	Histone deacetylase 9	1	Vorinostat
T38	P47898	5-hydroxytryptamine receptor 5A	1	Loxapine
T39	Q16790	Carbonic anhydrase 9	1	Ethoxzolamide
T40	P35218	Carbonic anhydrase 5A, mitochondrial	1	Ethoxzolamide
T41	Q8N1Q1	Carbonic anhydrase 13	1	Ethoxzolamide
T42	Q13547	Histone deacetylase 1	1	Vorinostat
T43	P46098	5-hydroxytryptamine receptor 3A	1	Loxapine
T44	P28566	5-hydroxytryptamine receptor 1E	1	Loxapine
T45	Q9BY41	Histone deacetylase 8	1	Vorinostat
T46	Q99720	Sigma non-opioid intracellular receptor 1	1	Benzatropine
T47	Q12809	Potassium voltage-gated channel subfamily H member 2	1	Domperidone
T48	P30939	5-hydroxytryptamine receptor 1F	1	Mianserin
T49	P35869	Aryl hydrocarbon receptor	1	Flutamide
T50	P43166	Carbonic anhydrase 7	1	Ethoxzolamide
T51	Q14432	cGMP-inhibited 3',5'-cyclic phosphodiesterase A	1	Papaverine
T52	P28222	5-hydroxytryptamine receptor 1B	1	Loxapine
T53	Q969S8	Histone deacetylase 10	1	Vorinostat
T54	P16066	Atrial natriuretic peptide receptor 1	1	Nitroprusside

T55	P23280	Carbonic anhydrase 6 (EC 4.2.1.1)	1	Ethoxzolamide
T56	Q8WUI4	Histone deacetylase 7	1	Vorinostat
T57	P56524	Histone deacetylase 4	1	Vorinostat
T58	Q96DB2	Histone deacetylase 11	1	Vorinostat
T59	P08185	Corticosteroid-binding globulin	1	Hydrocortisone
T60	O43570	Carbonic anhydrase 12	1	Ethoxzolamide
T61	P07550	Beta-2 adrenergic receptor	1	Isoetarine
T62	P21917	D(4) dopamine receptor	1	Loxapine
T63	Q9Y2D0	Carbonic anhydrase 5B, mitochondrial	1	Ethoxzolamide
T64	P00918	Carbonic anhydrase 2	1	Ethoxzolamide
T65	P22748	Carbonic anhydrase 4	1	Ethoxzolamide
T66	P34972	Cannabinoid receptor 2	1	JWH-015
T67	Q9UQL6	Histone deacetylase 5	1	Vorinostat
T68	P20648	Potassium-transporting ATPase α chain 1	1	Lansoprazole
T69	P10275	Androgen receptor	1	Flutamide
T70	Q9Y233	cAMP and cAMP-inhibited cGMP 3',5'-cyclic phosphodiesterase 10A	1	Papaverine
T71	P00915	Carbonic anhydrase 1	1	Ethoxzolamide
T72	P21554	Cannabinoid receptor 1	1	JWH-015
T73	Q92769	Histone deacetylase 2	1	Vorinostat
T74	Q07343	cAMP-specific 3',5'-cyclic phosphodiesterase 4B	1	Papaverine
T75	O15379	Histone deacetylase 3	1	Vorinostat

Appendix A.3 Mapping of 32 identified probes and targets in KEGG human pathways

Pathways are ranked by the number of mapped probes, probe targets that mapped into each pathway and the corresponding probes were listed in the corresponding pathway row. Target information for each target ID is listed in **Appendix A.2**.

Index	Pathway Name	Probes count	Targets in this pathway	Probes in this pathway
1	Calcium signaling pathway	9	T8, T2, T10, T27, T7, T29, T19, T13, T16, T21, T14, T30, T20, T15, T26, T1, T61, T38, T28	Mianserin, Domperidone, Benzatropine, Cyproheptadine, (Quipazine,N-methyl-,dimaleate), Loxapine, Isoetarine, Meclizine, Triprolidine
2	Inflammatory mediator regulation of TRP channels	8	T1, T7, T2, T13	Mianserin, Domperidone, Benzatropine, Cyproheptadine, (Quipazine,N-methyl-,dimaleate), Loxapine, Meclizine, Triprolidine
3	cGMP-PKG signaling pathway	8	T54, T4, T51, T20, T30, T11, T12, T26, T61, T28	Nitroprusside, Mianserin, Domperidone, Benzatropine, Cyproheptadine, Loxapine, Papaverine, Isoetarine
4	cAMP signaling pathway	8	T8, T54, T10, T27, T44, T52, T51, T24, T48, T21, T19, T30, T6, T22, T74, T61	Nitroprusside, Mianserin, Domperidone, Benzatropine, Cyproheptadine, Loxapine, Papaverine, Isoetarine
5	Gap junction	7	T2, T7, T13, T19, T30, T6	Mianserin, Domperidone, Benzatropine, Cyproheptadine, (Quipazine,N-methyl-,dimaleate), Loxapine, Isoetarine
6	Alcoholism	6	T32, T56, T75, T58, T18, T53, T42, T73, T45, T6, T67, T37, T19, T57	Vorinostat, Mianserin, Domperidone, Benzatropine, Cyproheptadine, Loxapine
7	Rap1 signaling pathway	6	T72, T6	Mianserin, Domperidone, Benzatropine, JWH-015, Cyproheptadine, Loxapine
8	Serotonergic synapse	6	T43, T44, T2, T24, T52, T7, T13, T48, T21, T16, T22, T38, T23	Mianserin, Domperidone, Benzatropine, Cyproheptadine, Quipazine, N-methyl-,dimaleate, Loxapine
9	Amphetamine addiction	5	T19, T18, T42	Loxapine, Benzatropine, Vorinostat, Cyproheptadine, Mianserin
10	Dopaminergic synapse	5	T5, T18, T27, T19, T6, T62	Loxapine, Domperidone, Benzatropine, Cyproheptadine, Mianserin
11	Cocaine addiction	5	T19, T18, T6	Loxapine, Domperidone, Benzatropine, Cyproheptadine, Mianserin
12	Parkinson's disease	5	T19, T18, T6	Loxapine, Domperidone, Benzatropine, Cyproheptadine, Mianserin
13	Morphine addiction	4	T74, T19, T51, T70	Loxapine, Papaverine, Mianserin, Cyproheptadine

14	Cholinergic synapse	4	T8, T9, T14, T10, T15	Loxapine, Benzatropine, Cyproheptadine, Mianserin
15	PI3K-Akt signaling pathway	4	T8, T10	Loxapine, Benzatropine, Cyproheptadine, Mianserin
16	Ras signaling pathway	3	T16	Loxapine, Mianserin, Cyproheptadine
17	Purine metabolism	2	T74, T54, T51, T70	Papaverine, Nitroprusside
18	Endocytosis	2	T61, T30	Loxapine, Isoetarine
19	AMPK signaling pathway	2	T28	Loxapine, Mianserin
20	Pathways in cancer	2	T69, T73, T42	Flutamide, Vorinostat
21	Oxidative phosphorylation	1	T68	Lansoprazole
22	Retrograde endocannabinoid signaling	1	T72	JWH-015
23	Alzheimer's disease	1	T35	Lansoprazole
24	Proximal tubule bicarbonate reclamation	1	T64, T65	Ethoxzolamide
25	Huntington's disease	1	T73, T42	Vorinostat
26	Nitrogen metabolism	1	T64, T65, T63, T50, T55, T33, T71, T39, T60, T40, T41	Ethoxzolamide
27	Cell cycle	1	T73, T42	Vorinostat
28	Transcriptional misregulation in cancer	1	T73, T42	Vorinostat
29	Longevity regulating pathway - multiple species	1	T73, T42	Vorinostat
30	MicroRNAs in cancer	1	T57, T42	Vorinostat
31	Notch signaling pathway	1	T73, T42	Vorinostat
32	MAPK signaling pathway	1	T35	Lansoprazole
33	Epstein-Barr virus infection	1	T57, T67, T73, T42	Vorinostat
34	Viral carcinogenesis	1	T32, T58, T75, T53, T42, T73, T45, T56, T67, T37, T57	Vorinostat

Appendix A.4 Combination Pairs

Each combination is represented by two compound names connected with an underscore.

Combination Number	Combination	Combination Number	Combination
1	Betamethasone_Lonidamine	81	Ruthenium red_Budesonide
2	Sodium Nitroprusside_Triamcinolone	82	Ruthenium red_3-tropanyl-indole-3-carboxylate hydrochloride
3	Sodium Nitroprusside_Betamethasone	83	Triprolidine hydrochloride_3-tropanyl-indole-3-carboxylate hydrochloride
4	Sodium Nitroprusside_Beclomethasone	84	Beclomethasone_Budesonide
5	Ethoxzolamide_Beclomethasone	85	Ethoxzolamide_JWH-015
6	Triprolidine hydrochloride_Betamethasone	86	Triprolidine hydrochloride_Domperidone
7	Domperidone_Isoetarine mesylate	87	Triprolidine hydrochloride_Quipazine,N-methyl-,dimaleate
8	Sodium Nitroprusside_Budesonide	88	Triamcinolone_3-tropanyl-indole-3-carboxylate hydrochloride
9	Isoetarine mesylate_m-Iodobenzylguanidine hemisulfate	89	Ethoxzolamide_Lansoprazole
10	Sodium Nitroprusside_Isoetarine mesylate	90	Beclomethasone_Betamethasone
11	Sodium Nitroprusside_Lansoprazole	91	Ethoxzolamide_Mianserin hydrochloride
12	Ethoxzolamide_Betamethasone	92	Ethoxzolamide_m-Iodobenzylguanidine hemisulfate
13	Sodium Nitroprusside_Mianserin hydrochloride	93	Budesonide_3-tropanyl-indole-3-carboxylate hydrochloride
14	Beclomethasone_Quipazine,N-methyl-,dimaleate	94	Ruthenium red_Lonidamine
15	Sodium Nitroprusside_Loxapine succinate	95	Triprolidine hydrochloride_Budesonide
16	Ethoxzolamide_Loxapine succinate	96	Triamcinolone_Cyproheptadine hydrochloride
17	Ethoxzolamide_Domperidone	97	3-tropanyl-indole-3-carboxylate hydrochloride_PD168,077 maleate
18	Ruthenium red_Betamethasone	98	Ethoxzolamide_PD168,077 maleate
19	3-tropanyl-indole-3-carboxylate hydrochloride_Isoetarine mesylate	99	Budesonide_Isoetarine mesylate
20	Benztropine mesylate_Isoetarine mesylate	100	Triamcinolone_Quipazine,N-methyl-,dimaleate
21	Isoetarine mesylate_Loxapine succinate	101	Ruthenium red_Benztropine mesylate
22	Domperidone_m-Iodobenzylguanidine hemisulfate	102	Triamcinolone_Budesonide

23	Sodium Nitroprusside_U-83836 dihydrochloride	103	Ruthenium red_Triprolidine hydrochloride
24	Tetradecylthioacetic acid_Budesonide	104	Sodium Nitroprusside_Cyproheptadine hydrochloride
25	Betamethasone_Quipazine,N-methyl-,dimaleate	105	3-tropanyl-indole-3-carboxylate hydrochloride_Papaverine hydrochloride
26	Tetradecylthioacetic acid_Betamethasone	106	Ethoxzolamide_Isoetarine mesylate
27	Tetradecylthioacetic acid_Isoetarine mesylate	107	Lonidamine_Benztropine mesylate
28	Isoetarine mesylate_Mianserin hydrochloride	108	3-tropanyl-indole-3-carboxylate hydrochloride_Mianserin hydrochloride
29	Isoetarine mesylate_Papaverine hydrochloride	109	Sodium Nitroprusside_Ethoxzolamide
30	Betamethasone_Isoetarine mesylate	110	Lansoprazole_Loxapine succinate
31	Triamcinolone_Benztropine mesylate	111	Ethoxzolamide_Papaverine hydrochloride
32	Domperidone_Lansoprazole	112	Ruthenium red_Quipazine,N-methyl-,dimaleate
33	Beclomethasone_Isoetarine mesylate	113	Mianserin hydrochloride_Papaverine hydrochloride
34	Sodium Nitroprusside_Lonidamine	114	Tetradecylthioacetic acid_Flutamide
35	Triprolidine hydrochloride_Beclomethasone	115	Mianserin hydrochloride_PD168,077 maleate
36	Triamcinolone_Lonidamine	116	Domperidone_Loxapine succinate
37	Beclomethasone_3-tropanyl-indole-3-carboxylate hydrochloride	117	Lonidamine_Domperidone
38	Betamethasone_3-tropanyl-indole-3-carboxylate hydrochloride	118	Benztropine mesylate_m-Iodobenzylguanidine hemisulfate
39	Beclomethasone_Domperidone	119	Flutamide_Loxapine succinate
40	Tetradecylthioacetic acid_Triamcinolone	120	Tetradecylthioacetic acid_Quipazine,N-methyl-,dimaleate
41	Sodium Nitroprusside_Triprolidine hydrochloride	121	Ruthenium red_Isoetarine mesylate
42	Triamcinolone_Domperidone	122	Tetradecylthioacetic acid_3-tropanyl-indole-3-carboxylate hydrochloride
43	Ethoxzolamide_Budesonide	123	Tetradecylthioacetic acid_Lonidamine
44	Domperidone_Papaverine hydrochloride	124	Domperidone_Mianserin hydrochloride
45	Isoetarine mesylate_PD168,077 maleate	125	Triprolidine hydrochloride_Benztropine mesylate
46	Ethoxzolamide_Triamcinolone	126	Ethoxzolamide_Ruthenium red
47	Sodium Nitroprusside_3-tropanyl-indole-3-carboxylate hydrochloride	127	Lansoprazole_m-Iodobenzylguanidine hemisulfate
48	Betamethasone_Benztropine mesylate	128	Loxapine succinate_Mianserin hydrochloride

49	Ethoxzolamide_Lonidamine	129	Benztropine mesylate_Papaverine hydrochloride
50	Triamcinolone_Isoetarine mesylate	130	Ruthenium red_Triamcinolone
51	Domperidone_PD168,077 maleate	131	Triamcinolone_Beclomethasone
52	Ethoxzolamide_Triprolidine hydrochloride	132	Domperidone_Flutamide
53	Loxapine succinate_m-lodobenzylguanidine hemisulfate	133	3-tropanyl-indole-3-carboxylate hydrochloride_Loxapine succinate
54	Budesonide_Quipazine,N-methyl-,dimaleate	134	Lonidamine_Quipazine,N-methyl-,dimaleate
55	Beclomethasone_Lonidamine	135	Tetradecylthioacetic acid_Benztropine mesylate
56	Sodium Nitroprusside_Domperidone	136	Cyproheptadine hydrochloride_Lonidamine
57	Ethoxzolamide_Benztropine mesylate	137	Tetradecylthioacetic acid_Triprolidine hydrochloride
58	Ruthenium red_Domperidone	138	Tetradecylthioacetic acid_Domperidone
59	Ethoxzolamide_Quipazine,N-methyl-,dimaleate	139	Budesonide_Benztropine mesylate
60	Loxapine succinate_PD168,077 maleate	140	Cyproheptadine hydrochloride_Quipazine,N-methyl-,dimaleate
61	3-tropanyl-indole-3-carboxylate hydrochloride_m-lodobenzylguanidine hemisu	141	Beclomethasone_Cyproheptadine hydrochloride
62	Budesonide_Lonidamine	142	Benztropine mesylate_Loxapine succinate
63	Triamcinolone_Triprolidine hydrochloride	143	Papaverine hydrochloride_PD168,077 maleate
64	Sodium Nitroprusside_Benztropine mesylate	144	Betamethasone_Cyproheptadine hydrochloride
65	Triprolidine hydrochloride_Isoetarine mesylate	145	Triprolidine hydrochloride_Flutamide
66	Triprolidine hydrochloride_Lonidamine	146	m-lodobenzylguanidine hemisulfate_PD168,077 maleate
67	Tetradecylthioacetic acid_(Z)-Gugglesterone	147	Lonidamine_3-tropanyl-indole-3-carboxylate hydrochloride
68	Ethoxzolamide_3-tropanyl-indole-3-carboxylate hydrochloride	148	Ethoxzolamide_Cyproheptadine hydrochloride
69	Beclomethasone_Benztropine mesylate	149	Tetradecylthioacetic acid_Cyproheptadine hydrochloride
70	Loxapine succinate_Papaverine hydrochloride	150	m-lodobenzylguanidine hemisulfate_Papaverine hydrochloride
71	Ruthenium red_Beclomethasone	151	Ruthenium red_Cyproheptadine hydrochloride
72	Ethoxzolamide_Tetradecylthioacetic acid	152	Benztropine mesylate_Domperidone
73	Lonidamine_Isoetarine mesylate	153	Sodium Nitroprusside_Tetradecylthioacetic acid

74	Triamcinolone_Betamethasone	154	Triprolidine hydrochloride_Cyproheptadine hydrochloride
75	Betamethasone_Budesonide	155	Sodium Nitroprusside_Ruthenium red
76	Lonidamine_Flutamide	156	Isoetarine mesylate_Lansoprazole
77	Triamcinolone_Flutamide	157	Flutamide_m-Iodobenzylguanidine hemisulfate
78	Budesonide_Domperidone	158	3-tropanyl-indole-3-carboxylate hydrochloride_Benztropine mesylate
79	Tetradecylthioacetic acid_Beclomethasone	159	Ruthenium red_Tetradecylthioacetic acid
80	Ethoxzolamide_Flutamide		

Appendix A.5 Synergistic Pairs

Compound combinations were run at least once on two different days. n= the total number of combination samples analyzed. An n = 2 indicates a combination was run only once on each of the two days.

Combination Number	Combination	Avg Percent Recovery	Std	n*	Avg Combi Ratio	Std	Median BCI
1	Betamethasone_Lonidamine	84.68	6.12	4	1.98	0.62	1.39
2	Sodium Nitroprusside_Triamcinolone	89.08	4.62	5	1.89	0.34	1.34
3	Sodium Nitroprusside_Betamethasone	96.89	4.67	5	1.81	0.39	1.27
4	Sodium Nitroprusside_Beclomethasone	94.51	2.60	5	1.78	0.26	1.26
5	Ethoxzolamide_Beclomethasone	86.85	2.49	5	1.72	0.11	1.23
6	Triprolidine hydrochloride_Betamethasone	91.41	2.80	4	1.59	0.28	1.22
7	Domperidone_Isoetarine mesylate	78.69	12.41	4	1.79	0.12	1.20
8	Sodium Nitroprusside_Budesonide	100.88	1.04	5	1.50	0.08	1.19
9	Isoetarine mesylate_m-Iodobenzylguanidine hemisulfate	71.92	6.76	4	1.86	0.06	1.17
10	Sodium Nitroprusside_Isoetarine mesylate	80.63	5.73	5	1.87	0.36	1.17
11	Sodium Nitroprusside_Lansoprazole	82.22	6.02	4	1.56	0.08	1.17
12	Ethoxzolamide_Betamethasone	77.30	33.72	6	1.55	0.02	1.15
13	Sodium Nitroprusside_Mianserin hydrochloride	75.42	8.10	4	1.63	0.20	1.14
14	Beclomethasone_Quipazine,N-methyl-,dimaleate	79.64	2.89	2	1.66	0.15	1.14
15	Sodium Nitroprusside_Loxapine succinate	80.46	2.19	4	1.49	0.03	1.13
16	Ethoxzolamide_Loxapine succinate	74.43	4.13	4	1.38	0.03	1.13
17	Ethoxzolamide_Domperidone	81.30	17.04	6	1.68	0.27	1.12
18	Ruthenium red_Betamethasone	79.17	7.01	4	1.51	0.34	1.12
19	3-tropanyl-indole-3-carboxylate hydrochloride_Isoetarine mesylate	69.62	3.52	4	1.76	0.08	1.12
20	Benztropine mesylate_Isoetarine mesylate	65.81	2.06	4	1.71	0.24	1.12
21	Isoetarine mesylate_Loxapine succinate	80.01	4.50	4	1.48	0.06	1.11
22	Domperidone_m-Iodobenzylguanidine hemisulfate	72.00	3.99	4	1.64	0.04	1.11
23	Sodium Nitroprusside_U-83836 dihydrochloride	84.34	2.71	5	1.45	0.12	1.11
24	Tetradecylthioacetic acid_Budesonide	96.68	1.02	4	1.43	0.07	1.11
25	Betamethasone_Quipazine,N-methyl-,dimaleate	83.07	0.57	2	1.50	0.12	1.11
26	Tetradecylthioacetic acid_Betamethasone	85.04	3.50	4	1.42	0.19	1.11
27	Tetradecylthioacetic acid_Isoetarine mesylate	88.66	4.19	2	1.39	0.08	1.10
28	Isoetarine mesylate_Mianserin hydrochloride	73.47	5.54	4	1.60	0.34	1.10

29	Isoetarine mesylate_Papaverine hydrochloride	74.95	6.28	4	1.57	0.02	1.10
30	Betamethasone_Isoetarine mesylate	82.64	9.58	2	1.48	0.04	1.10
31	Triamcinolone_Benztropine mesylate	74.81	0.50	2	1.31	0.30	1.09
32	Domperidone_Lansoprazole	79.79	12.84	4	1.51	0.11	1.08
33	Beclomethasone_Isoetarine mesylate	77.04	5.53	2	1.47	0.01	1.08
34	Sodium Nitroprusside_Lonidamine	66.90	22.35	5	1.39	0.09	1.08
35	Triprolidine hydrochloride_Beclomethasone	86.17	3.73	4	1.42	0.15	1.08
36	Triamcinolone_Lonidamine	58.55	8.25	4	1.72	0.49	1.07
37	Beclomethasone_3-tropanyl-indole-3-carboxylate_hydrochloride	78.29	3.15	2	1.53	0.01	1.07
38	Betamethasone_3-tropanyl-indole-3-carboxylate_hydrochloride	82.88	8.42	2	1.49	0.02	1.07
39	Beclomethasone_Domperidone	83.70	0.17	2	1.42	0.02	1.07
40	Tetradecylthioacetic acid_Triamcinolone	77.97	5.99	4	1.32	0.29	1.07
41	Sodium Nitroprusside_Triprolidine hydrochloride	86.41	5.97	5	1.46	0.25	1.07
42	Triamcinolone_Domperidone	73.41	1.45	2	1.25	0.01	1.07
43	Ethoxzolamide_Budesonide	87.19	3.45	6	1.29	0.03	1.07
44	Domperidone_Papaverine hydrochloride	75.42	11.88	4	1.58	0.19	1.07
45	Isoetarine mesylate_PD168,077 maleate	71.38	2.48	4	1.57	0.15	1.07
46	Ethoxzolamide_Triamcinolone	61.43	7.36	6	1.54	0.33	1.06
47	Sodium Nitroprusside_3-tropanyl-indole-3-carboxylate hydrochloride	68.86	12.35	5	1.53	0.13	1.06
48	Betamethasone_Benztropine mesylate	85.13	0.92	2	1.44	0.29	1.05
49	Ethoxzolamide_Lonidamine	64.69	6.30	6	1.61	0.37	1.05
50	Triamcinolone_Isoetarine mesylate	61.87	12.97	2	1.39	0.12	1.05
51	Domperidone_PD168,077 maleate	72.46	5.13	4	1.55	0.17	1.04
52	Ethoxzolamide_Triprolidine hydrochloride	77.82	7.22	6	1.29	0.17	1.04
53	Loxapine succinate_m-Iodobenzylguanidine hemisulfate	73.76	2.68	4	1.36	0.00	1.04
54	Budesonide_Quipazine,N-methyl-,dimaleate	85.41	12.39	2	1.25	0.12	1.04
55	Beclomethasone_Lonidamine	69.30	21.92	4	1.37	0.03	1.03
56	Sodium Nitroprusside_Domperidone	61.74	25.78	5	1.26	0.37	1.03
57	Ethoxzolamide_Benztropine mesylate	58.69	10.99	6	1.46	0.23	1.02
58	Ruthenium red_Domperidone	80.82	1.76	2	1.37	0.05	1.01
59	Ethoxzolamide_Quipazine,N-methyl-,dimaleate	60.16	7.97	4	1.59	0.17	1.01
60	Loxapine succinate_PD168,077 maleate	75.92	2.61	4	1.40	0.01	1.01
61	3-tropanyl-indole-3-carboxylate_hydrochloride_m-Iodobenzylguanidine hemisulfate	61.79	2.27	4	1.57	0.07	1.01

Appendix B Supporting Materials for the Drug Abuse Study

Appendix B contains the information and structure of the addictive drugs in the drug addiction study (see Section 3.2.2).

Appendix B.1 Dataset of 50 addictive drugs and their corresponding groups and identifiers

'# of targets' counts the number of targets of drugs recorded in either DrugBank v5 or STITCH v5.

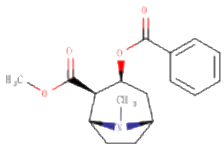
Index	Drug group and name		DrugBank ID	Pubchem ID	# of targets	Reference
1	CNS Stimulants	Cocaine	DB00907	446220	45	(Gawin and Ellinwood, 1988)
2		Methylphenidate	DB00422	4158	3	(Klein-Schwartz, 2002)
3		Methamphetamine	DB01577	10836	11	(Winslow, et al., 2007)
4		Amphetamine	DB00182	3007	17	(Kramer, et al., 1967)
5		Phenmetrazine	DB00830	4762	2	(Mellar and Hollister, 1982)
6		Phendimetrazine	DB01579	30487	3	(Bolin, et al., 2016)
7	CNS Depressants	Pentobarbital	DB00312	4737	27	(Griffiths, et al., 1979)
8		Zaleplon	DB00962	5719	2	(Dooley and Plosker, 2000)
9		Zolpidem	DB00425	5732	13	(Madrak and Rosenberg, 2001)
10		Glutethimide	DB01437	3487	16	(Jones and Mayberry, 1986)
11		Flunitrazepam	DB01544	3380	16	(Druid, et al., 2001)

12		Diazepam	DB00829	3016	24	(Woody, et al., 1975)
13		Lorazepam	DB00186	3958	17	(Troisi 2nd, et al., 1993)
14		Triazolam	DB00897	5556	20	(Fleming, 1983)
15		Alprazolam	DB00404	2118	17	(Rush, et al., 1993)
16		Chlordiazepoxide	DB00475	2712	16	(Hollister, et al., 1961)
17		Promethazine	DB01069	4927	19	(Tsay, et al., 2015)
18		Eszopiclone	DB00402	969472	17	(Hajak, et al., 2003)
19		Gamma Hydroxybutyric Acid (GHB)	DB01440	11266	2	(Galloway, et al., 2000)
20	Opioids	Meperidine	NA	4058	2	(Joranson, et al., 2000)
21		Fentanyl	DB00813	3345	3	(Gold, et al., 2006)
22		Methadone	DB00333	4095	11	(Cicero and Inciardi, 2005)
23		Loperamide	DB00836	3955	5	(Lasoff, et al., 2017)
24		Oxymorphone	DB01192	5284604	3	(Babalonis, et al., 2016)
25		Hydromorphone	DB00327	5284570	3	(Walsh, et al., 2008)
26		Hydrocodone	DB00956	5284569	2	(Babalonis, et al., 2016)
27		Oxycodone	DB00497	5284603	5	(Harris, et al., 2014)
28		Codeine	DB00318	5284371	3	(Kathiramalainathan, et al., 2000)
29		Morphine	DB00295	5288826	4	(Preston, et al., 1991)
30		Heroin	DB01452	5462328	3	(Buttner, et al., 2000)
31		Buprenorphine	DB00921	644073	4	(O'Connor, et al., 1988)
32	Cannabinoids	Cannabichromene	NA	30219	1	(Poklis, et al., 2010)
33		Dronabinol	DB00470	16078	3	(Calhoun, et al., 1998)

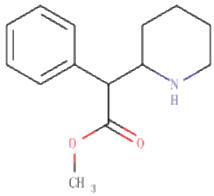
34		Cannabidiol	DB09061	644019	2	(Robson, 2011)
35		Cannabinol	NA	2543	2	(Yamamoto, et al., 2003)
36		Anandamide	NA	5281969	2	(Solinas, et al., 2007)
37		2-AG	NA	5282280	2	(Solinas, et al., 2007)
38	Steroids	Oxandrolone	DB00621	5878	1	(Bahrke and Yesalis, 2004)
39		Oxymetholone	DB06412	5281034	2	(Bahrke and Yesalis, 2004)
40		Nandrolone	DB13169	9904	2	(Kouvelas, et al., 2008)
41	Hallucinogens	Psilocin	NA	4980	11	(Ludwig and Levine, 1965)
42		Dimethyltryptamine	DB01488	8441	3	(Winstock, et al., 2014)
43		Psilocybin	DB11664	10624	5	(Passie, et al., 2002)
44		Lysergic Acid Diethylamide (LSD)	DB04829	5761	12	(Simpson, et al., 1997)
45		Ketamine	DB01221	3821	20	(Dotson, et al., 1995)
46		Phencyclidine	DB03575	6468	10	(Slavney, et al., 1977)
47		Midomafetamine	DB01454	1615	8	(Seger, 2010)
48		Mescaline	NA	4076	2	(Neiman, et al., 2000)
49		Dextrorphan	NA	5360697	1	(Schwartz, 2005)
50		Dextromethorphan	DB00514	5360696	21	(Boyer, 2004)

Appendix B.2 2D structures of the dataset of 50 addictive drugs

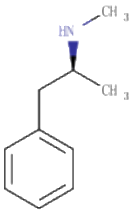
The names of drugs/chemicals are colored green, blue, red, cyan, light brown, black and magenta for the 6 CNS stimulants, 13 CNS depressants, 12 opioids, 7 cannabinoids, 4 anabolic steroids and 10 hallucinogens, respectively.



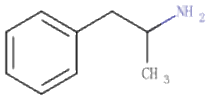
Cocaine



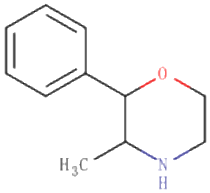
Methylphenidate



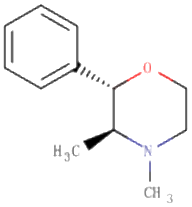
Methamphetamine



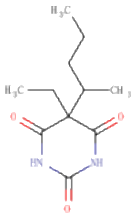
Amphetamine



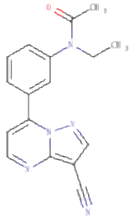
Phenmetrazine



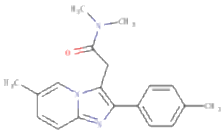
Phendimetrazine



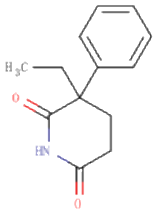
Pentobarbital



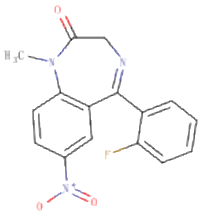
Zaleplon



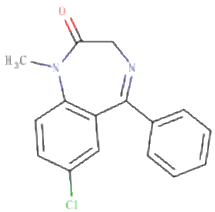
Zolpidem



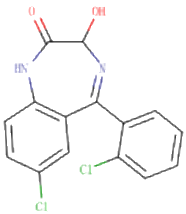
Glutethimide



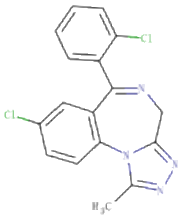
Flunitrazepam



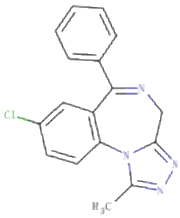
Diazepam



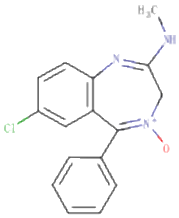
Lorazepam



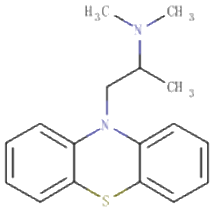
Triazolam



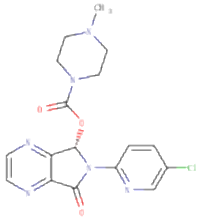
Alprazolam



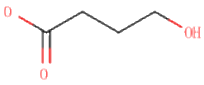
Chlordiazepoxide



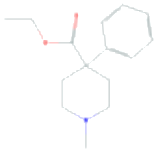
Promethazine



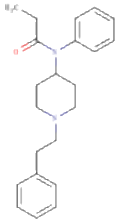
Eszopiclone



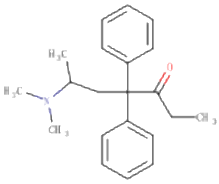
Gamma Hydroxybutyric Acid (GHB)



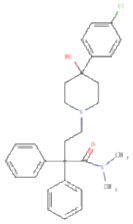
Meperidine



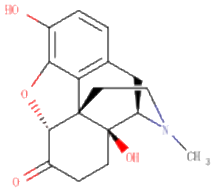
Fentanyl



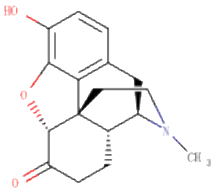
Methadone



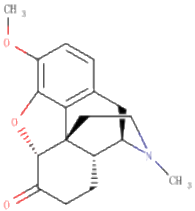
Loperamide



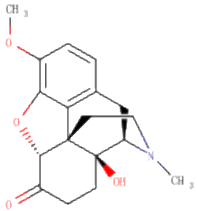
Oxymorphone



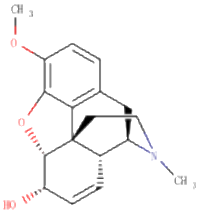
Hydromorphone



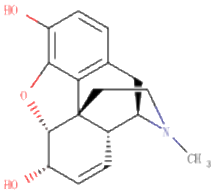
Hydrocodone



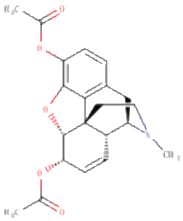
Oxycodone



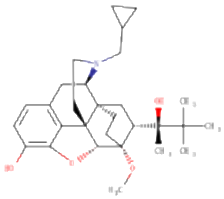
Codeine



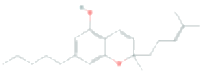
Morphine



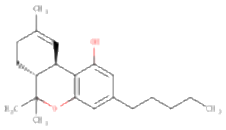
Heroin



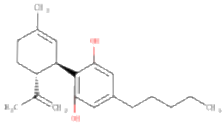
Buprenorphine



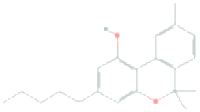
Cannabichromene



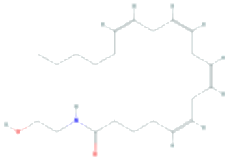
Dronabinol



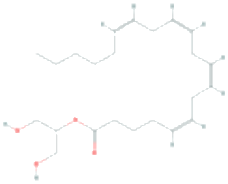
Cannabidiol



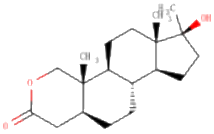
Cannabinol



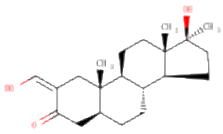
Anandamide



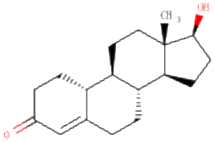
2-AG



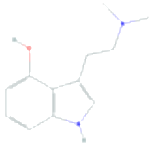
Oxandrolone



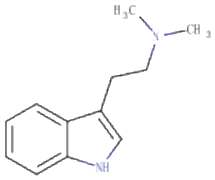
Oxymetholone



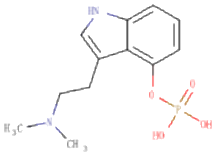
Nandrolone



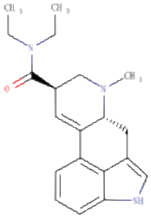
Psilocin



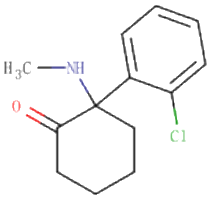
Dimethyltryptamine



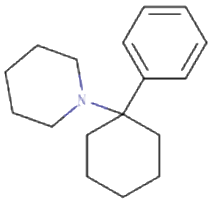
Psilocybin



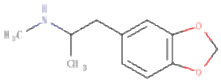
Lysergic Acid Diethylamide (LSD)



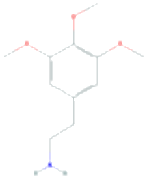
Ketamine



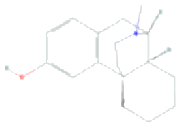
Phencyclidine



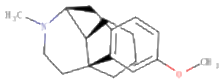
Midomafetamine



Mescaline



Dextrorphan



Dextromethorphan

Appendix C Supporting Materials for the NAFLD Study

Appendix C contains the information of differentially regulated pathways and supporting materials for drug prediction in the NAFLD study (see Section 3.3.2).

Appendix C.1 Categorization of differentially regulated pathways

C1: Insulin resistance and oxidative stress; C2: Cell Stress, apoptosis and lipotoxicity; C3: Inflammation; C4: Fibrosis

KEGG Pathway Group: Metabolism (number: 45, percentage: 32%)				
KEGG Pathway Subgroup	**KEGG ID**	**Pathway Name**	**NAFLD Pathway Category**	**References**
Amino acid metabolism (n = 9)	hsa00340	Histidine metabolism	C1	
	hsa00260	Glycine, serine and threonine metabolism	C2	
	hsa00360	Phenylalanine metabolism	C2	(Kim, et al., 2018)
	hsa00330	Arginine and proline metabolism	C3	(Dumas, et al., 2014)
	hsa00380	Tryptophan metabolism	C3	(Oates, et al., 2019)
	hsa00250	Alanine, aspartate and glutamate metabolism	C3, C4	(Oates, et al., 2019)
	hsa00290	Valine, leucine and isoleucine biosynthesis	C4	(Gaggini, et al., 2018)
	hsa00350	Tyrosine metabolism	C4	
	hsa00280	Valine, leucine and isoleucine degradation	C4	(Gaggini, et al., 2018)
Carbohydrate metabolism (n = 10)	hsa00052	Galactose metabolism	C1	(Basaranoglu, et al., 2013)
	hsa00020	Citrate cycle (TCA cycle)	C1	
	hsa00650	Butanoate metabolism	C1	(Endo, et al., 2013)
	hsa00562	Inositol phosphate metabolism	C1	(Kanehisa, et al., 2017)

	hsa00500	Starch and sucrose metabolism	C1	
	hsa00630	Glyoxylate and dicarboxylate metabolism	C1, C2	(Kanehisa, et al., 2017)
	hsa00030	Pentose phosphate pathway	C1, C2	(Jin, et al., 2018)
	hsa00520	Amino sugar and nucleotide sugar metabolism	C1, C2, C3	(Jensen, et al., 2018)
	hsa00051	Fructose and mannose metabolism	C1, C2, C3	(Jegatheesan and De Bandt, 2017)
	hsa00010	Glycolysis / Gluconeogenesis	C1, C4	(Zhao, et al., 2020)
Energy metabolism (n = 2)	hsa00190	Oxidative phosphorylation	C1	
	hsa00920	Sulfur metabolism	C1	
Glycan biosynthesis and metabolism (n = 8)	hsa00511	Other glycan degradation	C4	(Rostami and Parsian, 2013)
	hsa00533	Glycosaminoglycan biosynthesis - keratan sulfate	C4	(Rostami and Parsian, 2013)
	hsa00603	Glycosphingolipid biosynthesis - globo and isoglobo series	C4	(Rostami and Parsian, 2013)
	hsa00604	Glycosphingolipid biosynthesis - ganglio series	C4	(Rostami and Parsian, 2013)
	hsa00531	Glycosaminoglycan degradation	C4	(Rostami and Parsian, 2013)
	hsa00532	Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	C4	(Rostami and Parsian, 2013)
	hsa00601	Glycosphingolipid biosynthesis - lacto and neolacto series	C4	(Rostami and Parsian, 2013)
	hsa00534	Glycosaminoglycan biosynthesis - heparan sulfate / heparin	C4	(Rostami and Parsian, 2013)
Lipid metabolism (n = 5)	hsa00564	Glycerophospholipid metabolism	C2	
	hsa00561	Glycerolipid metabolism	C2	
	hsa00120	Primary bile acid biosynthesis	C2	
	hsa00140	Steroid hormone biosynthesis	C2	
	hsa00071	Fatty acid degradation	C2	
Metabolism of cofactors and vitamins (n = 4)	hsa00860	Porphyrin and chlorophyll metabolism	C1	
	hsa00770	Pantothenate and CoA biosynthesis	C1	
	hsa00760	Nicotinate and nicotinamide metabolism	C1	(Guarino and Dufour, 2019)
	hsa00670	One carbon pool by folate	C2	(Radziejewska, et al., 2019)
Metabolism of other amino acids (n = 2)	hsa00480	Glutathione metabolism	C1	(Liu, et al., 2015)
	hsa00450	Selenocompound metabolism	C1	

Metabolism of terpenoids and polyketides (n = 1)	hsa00900	Terpenoid backbone biosynthesis	C1	(Kuzuyama, 2017)
Nucleotide metabolism (n = 2)	hsa00230	Purine metabolism	C1	(Cai, et al., 2014)
	hsa00240	Pyrimidine metabolism	C1, C2	(Le, et al., 2013)
Xenobiotics biodegradation and metabolism (n = 2)	hsa00983	Drug metabolism - other enzymes	C2	(Naik, et al., 2013)
	hsa00980	Metabolism of xenobiotics by cytochrome P450	C2	(Naik, et al., 2013)
KEGG Pathway Group: Human Diseases (number: 31, percentage: 22%)				
Cancer: overview (n = 1)	hsa05200	Pathways in cancer	C5	
Cancer: specific types (n = 14)	hsa05210	Colorectal cancer	C5	
	hsa05216	Thyroid cancer	C5	
	hsa05214	Glioma	C5	
	hsa05215	Prostate cancer	C5	
	hsa05217	Basal cell carcinoma	C5	
	hsa05223	Non-small cell lung cancer	C5	
	hsa05222	Small cell lung cancer	C5	
	hsa05212	Pancreatic cancer	C5	
	hsa05218	Melanoma	C5	
	hsa05221	Acute myeloid leukemia	C5	
	hsa05213	Endometrial cancer	C5	
	hsa05220	Chronic myeloid leukemia	C5	
	hsa05211	Renal cell carcinoma	C5	
	hsa05219	Bladder cancer	C5	
Cardiovascular disease (n = 4)	hsa05416	Viral myocarditis	C5	
	hsa05414	Dilated cardiomyopathy (DCM)	C5	
	hsa05410	Hypertrophic cardiomyopathy (HCM)	C5	
	hsa05412	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	C5	
Endocrine and metabolic disease (n = 1)	hsa04940	Type I diabetes mellitus	C5	
Immune disease (n = 4)	hsa05322	Systemic lupus erythematosus	C5	
	hsa05332	Graft-versus-host disease	C5	
	hsa05330	Allograft rejection	C5	
	hsa05340	Primary immunodeficiency	C5	
Infectious disease: bacterial (n = 2)	hsa05130	Pathogenic Escherichia coli infection	C5	
	hsa05110	Vibrio cholerae infection	C5	

Infectious disease: parasitic (n = 1)	hsa05140	Leishmaniasis	C5	
Neurodegenerative disease (n = 4)	hsa05016	Huntington disease	C5	
	hsa05010	Alzheimer disease	C5	
	hsa05012	Parkinson disease	C5	
	hsa05020	Prion diseases	C5	
KEGG Pathway Group: Organismal Systems (number: 28, percentage: 20%)				
Circulatory system (n = 2)	hsa04260	Cardiac muscle contraction	C6	(Mangi, et al., 2017),(Ismaiel and Dumitrascu, 2019)
	hsa04270	Vascular smooth muscle contraction	C6	(Pasarin, et al., 2017)
Development and regeneration (n = 2)	hsa04360	Axon guidance	C3	(Taipale, et al., 2018)
	hsa04320	Dorso-ventral axis formation	C7	
Endocrine system (n = 5)	hsa04916	Melanogenesis	C1, C3	(Page, et al., 2011)
	hsa04614	Renin-angiotensin system	C1, C3, C4	(Paschos and Tziomalos, 2012; Simoes, et al., 2017)
	hsa03320	PPAR signaling pathway	C2	(Liss and Finck, 2017)
	hsa04912	GnRH signaling pathway	C2	(Kanehisa, et al., 2017)
	hsa04914	Progesterone-mediated oocyte maturation	C7	
Excretory system (n = 2)	hsa04964	Proximal tubule bicarbonate reclamation	C6	
	hsa04962	Vasopressin-regulated water reabsorption	C6	(Li, et al., 2019)
Immune system (n = 13)	hsa04664	Fc epsilon RI signaling pathway	C3	
	hsa04620	Toll-like receptor signaling pathway	C3	
	hsa04660	T cell receptor signaling pathway	C1, C3, C4	(Van Herck, et al., 2019)
	hsa04662	B cell receptor signaling pathway	C3	
	hsa04666	Fc gamma R-mediated phagocytosis	C3	
	hsa04650	Natural killer cell mediated cytotoxicity	C3	
	hsa04670	Leukocyte transendothelial migration	C3	
	hsa04062	Chemokine signaling pathway	C3	
	hsa04621	NOD-like receptor signaling pathway	C3	
	hsa04672	Intestinal immune network for IgA production	C3	

	hsa04612	Antigen processing and presentation	C3	
	hsa04622	RIG-I-like receptor signaling pathway	C3	
	hsa04640	Hematopoietic cell lineage	C3	
Nervous system (n = 2)	hsa04722	Neurotrophin signaling pathway	C6	(Davis, et al., 2012)·(Muirhead and Monaghan, 2012)
	hsa04720	Long-term potentiation	C6	(Ross, et al., 2012)
Sensory system (n = 2)	hsa04740	Olfactory transduction	C6	(Paz-Filho, et al., 2013)
	hsa04742	Taste transduction	C6	
KEGG Pathway Group: Environmental Information Processing (number: 14, percentage: 10%)				
Membrane transport (n = 1)	hsa02010	ABC transporters	C2	(Hardwick, et al., 2011; Naik, et al., 2013)
Signal transduction (n = 9)	hsa04070	Phosphatidylinositol signaling system	C1	(Matsuda, et al., 2013)
	hsa04330	Notch signaling pathway	C1	(Zhao, et al., 2018)·(Valenti, et al., 2013)
	hsa04370	VEGF signaling pathway	C2	
	hsa04012	ErbB signaling pathway	C2	
	hsa04310	Wnt signaling pathway	C2, C3, C4	(Zhao, et al., 2020)
	hsa04010	MAPK signaling pathway	C2, C4	(Zhao, et al., 2020)
	hsa04630	Jak-STAT signaling pathway	C3	(Riordan and Nadeau, 2014)
	hsa04350	TGF-beta signaling pathway	C4	(Feaver, et al., 2016)
	hsa04340	Hedgehog signaling pathway	C4	(Syn, et al., 2009)
Signaling molecules and interaction (n = 4)	hsa04512	ECM-receptor interaction	C2	
	hsa04514	Cell adhesion molecules (CAMs)	C2	
	hsa04060	Cytokine-cytokine receptor interaction	C3	(Braunersreuther, et al., 2012)
	hsa04080	Neuroactive ligand-receptor interaction	C6	
KEGG Pathway Group: Genetic Information Processing (number: 11, percentage: 8%)				
Folding, sorting and degradation (n = 4)	hsa04120	Ubiquitin mediated proteolysis	C1, C2, C3	(Luo, et al., 2018)
	hsa03050	Proteasome	C2	(Feaver, et al., 2016)
	hsa04130	SNARE interactions in vesicular transport	C2	

	hsa03018	RNA degradation	C2, C3, C4	
Replication and repair (n = 4)	hsa03450	Non-homologous end-joining	C2	
	hsa03410	Base excision repair	C2	
	hsa03430	Mismatch repair	C2	
	hsa03440	Homologous recombination	C2	
Transcription (n = 2)	hsa03022	Basal transcription factors	C2	
	hsa03020	RNA polymerase	C2	
Translation (n = 1)	hsa03010	Ribosome	C2	
KEGG Pathway Group: Cellular Processes (number: 10, percentage: 7%)				
Cell growth and death (n = 4)	hsa04115	p53 signaling pathway	C1, C2, C4	(Krstic, et al., 2018; Yan, et al., 2018)
	hsa04210	Apoptosis	C2, C3, C4	(Kanda, et al., 2018)
	hsa04110	Cell cycle	C2	
	hsa04114	Oocyte meiosis	C7	
Cell motility (n = 1)	hsa04810	Regulation of actin cytoskeleton	C1, C4	(Chambel, et al., 2015)
Cellular community – eukaryotes (n = 3)	hsa04520	Adherens junction	C2	
	hsa04510	Focal adhesion	C2	
	hsa04540	Gap junction	C2, C3, C4	(Hernandez-Guerra, et al., 2019)
Transport and catabolism (n = 2)	hsa04142	Lysosome	C2	(Du, et al., 2020)
	hsa04146	Peroxisome	C2, C3	(Orabona, et al., 2018)

Appendix C.2 Differentially regulated pathways of each comparison

PF vs. N&S					
Pathway Name	**NAFLD Pathway Category**	**KEGG Pathway Group**	**KEGG Pathway Subgroup**	**Adjusted p-value (FDR)**	**logFoldChange**
Ubiquitin mediated proteolysis	C2	Genetic Information Processing	Folding, sorting and degradation	8.3E-24	1.51
Fructose and mannose metabolism	C1, C3	Metabolism	Carbohydrate metabolism	8.3E-24	-1.53
Apoptosis	C2	Cellular Processes	Cell growth and death	2.2E-23	1.5
Glycosphingolipid biosynthesis - globo and isoglobo series	C4	Metabolism	Glycan biosynthesis and metabolism	1.5E-22	-1.5
Gap junction	C2, C3, C4	Cellular Processes	Cellular community - eukaryotes	1.6E-22	1.5
Glyoxylate and dicarboxylate metabolism	C1, C2	Metabolism	Carbohydrate metabolism	2E-22	-1.46
Wnt signaling pathway	C2, C3, C4	Environmental Information Processing	Signal transduction	4.4E-20	1.42
Pentose phosphate pathway	C1, C2	Metabolism	Carbohydrate metabolism	1.3E-19	-1.4
Amino sugar and nucleotide sugar metabolism	C1, C2, C3	Metabolism	Carbohydrate metabolism	1.6E-19	-1.4
T cell receptor signaling pathway	C3	Organismal Systems	Immune system	2.5E-19	1.39
TGF-beta signaling pathway	C4	Environmental Information Processing	Signal transduction	5.2E-19	1.36
Glycine, serine and threonine metabolism	C2	Metabolism	Amino acid metabolism	5.6E-19	-1.38

Ribosome	C2	Genetic Information Processing	Translation	2.9E-18	-1.29
Sulfur metabolism	C1	Metabolism	Energy metabolism	4E-18	-1.34
Tyrosine metabolism	C4	Metabolism	Amino acid metabolism	4E-18	-1.35
Galactose metabolism	C1	Metabolism	Carbohydrate metabolism	4.8E-18	-1.33
Oxidative phosphorylation	C1	Metabolism	Energy metabolism	1E-17	-1.29
Adherens junction	C2	Cellular Processes	Cellular community - eukaryotes	9.9E-17	1.3
Alanine, aspartate and glutamate metabolism	C3, C4	Metabolism	Amino acid metabolism	6.9E-16	-1.26
Cell cycle	C2	Cellular Processes	Cell growth and death	2.3E-15	1.24
Glycosaminoglycan biosynthesis - keratan sulfate	C4	Metabolism	Glycan biosynthesis and metabolism	8E-15	-1.2
Citrate cycle (TCA cycle)	C1	Metabolism	Carbohydrate metabolism	8.1E-15	-1.22
Fc epsilon RI signaling pathway	C3	Organismal Systems	Immune system	1.5E-14	1.21
Proteasome	C2	Genetic Information Processing	Folding, sorting and degradation	3.5E-14	-1.19
B cell receptor signaling pathway	C3	Organismal Systems	Immune system	3.6E-14	1.19
Jak-STAT signaling pathway	C3	Environmental Information Processing	Signal transduction	4.5E-14	1.19
Arginine and proline metabolism	C3	Metabolism	Amino acid metabolism	5.2E-14	-1.19
Pyrimidine metabolism	C1, C2	Metabolism	Nucleotide metabolism	5.6E-14	-1.19

Non-homologous end-joining	C2	Genetic Information Processing	Replication and repair	5.6E-14	1.15
Basal transcription factors	C2	Genetic Information Processing	Transcription	7.1E-14	1.1
Phenylalanine metabolism	C2	Metabolism	Amino acid metabolism	3.4E-13	-1.15
One carbon pool by folate	C2	Metabolism	Metabolism of cofactors and vitamins	3.9E-13	-1.15
Melanogenesis	C1, C3	Organismal Systems	Endocrine system	7E-13	1.14
Fc gamma R-mediated phagocytosis	C3	Organismal Systems	Immune system	1.1E-12	1.13
Histidine metabolism	C1	Metabolism	Amino acid metabolism	2.3E-12	-1.11
Other glycan degradation	C4	Metabolism	Glycan biosynthesis and metabolism	3.1E-12	-1.07
Toll-like receptor signaling pathway	C3	Organismal Systems	Immune system	3.1E-12	1.11
Regulation of actin cytoskeleton	C1, C4	Cellular Processes	Cell motility	3.3E-12	1.08
Natural killer cell mediated cytotoxicity	C3	Organismal Systems	Immune system	3.8E-12	1.1
Glycosphingolipid biosynthesis - lacto and neolacto series	C4	Metabolism	Glycan biosynthesis and metabolism	5.3E-12	-1.1
Purine metabolism	C1	Metabolism	Nucleotide metabolism	5.5E-12	-1.09
Leukocyte transendothelial migration	C3	Organismal Systems	Immune system	1.3E-11	1.08
Chemokine signaling pathway	C3	Organismal Systems	Immune system	3.2E-11	1.05

RNA polymerase	C2	Genetic Information Processing	Transcription	7.5E-11	-1.04
PPAR signaling pathway	C2	Organismal Systems	Endocrine system	9.5E-11	-1.02
VEGF signaling pathway	C2	Environmental Information Processing	Signal transduction	9.5E-11	1.03
GnRH signaling pathway	C2	Organismal Systems	Endocrine system	5.5E-10	0.99
NOD-like receptor signaling pathway	C3	Organismal Systems	Immune system	5.7E-10	0.99
Tryptophan metabolism	C3	Metabolism	Amino acid metabolism	7.1E-10	-0.99
Cytokine-cytokine receptor interaction	C3	Environmental Information Processing	Signaling molecules and interaction	9.7E-10	0.98
Glycolysis / Gluconeogenesis	C1, C4	Metabolism	Carbohydrate metabolism	1.8E-09	-0.97
Drug metabolism - other enzymes	C2	Metabolism	Xenobiotics biodegradation and metabolism	1.1E-08	-0.92
Glycosphingolipid biosynthesis - ganglio series	C4	Metabolism	Glycan biosynthesis and metabolism	1.2E-08	-0.9
Glycosaminoglycan degradation	C4	Metabolism	Glycan biosynthesis and metabolism	1.3E-08	-0.9
Butanoate metabolism	C1	Metabolism	Carbohydrate metabolism	1.7E-08	-0.91
Glycerophospholipid metabolism	C2	Metabolism	Lipid metabolism	2.9E-08	-0.89
Glycerolipid metabolism	C2	Metabolism	Lipid metabolism	3.3E-08	-0.88
Intestinal immune network for IgA production	C3	Organismal Systems	Immune system	4.6E-08	0.88

Phosphatidylinositol signaling system	C1	Environmental Information Processing	Signal transduction	8.3E-08	0.87
Glutathione metabolism	C1	Metabolism	Metabolism of other amino acids	1E-07	-0.85
Antigen processing and presentation	C3	Organismal Systems	Immune system	1.1E-07	0.85
Base excision repair	C2	Genetic Information Processing	Replication and repair	1.6E-07	-0.83
Focal adhesion	C2	Cellular Processes	Cellular community - eukaryotes	2.1E-07	0.82
Valine, leucine and isoleucine biosynthesis	C4	Metabolism	Amino acid metabolism	2.3E-07	0.82
p53 signaling pathway	C1, C2, C4	Cellular Processes	Cell growth and death	5.3E-07	0.81
Selenocompound metabolism	C1	Metabolism	Metabolism of other amino acids	5.3E-07	-0.81
SNARE interactions in vesicular transport	C2	Genetic Information Processing	Folding, sorting and degradation	2.9E-06	0.76
RNA degradation	C2, C3, C4	Genetic Information Processing	Folding, sorting and degradation	3.3E-06	0.71
Primary bile acid biosynthesis	C2	Metabolism	Lipid metabolism	3.4E-06	-0.76
Cell adhesion molecules (CAMs)	C2	Environmental Information Processing	Signaling molecules and interaction	4.3E-06	0.73
RIG-I-like receptor signaling pathway	C3	Organismal Systems	Immune system	8.6E-06	0.73
Peroxisome	C2, C3	Cellular Processes	Transport and catabolism	1.2E-05	-0.71
Steroid hormone biosynthesis	C2	Metabolism	Lipid metabolism	1.6E-05	-0.7

Mismatch repair	C2	Genetic Information Processing	Replication and repair	1.7E-05	0.69
Porphyrin and chlorophyll metabolism	C1	Metabolism	Metabolism of cofactors and vitamins	1.9E-05	-0.7
Axon guidance	C3	Organismal Systems	Development and regeneration	2.5E-05	0.68
Fatty acid degradation	C2	Metabolism	Lipid metabolism	3.6E-05	-0.67
Renin-angiotensin system	C1, C3, C4	Organismal Systems	Endocrine system	3.8E-05	0.67
Inositol phosphate metabolism	C1	Metabolism	Carbohydrate metabolism	3.9E-05	0.67
ErbB signaling pathway	C2	Environmental Information Processing	Signal transduction	5.2E-05	0.66
Notch signaling pathway	C1	Environmental Information Processing	Signal transduction	5.6E-05	0.65
Hedgehog signaling pathway	C4	Environmental Information Processing	Signal transduction	7.3E-05	0.65
ABC transporters	C2	Environmental Information Processing	Membrane transport	8.4E-05	-0.64
Metabolism of xenobiotics by cytochrome P450	C2	Metabolism	Xenobiotics biodegradation and metabolism	0.00021	-0.61
Terpenoid backbone biosynthesis	C1	Metabolism	Metabolism of terpenoids and polyketides	0.00034	-0.59
MAPK signaling pathway	C2, C4	Environmental Information Processing	Signal transduction	0.00038	0.58
Valine, leucine and isoleucine degradation	C4	Metabolism	Amino acid metabolism	0.00048	-0.57

Homologous recombination	C2	Genetic Information Processing	Replication and repair	0.00092	0.54
PLI vs. N&S					
Basal transcription factors	C2	Genetic Information Processing	Transcription	8.2E-17	1.55
Ribosome	C2	Genetic Information Processing	Translation	3.2E-16	-1.48
Ubiquitin mediated proteolysis	C2	Genetic Information Processing	Folding, sorting and degradation	7.5E-16	1.4
Glyoxylate and dicarboxylate metabolism	C1, C2	Metabolism	Carbohydrate metabolism	2.6E-14	-1.34
Oxidative phosphorylation	C1	Metabolism	Energy metabolism	3.9E-13	-1.33
RNA degradation	C2, C3, C4	Genetic Information Processing	Folding, sorting and degradation	4.7E-13	1.4
Apoptosis	C2	Cellular Processes	Cell growth and death	7E-13	1.25
Fructose and mannose metabolism	C1, C3	Metabolism	Carbohydrate metabolism	3E-12	-1.2
Other glycan degradation	C4	Metabolism	Glycan biosynthesis and metabolism	4.5E-12	-1.3
Non-homologous end-joining	C2	Genetic Information Processing	Replication and repair	4E-11	1.23
TGF-beta signaling pathway	C4	Environmental Information Processing	Signal transduction	2E-10	1.15
Glycosaminoglycan biosynthesis - keratan sulfate	C4	Metabolism	Glycan biosynthesis and metabolism	6E-09	-1.09
Sulfur metabolism	C1	Metabolism	Energy metabolism	4.1E-08	-1

Galactose metabolism	C1	Metabolism	Carbohydrate metabolism	4.2E-08	-1
Glycosphingolipid biosynthesis - globo and isoglobo series	C4	Metabolism	Glycan biosynthesis and metabolism	5.1E-08	-0.95
Valine, leucine and isoleucine biosynthesis	C4	Metabolism	Amino acid metabolism	1.1E-07	1.04
Alanine, aspartate and glutamate metabolism	C3, C4	Metabolism	Amino acid metabolism	6.3E-07	-0.93
PPAR signaling pathway	C2	Organismal Systems	Endocrine system	6.8E-07	-0.96
Glycosphingolipid biosynthesis - ganglio series	C4	Metabolism	Glycan biosynthesis and metabolism	8.6E-07	-0.97
Glycosaminoglycan degradation	C4	Metabolism	Glycan biosynthesis and metabolism	1.3E-06	-0.95
Gap junction	C2, C3, C4	Cellular Processes	Cellular community - eukaryotes	1.9E-06	0.84
Base excision repair	C2	Genetic Information Processing	Replication and repair	3.3E-06	-0.92
Pentose phosphate pathway	C1, C2	Metabolism	Carbohydrate metabolism	5E-06	-0.82
Glutathione metabolism	C1	Metabolism	Metabolism of other amino acids	5E-06	-0.9
Glycine, serine and threonine metabolism	C2	Metabolism	Amino acid metabolism	5E-06	-0.83
Proteasome	C2	Genetic Information Processing	Folding, sorting and degradation	1.4E-05	-0.82
Phenylalanine metabolism	C2	Metabolism	Amino acid metabolism	2.4E-05	-0.81
Tyrosine metabolism	C4	Metabolism	Amino acid metabolism	2.4E-05	-0.78

Adherens junction	C2	Cellular Processes	Cellular community - eukaryotes	3.3E-05	0.77
Cell cycle	C2	Cellular Processes	Cell growth and death	3.8E-05	0.77
Glycerophospholipid metabolism	C2	Metabolism	Lipid metabolism	3.8E-05	-0.82
Glycerolipid metabolism	C2	Metabolism	Lipid metabolism	5.3E-05	-0.8
Mismatch repair	C2	Genetic Information Processing	Replication and repair	5.3E-05	0.82
Lysosome	C2	Cellular Processes	Transport and catabolism	5.3E-05	-0.83
ECM-receptor interaction	C2	Environmental Information Processing	Signaling molecules and interaction	7.3E-05	-0.8
Jak-STAT signaling pathway	C3	Environmental Information Processing	Signal transduction	0.00026	0.7
Amino sugar and nucleotide sugar metabolism	C1, C2, C3	Metabolism	Carbohydrate metabolism	0.00031	-0.66
Pyrimidine metabolism	C1, C2	Metabolism	Nucleotide metabolism	0.00056	-0.66
Fc epsilon RI signaling pathway	C3	Organismal Systems	Immune system	0.00056	0.66
Renin-angiotensin system	C1, C3, C4	Organismal Systems	Endocrine system	0.00068	0.7
Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	C4	Metabolism	Glycan biosynthesis and metabolism	0.0007	-0.7
Toll-like receptor signaling pathway	C3	Organismal Systems	Immune system	0.00071	0.66
Citrate cycle (TCA cycle)	C1	Metabolism	Carbohydrate metabolism	0.00077	-0.64
Wnt signaling pathway	C2, C3, C4	Environmental Information Processing	Signal transduction	0.00088	0.6

PF vs. PLI					
Regulation of actin cytoskeleton	C1, C4	Cellular Processes	Cell motility	4.2E-09	1.24
Focal adhesion	C2	Cellular Processes	Cellular community - eukaryotes	1E-06	1.09
Cell adhesion molecules (CAMs)	C2	Environmental Information Processing	Signaling molecules and interaction	1E-06	1.08
T cell receptor signaling pathway	C3	Organismal Systems	Immune system	1.1E-05	0.88
ECM-receptor interaction	C2	Environmental Information Processing	Signaling molecules and interaction	1.5E-05	0.96
Histidine metabolism	C1	Metabolism	Amino acid metabolism	1.5E-05	-0.91
Wnt signaling pathway	C2, C3, C4	Environmental Information Processing	Signal transduction	3.1E-05	0.82
Chemokine signaling pathway	C3	Organismal Systems	Immune system	4.5E-05	0.87
Notch signaling pathway	C1	Environmental Information Processing	Signal transduction	5.1E-05	0.89
Fc gamma R-mediated phagocytosis	C3	Organismal Systems	Immune system	6.9E-05	0.83
Intestinal immune network for IgA production	C3	Organismal Systems	Immune system	7.9E-05	0.85
B cell receptor signaling pathway	C3	Organismal Systems	Immune system	7.9E-05	0.81
Hematopoietic cell lineage	C3	Organismal Systems	Immune system	9.9E-05	0.86
Axon guidance	C3	Organismal Systems	Development and regeneration	0.00011	0.84
Amino sugar and nucleotide sugar metabolism	C1, C2, C3	Metabolism	Carbohydrate metabolism	0.00012	-0.74

Leukocyte transendothelial migration	C3	Organismal Systems	Immune system	0.00013	0.79
Natural killer cell mediated cytotoxicity	C3	Organismal Systems	Immune system	0.00021	0.76
Antigen processing and presentation	C3	Organismal Systems	Immune system	0.00022	0.79
Selenocompound metabolism	C1	Metabolism	Metabolism of other amino acids	0.00032	-0.77
Gap junction	C2, C3, C4	Cellular Processes	Cellular community - eukaryotes	0.00036	0.67
Valine, leucine and isoleucine degradation	C4	Metabolism	Amino acid metabolism	0.00037	-0.77
Pantothenate and CoA biosynthesis	C1	Metabolism	Metabolism of cofactors and vitamins	0.00051	-0.76
Melanogenesis	C1, C3	Organismal Systems	Endocrine system	0.00053	0.7
RNA degradation	C2, C3, C4	Genetic Information Processing	Folding, sorting and degradation	0.00053	-0.69
VEGF signaling pathway	C2	Environmental Information Processing	Signal transduction	0.0006	0.71
Steroid hormone biosynthesis	C2	Metabolism	Lipid metabolism	0.00063	-0.74
Glycosaminoglycan biosynthesis - heparan sulfate / heparin	C4	Metabolism	Glycan biosynthesis and metabolism	0.00063	-0.75
Nicotinate and nicotinamide metabolism	C1	Metabolism	Metabolism of cofactors and vitamins	0.00069	-0.74
Starch and sucrose metabolism	C1	Metabolism	Carbohydrate metabolism	0.00089	-0.72
Arginine and proline metabolism	C3	Metabolism	Amino acid metabolism	0.00089	-0.67

Peroxisome	C2, C3	Cellular Processes	Transport and catabolism	0.00094	-0.71

Appendix C.3 Predicted drugs for 12 input signatures using CMap

The drugs (approved) are ranked in descending order based on CMAP score. The 1st rank represents the drug with the smallest CMap score, the 2nd the 2nd most, etc. Category (C1: Insulin Resistance and Oxidative Stress, C2: Cell Stress, Apoptosis and Lipotoxicity, C3: Inflammation, C4: Fibrosis)

DrugBank ID	Drug Name	Targets	Comparison	Category	Rank
DB02546	Vorinostat	HDAC1|HDAC2|HDAC3|HDAC6|HDAC8|acuC1	PF vs. N&S	C4	1
			PF vs. N&S	C2	2
			PF vs. PLI	C2	4
			PF vs. PLI	C3	5
			PF vs. N&S	C1	8
DB00947	Fulvestrant	ESR1	PF vs. N&S	C2	1
			PLI vs. N&S	C1	2
			PF vs. N&S	C1	4
			PLI vs. N&S	C2	4
			PLI vs. N&S	C4	10
DB01251	Gliquidone	ABCC8|KCNJ8	PF vs. PLI	C1	1
			PF vs. PLI	C2	1
			PF vs. PLI	C3	1
			PF vs. PLI	C4	1
DB01259	Lapatinib	EGFR|ERBB2	PF vs. PLI	C1	2
			PF vs. PLI	C3	2
			PF vs. PLI	C2	3
			PF vs. PLI	C4	7
DB00317	Gefitinib	EGFR	PF vs. PLI	C3	4
			PF vs. PLI	C1	5
			PF vs. PLI	C4	5
			PLI vs. N&S	C1	7

DB06774	Capsaicin	TRPV1|PHB2	PF vs. PLI	C2	2
			PF vs. N&S	C3	6
			PF vs. PLI	C4	6
			PF vs. N&S	C2	10
DB11672	Curcumin	PPARG|VDR|ABCC5|CBR1|GSTP1	PF vs. PLI	C2	7
			PF vs. PLI	C2	8
			PF vs. PLI	C3	8
			PF vs. PLI	C1	10
DB01590	Everolimus	MTOR	PF vs. N&S	C1	1
			PF vs. N&S	C3	1
			PF vs. N&S	C4	2
DB00390	Digoxin	ATP1A1	PF vs. N&S	C3	2
			PF vs. N&S	C4	4
			PF vs. N&S	C2	6
DB00288	Amcinonide	NR3C1|ANXA1	PF vs. PLI	C1	3
			PF vs. PLI	C4	4
			PF vs. PLI	C2	6
DB01175	Escitalopram	SLC6A4|SLC6A3|SLC6A2|ADRA1A|CHRM1|HRH1	PF vs. N&S	C1	2
			PLI vs. N&S	C1	5
			PLI vs. N&S	C2	10
DB00920	Ketotifen	HRH1|PGD	PF vs. N&S	C1	3
			PF vs. PLI	C1	6
			PF vs. N&S	C2	9
DB00420	Promazine	DRD2|HTR2A|HTR2C|DRD1|DRD4|CHRM4|CHRM2|CHRM3|CHRM1|ADRA1B|CHRM5|ADRA1A|HRH1|ADRA1D	PF vs. PLI	C1	4
			PF vs. PLI	C4	8
			PF vs. PLI	C2	10
DB01380	Cortisone acetate	NR3C1	PF vs. PLI	C3	6
			PF vs. PLI	C2	9
			PF vs. PLI	C4	10
DB00850	Perphenazine	DRD2|DRD1|CALM1	PLI vs. N&S	C4	8
			PLI vs. N&S	C1	10

			PLI vs. N&S	C3	10
DB01183	Naloxone	OPRM1|OPRD1|OPRK1|CREB1|ESR1|TLR4|CES1	PLI vs. N&S	C4	2
			PLI vs. N&S	C1	4
DB09167	Dosulepin	HTR1A|HTR2A|HRH1|CHRM1|CHRM2|CHRM3|CHRM4|CHRM5|ADRA2A|ADRA1A|SLC6A2|SLC6A4	PLI vs. N&S	C2	2
			PLI vs. N&S	C3	4
DB01396	Digitoxin	ATP1A1	PLI vs. N&S	C3	1
			PF vs. N&S	C4	7
DB00270	Isradipine	CACNA1C|CACNA2D1|CACNB2|CACNA1H|CACNA2D2|CACNA1D|CACNA1S	PF vs. PLI	C3	3
			PF vs. PLI	C2	5
DB00283	Clemastine	HRH1	PLI vs. N&S	C3	2
			PLI vs. N&S	C4	6
DB01623	Thiothixene	DRD2|DRD1|HTR2A	PLI vs. N&S	C2	3
			PF vs. N&S	C1	7
DB00933	Mesoridazine	HTR2A|DRD2	PF vs. N&S	C3	5
			PF vs. N&S	C4	5
DB04910	Oxibendazole	TUBB4B	PLI vs. N&S	C4	4
			PLI vs. N&S	C3	6
DB00539	Toremifene	ESR1|SHBG	PF vs. N&S	C4	3
			PF vs. PLI	C1	8
DB00136	Calcitriol	VDR|HOXA10	PF vs. N&S	C4	6
			PLI vs. N&S	C4	7
DB00737	Meclizine	HRH1|NR1I3	PF vs. N&S	C2	4
			PF vs. N&S	C1	10

DB00613	Amodiaquine	HNMT	PLI vs. N&S	C1	6
			PLI vs. N&S	C4	9
DB04946	Iloperidone	HTR2A|DRD2|DRD1|DRD3|DRD4|HTR1A|HTR6|HTR7|ADRA1A|HRH1|ADRA2C	PF vs. N&S	C2	7
			PLI vs. N&S	C2	9
DB00783	Estradiol	ESR1|ESR2|NR1I2|CHRNA4|NCOA2|GPER1|MT-ATP6|BECN1|HSD17B2|ESRRG	PLI vs. N&S	C2	8
			PLI vs. N&S	C1	9
DB00836	Loperamide	OPRM1|OPRD1|OPRK1|CACNA1A|POMC|CALM1|NR1I3	PF vs. N&S	C3	9
			PLI vs. N&S	C3	9
DB01392	Yohimbine	ADRA2A|ADRA2B|ADRA2C|HTR1A|HTR1B|HTR1D|DRD2|DRD3|HTR2A|HTR2C|KCNJ1|HTR2B	PLI vs. N&S	C1	1
DB00711	Diethylcarbamazine	ALOX5|PTGS1	PLI vs. N&S	C2	1
DB01018	Guanfacine	ADRA2A|ADRA2B	PLI vs. N&S	C4	1
DB01357	Mestranol	ESR1	PF vs. PLI	C4	2
DB00670	Pirenzepine	CHRM1	PF vs. N&S	C2	3
DB00768	Olopatadine	HRH1|HRH2|HRH3|S100A1|S100A12|S100B|S100A13|S100A2	PF vs. N&S	C3	3
DB01406	Danazol	ESR1|AR|PGR|GNRHR|GNRHR2|CCL2	PF vs. PLI	C4	3
DB00441	Gemcitabine	RRM1|TYMS|CMPK1	PLI vs. N&S	C1	3
DB01193	Acebutolol	ADRB1|ADRB2	PLI vs. N&S	C3	3
DB06786	Halcinonide	SMO	PLI vs. N&S	C4	3
DB00630	Alendronic acid	FDPS|PTPN4|PTPRS|PTPRE|ATP6V1A	PF vs. N&S	C3	4
DB00796	Candesartan cilexetil	AGTR1	PF vs. N&S	C1	5

DB00246	Ziprasidone	DRD2|DRD1|DRD5|HTR2A|DRD3|DRD4|HTR1A|HTR1B|HTR1D|HTR1E|HTR2C|HTR3A|HTR6|HTR7|HRH1|ADRA1A|ADRA1B|ADRA2A|ADRA2B|ADRA2C|CHRM1|CHRM2|CHRM3|CHRM4|CHRM5	PF vs. N&S	C2	5
DB00482	Celecoxib	PTGS2|PDPK1|CA2|CA3|ABCB5|ABCG2|ABCB1	PLI vs. N&S	C2	5
DB00594	Amiloride	SCNN1A|SCNN1B|SCNN1G|SCNN1D|AOC1|ASIC2|ASIC1|SLC9A1|PLAU	PLI vs. N&S	C3	5
DB00656	Trazodone	HTR2A|HTR2C|SLC6A4|HTR1A|HRH1|ADRA1A|ADRA2A	PLI vs. N&S	C4	5
DB00890	Dienestrol	ESR1|SHBG	PF vs. N&S	C1	6
DB00928	Azacitidine	DNMT1	PLI vs. N&S	C2	6
DB01118	Amiodarone	KCNH2|ADRB1|CACNA1H|CACNA2D2|THRA|THRB|PPARG	PF vs. N&S	C3	7
DB00584	Enalapril	ACE	PF vs. PLI	C1	7
DB00807	Proparacaine	SCN10A	PF vs. PLI	C3	7
DB00881	Quinapril	ACE	PLI vs. N&S	C2	7
DB00960	Pindolol	ADRB1|ADRB2|HTR1A|HTR1B|ADRB3	PLI vs. N&S	C3	7
DB06228	Rivaroxaban	F10	PF vs. N&S	C2	8
DB02789	Pregnenolone	SULT2B1|NR1I2	PF vs. N&S	C3	8
DB01179	Podofilox	TOP2A|TUBA4A|TUBB	PF vs. N&S	C4	8
DB00585	Nizatidine	HRH2	PLI vs. N&S	C1	8
DB00458	Imipramine	SLC6A2|HTR2A|SLC6A4|HRH1|ADRA1A|ADRA1D|CHRM1|CHRM2|CHRM3|CHRM4|CHRM5|KCND2|KCND3|HTR2C|ADRA1B|HTR7|DRD1|DRD2|KCNH2|SLC6A3|HTR1A|HTR6|KCNH1|ORM2	PLI vs. N&S	C3	8
DB00748	Carbinoxamine	HRH1	PF vs. N&S	C1	9
DB00903	Etacrynic acid	ATP1A1|SLC12A1|LEF1|GSTP1	PF vs. N&S	C4	9
DB00834	Mifepristone	PGR|NR3C1|KLK3|NR1I2	PF vs. PLI	C1	9

DB01138	Sulfinpyrazone	ABCC2|ABCC1|SLC22A12|NR1I2	PF vs. PLI	C3	9
DB00481	Raloxifene	ESR1|ESR2|SERPINB9|TFF1	PF vs. PLI	C4	9
DB01438	Phenazopyridine	SCN1A	PF vs. N&S	C3	10
DB09242	Moxonidine	ADRA2A|NISCH	PF vs. N&S	C4	10
DB09074	Olaparib	PARP1|PARP2|PARP3	PF vs. PLI	C3	10

Appendix C.4 Predicted small molecules for 12 input signatures using CMap

The small molecules (not approved) are ranked in descending order based on CMAP score. The 1st rank represents the drug with the smallest CMap score, the 2nd the 2nd most, etc. Category (C1: Insulin Resistance and Oxidative Stress, C2: Cell Stress, Apoptosis and Lipotoxicity, C3: Inflammation, C4: Fibrosis)

DrugBank ID	Drug Name	Targets	Comparison	Category	Rank
DB08597	6-[4-(2-piperidin-1-ylethoxy)phenyl]-3-pyridin-4-ylpyrazolo[1,5-a]pyrimidine	ACVR1|FKBP1A	PF vs. N&S	C1	1
			PLI vs. N&S	C1	1
			PF vs. N&S	C2	1
			PF vs. N&S	C3	1
			PLI vs. N&S	C3	1
			PLI vs. N&S	C4	1
			PLI vs. N&S	C2	8
DB08059	Wortmannin	PIK3CG|PLK1|PIK3R1|PIK3CA	PLI vs. N&S	C3	3
			PF vs. PLI	C4	3
			PF vs. PLI	C3	5
			PF vs. N&S	C1	7
			PLI vs. N&S	C1	7
			PF vs. N&S	C2	9
			PLI vs. N&S	C4	10

DB08607	(5R)-5-(4-{[(2R)-6-HYDROXY-2,5,7,8-TETRAMETHYL-3,4-DIHYDRO-2H-CHROMEN-2-YL]METHOXY}BENZYL)-1,3-THIAZOLIDINE-2,4-DIONE	CYP2C8	PF vs. PLI	C2	1
			PF vs. PLI	C3	1
			PF vs. PLI	C1	2
			PF vs. PLI	C4	2
			PF vs. N&S	C4	4
			PLI vs. N&S	C4	9
DB02424	Geldanamycin	HSP90AB1|HSP90AA1|HSP90B1	PF vs. N&S	C3	3
			PLI vs. N&S	C4	3
			PF vs. N&S	C1	4
			PLI vs. N&S	C3	6
			PF vs. N&S	C4	6
DB04297	7-[4-(Dimethylamino)Phenyl]-N-Hydroxy-4,6-Dimethyl-7-Oxo-2,4-Heptadienamide	HDAC8|acuC1|HDAC7	PF vs. N&S	C4	2
			PF vs. PLI	C2	5
			PF vs. N&S	C1	6
			PF vs. PLI	C2	7
			PF vs. PLI	C3	9
DB02656	LY-294002	PIM1|PIK3CG	PF vs. PLI	C4	5
			PF vs. PLI	C1	6
			PF vs. PLI	C3	7
			PLI vs. N&S	C3	8
			PF vs. N&S	C4	9
DB07863	2-chloro-5-nitro-N-phenylbenzamide	NCOA2|PPARG|RXRA	PF vs. N&S	C1	2
			PLI vs. N&S	C1	4
			PF vs. N&S	C2	6
			PF vs. N&S	C4	7
DB12445	Nitroaspirin	PTGS1	PF vs. PLI	C2	2
			PF vs. PLI	C3	3
			PF vs. PLI	C1	7
			PF vs. PLI	C4	8
DB00466	Picrotoxin		PF vs. PLI	C2	3

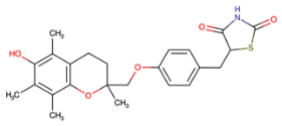
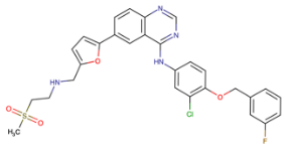
		GABRR1|GABRA1|GLRA2|GLRA3|GLRA1	PF vs. PLI	C3	4
			PF vs. PLI	C1	8
			PF vs. PLI	C4	9
DB08142	AT-7519	CDK2|CDK1	PF vs. PLI	C2	8
			PF vs. PLI	C3	8
			PF vs. PLI	C1	9
			PLI vs. N&S	C2	10
DB05022	Amonafide	TOP2A|TOP2B	PF vs. N&S	C4	1
			PF vs. N&S	C3	2
			PF vs. N&S	C1	3
DB02932	(R)-Bicalutamide	AR	PF vs. PLI	C1	1
			PF vs. PLI	C4	1
			PF vs. PLI	C2	4
DB03701	Vanoxerine	SLC6A3	PLI vs. N&S	C4	2
			PLI vs. N&S	C3	4
			PLI vs. N&S	C1	5
DB04581	1-benzylimidazole	QPCT	PF vs. PLI	C1	4
			PF vs. PLI	C4	7
			PF vs. PLI	C2	10
DB03496	Alvocidib	CDK2|CDK5|CDK9|CDK1|CDK6|EGFR|CDK4|CDK8|CDK7|PYGM|PYGB|PYGL	PLI vs. N&S	C2	6
			PF vs. PLI	C4	6
			PF vs. N&S	C2	10
DB04017	Clorgiline	MAOA	PLI vs. N&S	C2	1
			PLI vs. N&S	C1	3
DB03467	Naringenin	ttgR|ESR1|AKR1C1|CYP1B1|KANSL3|SHBG|CYP19A1|ESR2	PLI vs. N&S	C2	2
			PF vs. N&S	C2	3
DB07859	4-(4-CHLOROPHENYL)-4-[4-(1H-PYRAZOL-4-YL)PHENYL]PIPERIDINE	PRKACA|PKIA|AKT2|GSK3B	PF vs. PLI	C1	3
			PF vs. PLI	C4	4
DB03783	Phenacetin	PTGS1	PF vs. N&S	C2	5

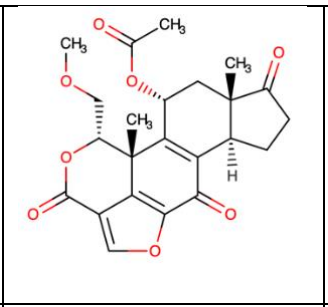
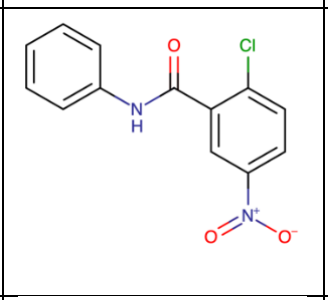
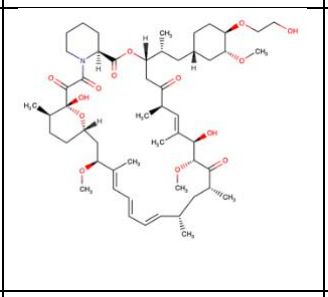
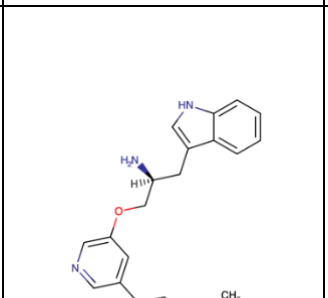
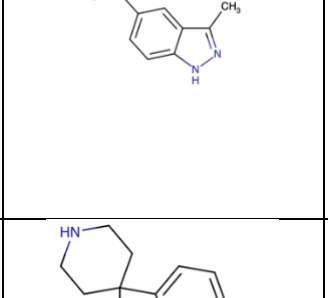
			PLI vs. N&S	C2	5
DB02860	Calyculin A	PPP1CC	PF vs. N&S	C2	2
			PLI vs. N&S	C2	9
DB13877	Iniparib	PARP1	PLI vs. N&S	C3	2
			PF vs. N&S	C4	10
DB12116	Epigallocatechin gallate	AHR|DNMT1|DHFRL1	PLI vs. N&S	C4	5
			PLI vs. N&S	C3	7
DB02587	Colforsin	ADCY2|GNAS|ADCY5|CFTR	PF vs. N&S	C4	3
			PF vs. PLI	C1	10
DB08784	2-(4-CHLORO-PHENYLAMINO)-NICOTINIC ACID	NMRAL1	PLI vs. N&S	C2	3
			PLI vs. N&S	C1	10
DB13061	MLN8054	AURKA	PF vs. N&S	C2	4
			PF vs. N&S	C1	10
DB06075	Linsitinib	IGF1R	PF vs. N&S	C3	6
			PF vs. N&S	C4	8
DB04348	Taurocholic Acid	CEL|FABP6|NR1H4	PF vs. PLI	C1	5
			PF vs. PLI	C4	10
DB14061	Hycanthone	ABCB1	PLI vs. N&S	C4	6
			PF vs. PLI	C2	9
DB12742	Amuvatinib	KIT|MET|RET|PDGFRA|FLT3|RAD51	PF vs. N&S	C1	8
			PF vs. N&S	C3	9
DB08435	(5E,14E)-11-oxoprosta-5,9,12,14-tetraen-1-oic acid	PPARG	PF vs. PLI	C3	2
DB04175	Mdl-29951	FBP1	PLI vs. N&S	C1	2
DB11582	Thiocolchicoside	GABRA1|GLRA1|TNFSF11	PF vs. N&S	C3	4
DB01103	Quinacrine	PLA2G6|PLA2G4A|PLCL1	PLI vs. N&S	C2	4
DB07697	1-(2,3-dihydro-1,4-benzodioxin-6-ylsulfonyl)-4-[(4-methoxyphen	PKM	PLI vs. N&S	C4	4

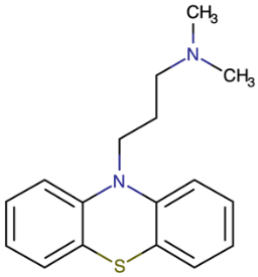
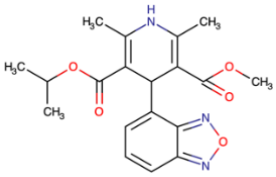
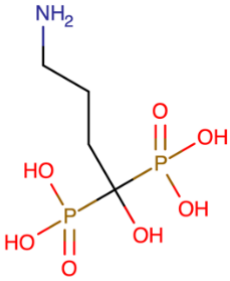
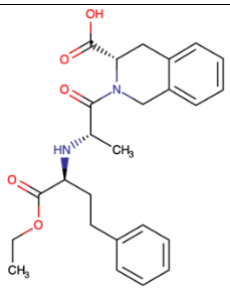
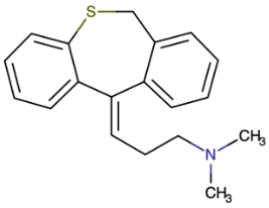
	yl)sulfonyl]piperazine				
DB12518	Raclopride	DRD2	PF vs. N&S	C1	5
DB05913	OSI-930	KIT|FLT1	PF vs. N&S	C3	5
DB02665	(1R,2S)-2-Phenylcyclopropanaminium	PRSS1	PF vs. N&S	C4	5
DB08073	(2S)-1-(1H-INDOL-3-YL)-3-{[5-(3-METHYL-1H-INDAZOL-5-YL)PYRIDIN-3-YL]OXY}PROPAN-2-AMINE	AKT2|GSK3B|PRKACA|PKIA	PLI vs. N&S	C3	5
DB04690	Camptothecin	TOP1	PF vs. PLI	C2	6
DB08167	Methylthioninium	ACHE	PF vs. PLI	C3	6
DB12200	Tivantinib	MET	PLI vs. N&S	C1	6
DB13520	Metergoline	SCN2A	PF vs. N&S	C2	7
DB07129	(2R)-1-(2,6-dimethylphenoxy)propan-2-amine	PLAU	PF vs. N&S	C3	7
DB08348	N~2~,N~2~-DIMETHYL-N~1~-(6-OXO-5,6-DIHYDROPHENANTHRIDIN-2-YL)GLYCINAMIDE	eta|EEF2|chxA|PARP3|PARP15	PLI vs. N&S	C2	7
DB09186	Nisoxetine	SLC6A4	PLI vs. N&S	C4	7
DB12191	Obatoclax	BCL2	PF vs. N&S	C2	8
DB11781	Tosedostat	NPEPPS|LTA4H	PF vs. N&S	C3	8
DB04149	(R)-Rolipram	PDE4B|PDE4D	PLI vs. N&S	C1	8
DB06393	Xaliproden	HTR1A	PLI vs. N&S	C4	8
DB12693	Ritanserin	HTR2A	PF vs. N&S	C1	9

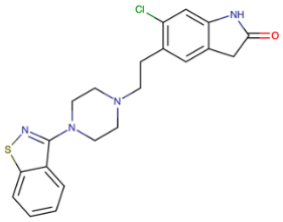
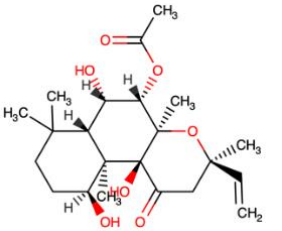
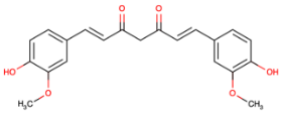
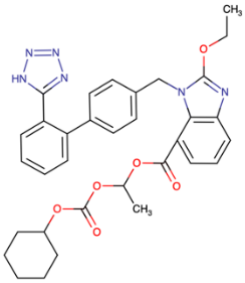
DB04513	N-(6-Aminohexyl)-5-Chloro-1-Naphthalenesulfonamide	CALM1|TNNC1|TNNI3	PLI vs. N&S	C1	9
DB05134	Tanespimycin	HSP90AA1|HSP90AB1	PLI vs. N&S	C3	9
DB03880	Batimastat	MMP8|MMP12|MMP16|ADAM28|ADAMTS5	PF vs. N&S	C3	10
DB05482	7-ethyl-10-hydroxycamptothecin	TOP1	PF vs. PLI	C3	10
DB08437	Puromycin	RPL10L|RPL13A|RPL23|RPL15|RPL19|RPL23A|RSL24D1|RPL26L1|RPL8|RPL37|RPL3|RPL11|aat	PLI vs. N&S	C3	10

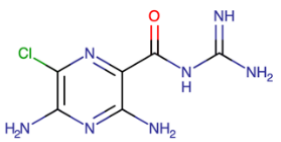
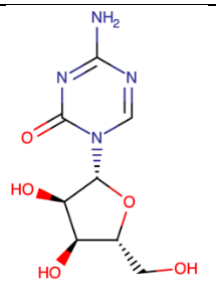
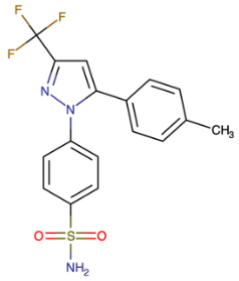
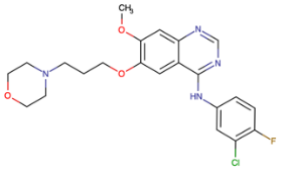
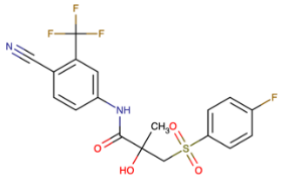
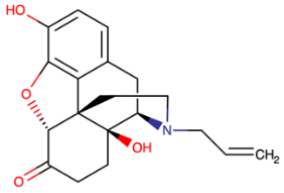
Appendix C.5 Ranking of the 49 drugs predicted by CMap using network proximity z-score

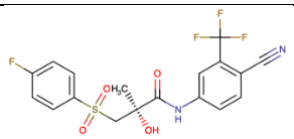
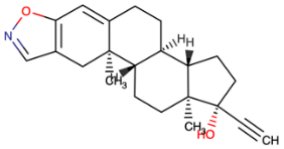
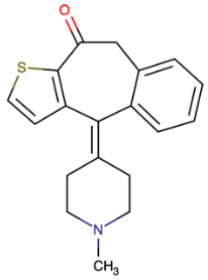
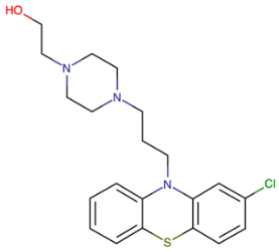
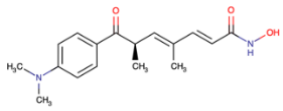
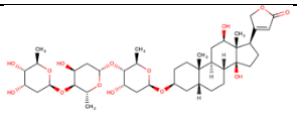
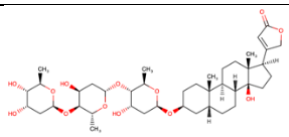
Rank	Drug	DrugBank ID	Known Targets	2D Structure	Z-score
1	Troglitazone	DB00197	PPARG, SERPINE1, ACSL4, SLC29A1, ESRRG, ESRRA, PPARD, PPARA, GSTP1		-3.60
2	Lapatinib	DB01259	ERBB2, EGFR		-2.58

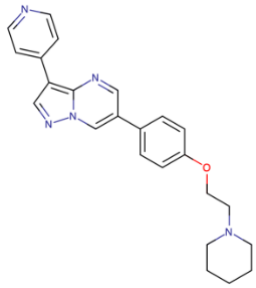
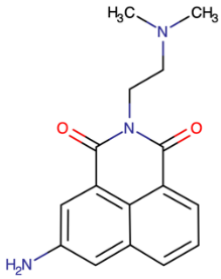
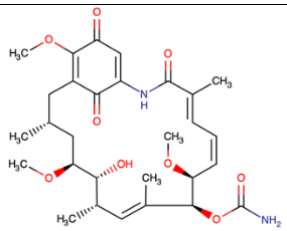
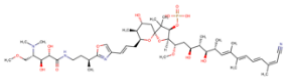
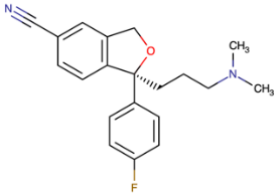
3	Wortmannin	DB08059	PIK3CG, PLK1, PIK3R1, PIK3CA		-2.32
4	2-chloro-5-nitro-N-phenylbenzamide	DB07863	NCOA2, PPARG, RXRA		-2.04
5	Everolimus	DB01590	MTOR		-1.95
6	(2S)-1-(1H-INDOL-3-YL)-3-{[5-(3-METHYL-1H-INDAZOL-5-YL)PYRIDIN-3-YL]OXY}PROPAN-2-AMINE	DB08073	AKT2, GSK3B, PRKACA, PKIA		-1.70
7	4-(4-CHLOROPHENYL)-4-[4-(1H-PYRAZOL-4-	DB07859	PRKACA, PKIA, AKT2, GSK3B		-1.70

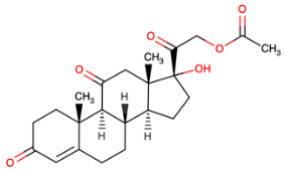
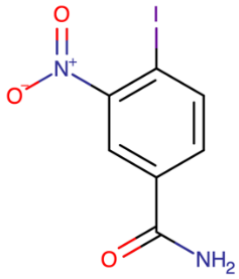
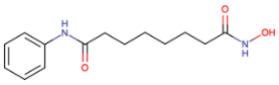
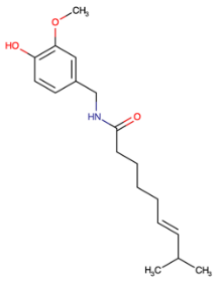
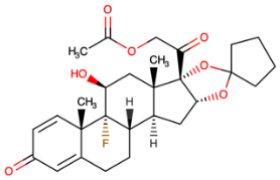
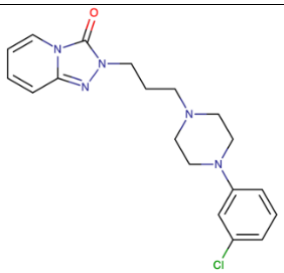
	YL)PHENYL]PIPERIDINE				
8	Promazine	DB00420	DRD2, HTR2A, HTR2C, DRD1, DRD4, CHRM4, CHRM2, CHRM3, CHRM1, ADRA1B, CHRM5, ADRA1A, HRH1, ADRA1D		-1.42
9	Isradipine	DB00270	CACNA1C, CACNA2D1, CACNB2, CACNA1H, CACNA2D2, CACNA1D, CACNA1S		-1.08
10	Alendronic acid	DB00630	FDPS, PTPN4, PTPRS, PTPRE, ATP6V1A		-1.01
11	Quinapril	DB00881	ACE		-0.86
12	Dosulepin	DB09167	HTR1A, HTR2A, HRH1, CHRM1, CHRM2, CHRM3, CHRM4, CHRM5, ADRA2A,		-0.76

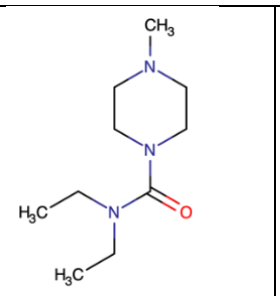
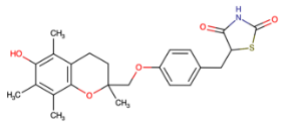
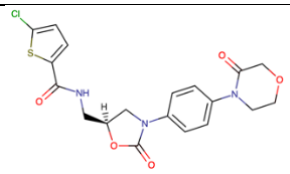
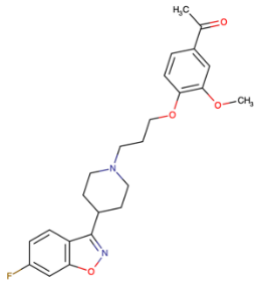
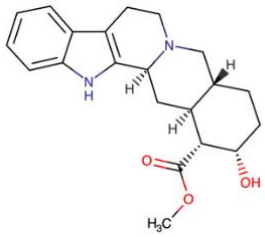
			ADRA2B, ADRA2C, ADRA1A, ADRA1B, ADRA1D, SLC6A2, SLC6A4		
13	Ziprasidone	DB00246	DRD2, DRD1, DRD5, HTR2A, DRD3, DRD4, HTR1A, HTR1B, HTR1D, HTR1E, HTR2C, HTR3A, HTR6, HTR7, HRH1, ADRA1A, ADRA1B, ADRA2A, ADRA2B, ADRA2C, CHRM1, CHRM2, CHRM3, CHRM4, CHRM5		-0.76
14	Colforsin	DB02587	ADCY2, GNAS, ADCY5, CFTR		-0.47
15	Curcumin	DB11672	PPARG, VDR, ABCC5, CBR1, GSTP1		-0.45
16	Candesartan cilexetil	DB00796	AGTR1		-0.13

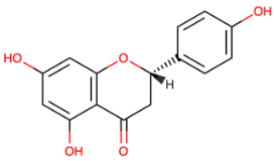
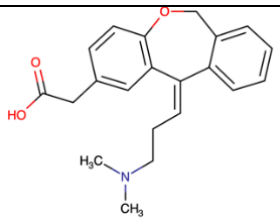
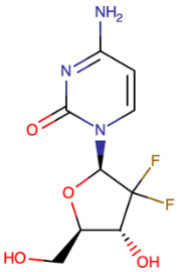
17	Amiloride	DB00594	SCNN1A, SCNN1B, SCNN1G, SCNN1D, AOC1, ASIC2, ASIC1, SLC9A1, PLAU		-0.09
18	Azacitidine	DB00928	DNMT1		-0.05
19	Celecoxib	DB00482	PTGS2, PDPK1, CA2, CA3, CDH11		-0.03
20	Gefitinib	DB00317	EGFR		0.00
21	Bicalutamide	DB01128	AR		0.00
22	Naloxone	DB01183	OPRM1, OPRD1, OPRK1, CREB1, ESR1, TLR4, CES1		0.00

23	(R)-Bicalutamide	DB02932	AR		0.00
24	Danazol	DB01406	ESR1, AR, PGR, GNRHR, GNRHR2, CCL2		0.00
25	Ketotifen	DB00920	HRH1, PGD		0.00
26	Perphenazine	DB00850	DRD2, DRD1, CALM1		0.00
27	Trichostatin A	DB04297	HDAC8, acuC1, HDAC7		0.25
28	Digoxin	DB00390	ATP1A1		0.31
29	Digitoxin	DB01396	ATP1A1		0.31

30	6-[4-(2-piperidin-1-ylethoxy)phenyl]-3-pyridin-4-ylpyrazolo[1,5-a]pyrimidine	DB08597	ACVR1, FKBP1A		0.36
31	Amonafide	DB05022	TOP2A, TOP2B		0.38
32	Geldanamycin	DB02424	HSP90AA1, HSP90AB1, HSP90B1		0.43
33	Calyculin A	DB02860	PPP1CC		0.44
34	Escitalopram	DB01175	SLC6A4, CHRM1, HRH1, HTR1A, HTR2A, ADRA1A, ADRA1B, ADRA1D, HTR2C, ADRA2A, ADRA2B, ADRA2C, DRD2, SLC6A2, SLC6A3		0.47

35	Cortisone acetate	DB01380	NR3C1		0.48
36	Iniparib	DB13877	PARP1		0.49
37	Vorinostat	DB02546	HDAC8, HDAC1, HDAC2, HDAC3, HDAC6, acuC1		0.51
38	Capsaicin	DB06774	TRPV1, PHB2		0.57
39	Amcinonide	DB00288	NR3C1, ANXA1		0.58
40	Trazodone	DB00656	HTR2A, HTR2C, SLC6A4, HTR1A, HRH1, ADRA1A, ADRA2A, Htr2c		0.72

41	Diethylcarbamazine	DB00711	ALOX5, PTGS1		0.77
42	Calcitriol	DB00136	VDR, HOXA10		0.80
43	Rivaroxaban	DB06228	F10		1.01
44	Iloperidone	DB04946	HTR2A, DRD2, DRD1, DRD3, DRD4, HTR1A, HTR6, HTR7, ADRA1A, HRH1, ADRA2C		1.04
45	Yohimbine	DB01392	ADRA2A, ADRA2C, ADRA2B, HTR1A, HTR1B, HTR1D, DRD2, DRD3, HTR2A, HTR2C, KCNJ1, KCNJ10, KCNJ11, KCNJ12, KCNJ14, KCNJ15, KCNJ8, HTR2B		1.09

46	Naringenin	DB03467	ttgR, ESR1, AKR1C1, CYP1B1, KANSL3, SHBG, CYP19A1, ESR2		1.61
47	Olopatadine	DB00768	HRH1, HRH2, HRH3, S100A1, S100A12, S100B, S100A13, S100A2		1.65
48	Gemcitabine	DB00441	RRM1, TYMS, CMPK1		1.89

Appendix C.6 Drug information of the selected 8 drugs for modulating steatosis

DrugBank ID	Drug Name	Steatosis-related pathways (targets)	CMap prediction	Network proximity
DB01590	Everolimus	PI3K-Akt signaling pathway (mTOR), Type II diabetes mellitus(mTOR), Insulin signaling pathway(mTOR)	C1, C3, C4	Z-score: -1.95 Rank: 5
Indication	Postmenopausal women with advanced hormone receptor-positive, HER2-negative breast cancer (advanced HR+BC), progressive neuroendocrine tumors of pancreatic origin (PNET) with unresectable, locally advanced or metastatic disease, advanced renal cell carcinoma (RCC), renal angiomyolipoma and tuberous sclerosis complex (TSC).			
MOA	Everolimus is a mTOR inhibitor that binds with high affinity to the FK506 binding protein-12 (FKBP-12), thereby forming a drug complex that inhibits the activation of mTOR. The result of everolimus inhibition of			

	mTOR is a reduction in cell proliferation, angiogenesis, and glucose uptake.
Adverse effects	Stomatitis, infections, asthenia, fatigue, cough, and diarrhea.
Contraindications	Hypersensitivity to everolimus, to other rapamycin derivatives, or to any of the excipients.
Notes	Everolimus inhibited hepatic lipid accumulation and improved metabolic parameters in a fast food induced mice model of NASH, however, inflammatory and fibrotic responses still exhibited despite the reduced hepatic steatosis (Love, et al., 2017). Everolimus is an approved immunosuppressant for liver transplantation (Yee and Tan, 2017). However, in patients the drug is shown to promote hyperglycemia (Xu, et al., 2016) and hyperlipidemia (Kasiske, et al., 2008).

DrugBank ID	Drug Name	Steatosis-related pathways (targets)	CMap prediction	Network proximity
DB00482	Celecoxib	Insulin signaling pathway (PDPK1), PI3K-Akt signaling pathway (PDPK1), TNF signaling pathway (PDPK1), PPAR signaling pathway (PDPK1)	C2	Z-score: -0.03 Rank: 19
Indication	Rheumatoid arthritis, osteoarthritis, ankylosing spondylitis, short-term pain, menstrual cramps			
MOA	Celecoxib is a selective noncompetitive inhibitor of cyclooxygenase-2 (COX-2) enzyme, which reduces the synthesis of metabolites that include prostaglandin E2 (PGE2), prostacyclin (PGI2), thromboxane (TXA2), prostaglandin D2 (PGD2), and prostaglandin F2 (PGF2). Resultant inhibition of these mediators leads to the alleviation of pain and inflammation. (Gong, et al.) Celecoxib exerts anticancer effects by binding to the cadherin-11 (CDH11) protein. (Zhu, et al.) In addition, celecoxib has been found to inhibit carbonic anhydrase enzymes 2 and 3, further enhancing its anticancer effects. (Nishimori, et al.; Weber, et al.) Celecoxib may cause an increased risk of thrombotic events. The risk of thrombosis resulting from COX-2 inhibition is caused by the vasoconstricting actions of thromboxane A2, leading to enhanced platelet aggregation.			

Adverse effects	Cardiovascular risk, gastrointestinal risk, renal effects, advanced renal disease, anaphylactoid reactions, skin reactions
Contraindications	Hypersensitivity to celecoxib, patients who have demonstrated allergic-type reactions to sulfonamides, patients who have experienced asthma, urticaria, or allergic-type reactions after taking aspirin or other NSAIDs, treatment of peri-operative pain in the setting of coronary artery bypass graft (CABG) surgery.
Notes	Celecoxib attenuates liver steatosis and inflammation in NAFLD in a rat model (Chen, et al., 2011). Celecoxib partially restores autophagic flux via downregulation of COX-2 and alleviates steatosis in vitro and in vivo (Liu, et al., 2018). Aspirin is a non-selective COX inhibitor which was shown to be protective of NAFLD progression in a retrospective study (Simon, et al., 2019).

DrugBank ID	Drug Name	Steatosis-related pathways (targets)	CMap prediction	Network proximity
DB09167	Dosulepin	PI3K-Akt signaling pathway (CHRM1, CHRM2)	C2, C3	Z-score: -0.76 Rank: 12
Indication	Depressive illness			
MOA	By binding to noradrenaline transporter (NAT) and serotonin transporter (SERT) in an equipotent manner and inhibiting the reuptake activity, dosulepin increases the free levels of noradrenaline and 5HT at the synaptic cleft. Dosulepin displays affinity towards α2-adrenoceptors and to a lesser extent, α1-adrenoceptors (Gillman). Inhibition of presynaptic α2-adrenoceptors by dosulepin facilitates noradrenaline release and further potentiates the antidepressant effects. It also downregulates central β-adrenoceptors by causing a decline in the number of receptors and reduces noradrenaline-induced cyclic AMP formation. Dosulepin binds to 5HT1A and 5HT2A receptors in the cerebral cortex and hippocampus as an antagonist. 5HT1A receptors are autoreceptors that inhibit 5HT release and 5HT2A receptors are Gi/Go-coupled receptors that reduces dopamine release upon activation. Antagonism at 5HT2A receptors may also improve sleep patterns. Dosulepin also binds to muscarinic acetylcholine receptors and causes antimuscarinic side effects such as dry mouth. By acting as an antagonist at histamine type 1 (H1) receptors, dosulepin mediates a sedative effect.			

Adverse effects	central nervous system effects, anticholinergic effects, cardiovascular effects, gastrointestinal system and blurred vision.
Contraindications	Epilespsy, TCAs should not be used concomitantly or within 14 days of treatment with monoamine oxidase inhibitors, acute recovery phase, liver failure, hypersensitivity to dosulepin
Notes	Tricyclic antidepressants (TCA) are not as commonly used due to safety concerns (Peretti, et al., 2000). The association between this class of drug and liver injury is weak (Cosmin Sebastian Voican, et al., 2014). One study found that another TCA, Amineptine, to cause microvesicular steatosis in mice (Le Dinh, et al., 1988). However, a recent retrospective study found that TCA use decreased fibrosis progression in hepatitis C patients (Chen, et al., 2018).

DrugBank ID	Drug Name	Steatosis-related pathways (targets)	CMap prediction	Network proximity
DB00881	Quinapril	Renin-angiotensin system (ACE)	C2	Z-score: -0.86 Rank: 11
Indication	Hypertension, heart failure			
MOA	Angiotensin II constricts coronary blood vessels and is positively inotropic, which under normal circumstances, would increase vascular resistance and oxygen consumption.4 This action can eventually lead to myocyte hypertrophy and vascular smooth muscle cell proliferation.4 Angiotensin II also stimulates production of plasminogen activator inhibitor-1 (PAI-1), increasing the risk of thrombosis.2 Quinaprilat prevents the conversion of angiotensin I to angiotensin II by inhibition of angiotensin converting enzyme, and also reduces the breakdown of bradykinin.1,2 Reduced levels of angiotensin II lead to lower levels of PAI-1, reducing the risk of thrombosis, especially after a myocardial infarction.2			
Adverse effects	Head and neck angioedema, intestinal angioedema, anaphylactoid reactions, hepatic failure (rare), hypotension (rare), neutropenia/agranulocytosis.			
Contraindications	Hypersensitive to quinapril and in patients with a history of angioedema related to previous treatment with an ACE inhibitor			
Notes	Quinapril reduces markers of vascular oxidative stress and may attenuate the progression of the pathophysiology seen in the metabolic			

	syndrome (Khan, et al., 2004). Quinapril treatment increases insulin-stimulated endothelial function and adiponectin gene expression in type 2 diabetes patients (Hermann, et al., 2006), can be considered in combination therapy.

DrugBank ID	Drug Name	Steatosis-related pathways (targets)	CMap prediction	Network proximity
DB00246	Ziprasidone	PI3K-Akt signaling pathway (CHRM1, CHRM2)	C2	Z-score: -0.76 Rank: 12
Indication	Oral form: schizophrenia, bipolar I disorder. Injectable formulation: acute agitation in schizophrenia.			
MOA	Ziprasidone binds to serotonin-2A (5-HT2A) and dopamine D2 receptors, with a higher 5-HT2A/D2 receptor affinity ratio when compared to other antipsychotics. (Stahl and Shayegan) Ziprasidone offers enhanced modulation of mood, notable negative symptom relief, overall cognitive improvement and reduced motor dysfunction which is linked to it's potent interaction with 5-HT2C, 5-HT1D, and 5-HT1A receptors in brain tissue. (Stahl and Shayegan) Ziprasidone can bind moderately to norepinephrine and serotonin reuptake sites which may contribute to its antidepressant and anxiolytic activity. (Stahl and Shayegan)			
Adverse effects	Schizophrenia, bipolar mania, somnolence, respiratory tract infection, extrapyramidal symptoms, dystonia, vital sign changes, weight gain, ECG changes.			
Contraindications	Hypersensitivity, dementia-related psychosis, QT syndrome, cardiac arrhythamias			
Notes	Anti-psychotics as class are associated with weight gain, metabolic syndrome, and NAFLD (Xu and Zhuang, 2019). However, compared to other antipsychotics, ziprasidone had least impact on weight gain, minimal risk for hyperlipidemia, and actually decreased hepatic glucose production (Xu and Zhuang, 2019).			
DrugBank ID	Drug Name	Steatosis-related pathways (targets)	CMap prediction	Network proximity
DB00270	Isradipine	Type II diabetes mellitus (CACNA1C, CACNA1D)	C2, C3	Z-score: -1.08 Rank: 9
Indication	Mild to moderate essential hypertension			

MOA	Isradipine binds directly to inactive calcium channels stabilizing their inactive conformation. Since arterial smooth muscle depolarizations are longer in duration than cardiac muscle depolarizations, inactive channels are more prevalent in smooth muscle cells. Alternative splicing of the α-1 subunit of the channel gives isradipine additional arterial selectivity. At therapeutic sub-toxic concentrations, isradipine has little effect on cardiac myocytes and conduction cells.
Adverse effects	Headache, edema, dizziness, palpitation, flushing, tachycardia, chest pain, rash
Contraindications	Hypersensitivity to isradipine or other calcium channel blockers; hypotension (<90 mm Hg systolic).
Notes	Calcium channel blockers have been suggested for restoring autophagic flux and treating metabolic pathologies in mouse models of obesity (Park, et al., 2014). However, a study found that it had no effect on metabolic syndrome in patients (Widimsky and Sirotiakova, 2006).

DrugBank ID	Drug Name	Steatosis-related pathways (targets)	CMap prediction	Network proximity
DB00420	Promazine	PI3K-Akt signaling pathway (CHRM1, CHRM2)	C1, C2, C4	Z-score: -1.42 Rank: 8
Indication	Moderate and severe psychomotor agitation; agitation or restlessness in the elderly.			
MOA	Promazine is an antagonist at types 1, 2, and 4 dopamine receptors, 5-HT receptor types 2A and 2C, muscarinic receptors 1 through 5, α(1)-receptors, and histamine H1-receptors. Promazine's antipsychotic effect is due to antagonism at dopamine and serotonin type 2 receptors, with greater activity at serotonin 5-HT2 receptors than at dopamine type-2 receptors. This may explain the lack of extrapyramidal effects. Promazine does not appear to block dopamine within the tubero-infundibular tract, explaining the lower incidence of hyperprolactinemia than with typical antipsychotic agents or risperidone. Antagonism at muscarinic receptors, H1-receptors, and α(1)-receptors also occurs with promazine.			
Adverse effects	Nervous system side effects, tardive dyskinesia, fever, hypotension, insomnia, nausea, vomiting, constipation, and diarrhea.			
Notes	Currently not approved for use in the United States. Studies show that antipsychotics including promazine might induce metabolic syndrome and NAFLD (Xu and Zhuang, 2019).			

DrugBank ID	Drug Name	Steatosis-related pathways (targets)	CMap prediction	Network proximity
DB11672	Curcumin	PPAR signaling pathway (PPARG)	C1, C2, C3	Z-score: -0.45 Rank: 15
Indication	NA			
MOA	Curcumin acts as a scavenger of oxygen species, such as hydroxyl radical, superoxide anion, and singlet oxygen and inhibit lipid peroxidation as well as peroxide-induced DNA damage. Curcumin mediates potent anti-inflammatory agent and anti-carcinogenic actions via modulating various signaling molecules. It suppresses a number of key elements in cellular signal transduction pathways pertinent to growth, differentiation, and malignant transformation; it was demonstrated in vitro that curcumin inhibits protein kinases, c-Jun/AP-1 activation, prostaglandin biosynthesis, and the activity and expression of the enzyme cyclooxygenase (COX)-2.			
Adverse effects	Diarrhea, headache, rash, yellow stool, nausea, diarrhea			
Contraindications	NA			
Notes	Clinical trial showed that curcumin supplementation was associated with significant decrease in hepatic fibrosis, nuclear-kappa B activity, hepatic steatosis and serum level of enzymes, and tumor necrosis-α. However, the curcumin supplementation plus lifestyle modification is not superior to lifestyle modification alone in amelioration of inflammation (Saadati, et al., 2019). Another clinical trial showed that nano-curcumin improves glucose indices, lipids, inflammation and nesfatin in overweight and obese patients with NAFLD (Jazayeri-Tehrani, et al., 2019). It is demonstrated that NAFLD severity is reduced with the use of curcumin (White and Lee, 2019). There have been numerous clinical trials for many different indications that used this compound which have not been successful (Nelson, et al., 2017). Curcumin also has overall unfavorable pharmacokinetics/pharmacodynamics (Nelson, et al., 2017).			

Appendix C.7 Drugs in clinical trials for NAFLD and NASH

DrugBank ID	Drug Name	Drug Type*	Condition	Phase	Recruitment Status	Result
DB00169	Cholecalciferol	SMD	NAFLD	4	Completed	Not submitted
DB09038	Empagliflozin	SMD	NAFLD	4	Completed	Not submitted
DB01120	Gliclazide	SMD	NAFLD	4	Completed	Not submitted
DB09198	Lobeglitazone	SMD	NAFLD	4	Completed	Not submitted
DB00331	Metformin	SMD	NAFLD	4	Completed	Not submitted
DB09539	Omega-3-acid ethyl esters	SMD	NAFLD	4	Completed	Influences bio-markers
DB01132	Pioglitazone	SMD	NAFLD	4	Completed	Needs future work
DB01261	Sitagliptin	SMD	NAFLD	4	Completed	Not submitted
DB11094	Vitamin D	SMD	NAFLD	4	Completed	Not submitted
DB01276	Exenatide	BD	NAFLD	4	Completed	Safe and effective
DB13961	Fish oil	BD	NAFLD	4	Completed	Not effective
DB00047	Insulin glargine	BD	NAFLD	4	Completed	Not submitted
DB06655	Liraglutide	BD	NAFLD	4	Completed	Not submitted
DB12625	Evogliptin	SMD	NAFLD	4	Enrolling by Invitation	-
DB00381	Amlodipine	SMD	NAFLD	4	Recruiting	-
DB00222	Glimepiride	SMD	NAFLD	4	Recruiting	-
DB00790	Perindopril	SMD	NAFLD	4	Recruiting	-
DB00191	Phentermine	SMD	NAFLD	4	Recruiting	-
DB00966	Telmisartan	SMD	NAFLD	4	Recruiting	-
DB11824	Tofogliflozin	SMD	NAFLD	4	Recruiting	-
DB01220	Rifaximin	SMD	NAFLD	4	Terminated	-
DB00847	Cysteamine	SMD	NAFLD	3	Completed	Not effective
DB11994	Diacerein	SMD	NAFLD	3	Completed	Not submitted
DB12539	Oltipraz	SMD	NAFLD	3	Completed	Not submitted
DB05408	Emricasan	SMD	NAFLD	2	Completed	Safe and effective
DB01039	Fenofibrate	SMD	NAFLD	2	Completed	Not submitted
DB12030	Fluorescein lisicol	SMD	NAFLD	2	Completed	Not submitted
DB05123	Gemcabene	SMD	NAFLD	2	Completed	Not submitted
DB12866	Pradigastat	SMD	NAFLD	2	Completed	Safe and effective

DB09298	Silibinin	SMD	NAFLD	2	Completed	Not submitted
DB12435	Tipelukast	SMD	NAFLD	2	Completed	Not submitted
DB09046	Metreleptin	BD	NAFLD	2	Completed	Improves bio-markers
DB15194	Cotadutide	SMD	NAFLD	2	Recruiting	-
DB14801	Lanifibranor	SMD	NAFLD	2	Recruiting	-
DB13928	Semaglutide	SMD	NAFLD	2	Recruiting	-
DB08869	Tesamorelin	SMD	NAFLD	2	Recruiting	-
DB00052	Somatotropin	BD	NAFLD	2	Recruiting	-
DB01025	Amlexanox	SMD	NAFLD	2	Active Not Recruiting	-
DB08887	Icosapent ethyl	SMD	NAFLD	2	Active Not Recruiting	-
DB15365	Pegbelfermin	BD	NAFLD	2	Active Not Recruiting	-
DB15212	Pemafibrate	SMD	NAFLD	2	Active Not Recruiting	-
DB00284	Acarbose	SMD	NAFLD	2	Terminated	-
DB00973	Ezetimibe	SMD	NAFLD	2	Terminated	-
DB00451	Levothyroxine	SMD	NAFLD	2	Terminated	-
DB05063	Mitoquinone	SMD	NAFLD	2	Terminated	-
DB01586	Ursodeoxychol ic acid	SMD	NAFLD	2	Terminated	-
DB06695	Dabigatran etexilate	SMD	NAFLD	1	Completed	-
DB00627	Niacin	SMD	NAFLD	NA	Completed	Effective
DB11627	Hepatitis B Vaccine (Recombinant)	BD	NAFLD	NA	Recruiting	-
DB04876	Vildagliptin	SMD	NAFLD	NA	Unknown Status	-
DB00316	Acetaminophe n	SMD	NAFLD	NA	Withdrawn	-
DB06817	Raltegravir	SMD	NASH	4	Enrolling by Invitation	-
DB00678	Losartan	SMD	NASH	3	Completed	Not submitted
DB14916	Selonsertib	SMD	NASH	3	Completed	Safe and effective
DB11860	Aramchol	SMD	NASH	3	Recruiting	-
DB11758	Cenicriviroc	SMD	NASH	3	Recruiting	-
DB06292	Dapagliflozin	SMD	NASH	3	Recruiting	-
DB05187	Elafibranor	SMD	NASH	3	Recruiting	-

DB05990	Obeticholic acid	SMD	NASH	3	Active Not Recruiting	-
DB01076	Atorvastatin	SMD	NASH	2	Completed	Safe and effective
DB15125	Belapectin	SMD	NASH	2	Completed	Not submitted
DB06756	Glycine betaine	SMD	NASH	2	Completed	Not submitted
DB12720	Nivocasan	SMD	NASH	2	Completed	Not submitted
DB00338	Omeprazole	SMD	NASH	2	Completed	Not submitted
DB00806	Pentoxifylline	SMD	NASH	2	Completed	Effective
DB09308	Solithromycin	SMD	NASH	2	Completed	Not submitted
DB13946	Testosterone undecanoate	SMD	NASH	2	Completed	Not submitted
DB05185	TRO19622	SMD	NASH	2	Completed	Not submitted
DB00412	Rosiglitazone	SMD	NASH	2	Completed	Not submitted
DB12152	Simtuzumab	BD	NASH	2	Completed	Safe and effective
DB15168	Cilofexor	SMD	NASH	2	Active Not Recruiting	-
DB15171	Tirzepatide	BD	NASH	2	Not Yet Recruiting	-
DB00945	Acetylsalicylic acid	SMD	NASH	2	Recruiting	-
DB12885	CF-102	SMD	NASH	2	Recruiting	-
DB00930	Colesevelam	SMD	NASH	2	Unknown Status	-
DB05372	CP-945598	SMD	NASH	1	Completed	Not submitted
DB11242	Gelatin	SMD	NASH	1	Completed	Not submitted
DB15373	Tilmanocept	SMD	NASH	1	Completed	Not submitted
DB00160	Alanine	SMD	NASH	NA	Completed	Not submitted
DB00811	Ribavirin	SMD	NASH	NA	Unknown Status	-
DB00008	Peginterferon alfa-2a	BD	NASH	NA	Unknown Status	-
DB00022	Peginterferon alfa-2b	BD	NASH	NA	Unknown Status	-

* SMD: Small molecule drug, BD: Biotechdrug

NA: used to describe trials without FDA-defined phases, including trials of devices or behavioral interventions.

Appendix C.8 Targets and pathways of 23 predicted drugs and 20 clinical trial drugs for NAFLD

Only the predefined 11 NAFLD associated pathways and the targets and drugs mapped on these pathways are listed here.

Predicted Drugs						
DrugBank ID	**Drug Name**	**Drug Type**	**Network Z-score**	**CMap Comparison**	**Target Gene Name**	**Pathway Name**
DB00197	Troglitazone	SMD	-3.60	PF vs. PLI	ACSL4	PPAR signaling pathway
						Fatty acid biosynthesis
						Adipocytokine signaling pathway
					PPARA	PPAR signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
						Adipocytokine signaling pathway
					PPARG	PPAR signaling pathway
					PPARD	PPAR signaling pathway
DB01259	Lapatinib	SMD	-2.58	PF vs. PLI	EGFR	PI3K-Akt signaling pathway
					ERBB2	PI3K-Akt signaling pathway
DB08059	Wortmannin	SMD	-2.32	PLI vs. N&S	PIK3CA	Insulin signaling pathway
						PI3K-Akt signaling pathway
						Apoptosis
						TNF signaling pathway
						Type II diabetes mellitus
						Non-alcoholic fatty liver disease (NAFLD)
					PIK3R1	Insulin signaling pathway
						PI3K-Akt signaling pathway
						Apoptosis
						TNF signaling pathway
						Type II diabetes mellitus
						Non-alcoholic fatty liver disease (NAFLD)
					PIK3CG	PI3K-Akt signaling pathway
DB07863	2-chloro-5-nitro-N-phenylbenzamide	SMD	-2.04	PF vs. N&S	RXRA	PPAR signaling pathway
						PI3K-Akt signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
						Adipocytokine signaling pathway

					PPARG	PPAR signaling pathway
DB01590	Everolimus	SMD	-1.95	PF vs. N&S	MTOR	Adipocytokine signaling pathway
						Type II diabetes mellitus
						Insulin signaling pathway
						PI3K-Akt signaling pathway
DB08073	(2S)-1-(1H-INDOL-3-YL)-3-{[5-(3-METHYL-1H-INDAZOL-5-YL)PYRIDIN-3-YL]OXY}PROPAN-2-AMINE	SMD	-1.70	PLI vs. N&S	PRKACA	Insulin signaling pathway
					GSK3B	Non-alcoholic fatty liver disease (NAFLD)
						Insulin signaling pathway
						PI3K-Akt signaling pathway
					AKT2	Adipocytokine signaling pathway
						Insulin signaling pathway
						PI3K-Akt signaling pathway
						Apoptosis
						TNF signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
DB07859	4-(4-CHLOROPHENYL)-4-[4-(1H-PYRAZOL-4-YL)PHENYL]PIPERIDINE	SMD	-1.70	PF vs. PLI	PRKACA	Insulin signaling pathway
					GSK3B	Non-alcoholic fatty liver disease (NAFLD)
						Insulin signaling pathway
						PI3K-Akt signaling pathway
					AKT2	Adipocytokine signaling pathway
						Insulin signaling pathway
						PI3K-Akt signaling pathway
						Apoptosis
						TNF signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
DB00420	Promazine	SMD	-1.42	PF vs. PLI	CHRM2	PI3K-Akt signaling pathway
					CHRM1	PI3K-Akt signaling pathway
DB00270	Isradipine	SMD	-1.08	PF vs. PLI	CACNA1C	Type II diabetes mellitus
					CACNA1D	Type II diabetes mellitus
DB00630	Alendronic acid	SMD	-1.01	PF vs. N&S	ATP6V1A	Oxidative phosphorylation
DB00246	Ziprasidone	SMD	-0.76	PF vs. N&S	CHRM1	PI3K-Akt signaling pathway
					CHRM2	PI3K-Akt signaling pathway
DB09167	Dosulepin	SMD	-0.76	PLI vs. N&S	CHRM2	PI3K-Akt signaling pathway
					CHRM1	PI3K-Akt signaling pathway
DB11672	Curcumin	SMD	-0.45	PF vs. PLI	PPARG	PPAR signaling pathway
DB00482	Celecoxib	SMD	-0.03	PLI vs. N&S	PTGS2	TNF signaling pathway
					PDPK1	PPAR signaling pathway
						Apoptosis

						Insulin signaling pathway
						PI3K-Akt signaling pathway
DB01406	Danazol	SMD	0.00	PF vs. PLI	CCL2	TNF signaling pathway
DB00850	Perphenazine	SMD	0.00	PLI vs. N&S	CALM1	Insulin signaling pathway
DB01183	Naloxone	SMD	0.00	PLI vs. N&S	CREB1	TNF signaling pathway
						PI3K-Akt signaling pathway
					TLR4	PI3K-Akt signaling pathway
DB00317	Gefitinib	SMD	0.00	PF vs. PLI; PLI vs. N&S	EGFR	PI3K-Akt signaling pathway
DB02424	Geldanamycin	SMD	0.43	PF vs. N&S	HSP90AB1	Protein processing in endoplasmic reticulum
						PI3K-Akt signaling pathway
					HSP90AA1	Protein processing in endoplasmic reticulum
						PI3K-Akt signaling pathway
					HSP90B1	Protein processing in endoplasmic reticulum
						PI3K-Akt signaling pathway
DB02860	Calyculin A	SMD	0.44	PF vs. N&S	PPP1CC	Insulin signaling pathway
DB01175	Escitalopram	SMD	0.47	PF vs. N&S; PLI vs. N&S	CHRM1	PI3K-Akt signaling pathway
DB13877	Iniparib	SMD	0.49	PLI vs. N&S	PARP1	Apoptosis
DB01392	Yohimbine	SMD	1.09	PLI vs. N&S	KCNJ11	Type II diabetes mellitus
Clinical Trial Drugs						
DB00412	Rosiglitazone	SMD	-3.27	NASH	RXRA	PPAR signaling pathway
						PI3K-Akt signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
						Adipocytokine signaling pathway
					RXRG	PPAR signaling pathway
						Adipocytokine signaling pathway
					ACSL4	PPAR signaling pathway
						Fatty acid biosynthesis
						Adipocytokine signaling pathway
					PPARA	PPAR signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
						Adipocytokine signaling pathway
					RXRB	PPAR signaling pathway
						Adipocytokine signaling pathway

					PPARG	PPAR signaling pathway
					PPARD	PPAR signaling pathway
DB00052	Somatotropin	BD	-2.66	NAFLD	PRLR	PI3K-Akt signaling pathway
					GHR	PI3K-Akt signaling pathway
DB01132	Pioglitazone	SMD	-2.59	NAFLD	PPARD	PPAR signaling pathway
					PPARG	PPAR signaling pathway
					PPARA	PPAR signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
						Adipocytokine signaling pathway
DB05187	Elafibranor	SMD	-2.59	NASH	PPARD	PPAR signaling pathway
					PPARG	PPAR signaling pathway
					PPARA	PPAR signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
						Adipocytokine signaling pathway
DB00451	Levothyroxine	SMD	-2.36	NAFLD	ITGB3	PI3K-Akt signaling pathway
					ITGAV	PI3K-Akt signaling pathway
DB01039	Fenofibrate	SMD	-1.89	NAFLD	PPARA	PPAR signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
						Adipocytokine signaling pathway
DB09198	Lobeglitazone	SMD	-1.70	NAFLD	PPARG	PPAR signaling pathway
DB13961	Fish oil	BD	-1.68	NAFLD	PTGS2	TNF signaling pathway
					SREBF1	Insulin signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
					NFKB1	Adipocytokine signaling pathway
						Apoptosis
						TNF signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
						PI3K-Akt signaling pathway
					CACNA1D	Type II diabetes mellitus
					PPARD	PPAR signaling pathway
					PPARA	PPAR signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
						Adipocytokine signaling pathway
					CACNA1C	Type II diabetes mellitus
					PPARG	PPAR signaling pathway
DB00966	Telmisartan	SMD	-1.46	NAFLD	PPARG	PPAR signaling pathway
DB00381	Amlodipine	SMD	-1.28	NAFLD	CACNA1B	Type II diabetes mellitus
					CACNA1C	Type II diabetes mellitus
DB05408	Emricasan	SMD	-1.24	NAFLD	CASP7	TNF signaling pathway

						Apoptosis
						Non-alcoholic fatty liver disease (NAFLD)
					CASP3	Apoptosis
						TNF signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
DB00945	Acetylsalicylic acid	SMD	-0.83	NASH	PRKAB1	Insulin signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
						Adipocytokine signaling pathway
					PRKAB2	Insulin signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
						Adipocytokine signaling pathway
					PRKAG3	Insulin signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
						Adipocytokine signaling pathway
					PTGS2	TNF signaling pathway
					PRKAA2	Non-alcoholic fatty liver disease (NAFLD)
						Insulin signaling pathway
						Adipocytokine signaling pathway
						PI3K-Akt signaling pathway
					PRKAA1	Non-alcoholic fatty liver disease (NAFLD)
						Insulin signaling pathway
						Adipocytokine signaling pathway
						PI3K-Akt signaling pathway
					MYC	PI3K-Akt signaling pathway
					MAPK3	Insulin signaling pathway
						PI3K-Akt signaling pathway
						TNF signaling pathway
						Apoptosis
						Type II diabetes mellitus
					MAPK1	Insulin signaling pathway
						PI3K-Akt signaling pathway
						TNF signaling pathway
						Apoptosis
						Type II diabetes mellitus
					NFKBIA	Adipocytokine signaling pathway
						Apoptosis
						TNF signaling pathway
					PRKAG1	Insulin signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)

						Adipocytokine signaling pathway
					TP53	Apoptosis
						PI3K-Akt signaling pathway
					IKBKB	Adipocytokine signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
						Apoptosis
						TNF signaling pathway
						Type II diabetes mellitus
						PI3K-Akt signaling pathway
						Insulin signaling pathway
					PRKAG2	Insulin signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
						Adipocytokine signaling pathway
					CCND1	PI3K-Akt signaling pathway
					HSPA5	Protein processing in endoplasmic reticulum
					CASP3	Apoptosis
						TNF signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
DB01120	Gliclazide	SMD	-0.54	NAFLD	VEGFA	PI3K-Akt signaling pathway
					ABCC8	Type II diabetes mellitus
DB09539	Omega-3-acid ethyl esters	SMD	-0.23	NAFLD	SREBF1	Insulin signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
DB00338	Omeprazole	SMD	-0.12	NASH	ATP4A	Oxidative phosphorylation
DB00316	Acetaminophen	SMD	-0.04	NAFLD	PTGS2	TNF signaling pathway
DB00047	Insulin glargine	BD	0.29	NAFLD	IGF1R	PI3K-Akt signaling pathway
					INSR	Non-alcoholic fatty liver disease (NAFLD)
						Insulin signaling pathway
						Type II diabetes mellitus
						PI3K-Akt signaling pathway
DB00331	Metformin	SMD	0.49	NAFLD	PRKAB1	Insulin signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
						Adipocytokine signaling pathway
DB11994	Diacerein	SMD	0.61	NAFLD	CYP2E1	Non-alcoholic fatty liver disease (NAFLD)
					NR1H3	PPAR signaling pathway
						Non-alcoholic fatty liver disease (NAFLD)
DB01025	Amlexanox	SMD	0.84	NAFLD	IL3	Apoptosis
						PI3K-Akt signaling pathway
					FGF1	PI3K-Akt signaling pathway

Bibliography

Abdulrahman, B.A., Khweek, A.A., Akhter, A., Caution, K., Kotrange, S., Abdelaziz, D.H., Newland, C., Rosales-Reyes, R., Kopp, B., McCoy, K., Montione, R., Schlesinger, L.S., Gavrilin, M.A., Wewers, M.D., Valvano, M.A. and Amer, A.O. Autophagy stimulation by rapamycin suppresses lung inflammation and infection by Burkholderia cenocepacia in a model of cystic fibrosis. *Autophagy* 2011;7(11):1359-1370.

Al-Hasani, R. and Bruchas, M.R. Molecular mechanisms of opioid receptor-dependent signaling and behavior. *Anesthesiology* 2011;115(6):1363-1381.

Alaimo, S., Bonnici, V., Cancemi, D., Ferro, A., Giugno, R. and Pulvirenti, A. DT-Web: a web-based application for drug-target interaction and drug combination prediction through domain-tuned network-based inference. *BMC Syst Biol* 2015;9 Suppl 3:S4.

Allen, M.D. and Zhang, J. Subcellular dynamics of protein kinase A activity visualized by FRET-based reporters. *Biochem Biophys Res Commun* 2006;348(2):716-721.

Alonso-Lopez, D., Campos-Laborie, F.J., Gutierrez, M.A., Lambourne, L., Calderwood, M.A., Vidal, M. and De Las Rivas, J. APID database: redefining protein-protein interaction experimental evidences and binary interactomes. *Database (Oxford)* 2019;2019.

Alonso-Lopez, D., Gutierrez, M.A., Lopes, K.P., Prieto, C., Santamaria, R. and De Las Rivas, J. APID interactomes: providing proteome-based interactomes with controlled quality for multiple species and derived networks. *Nucleic Acids Res* 2016;44(W1):W529-535.

An, J.Y., You, Z.H., Zhou, Y. and Wang, D.F. Sequence-based Prediction of Protein-Protein Interactions Using Gray Wolf Optimizer-Based Relevance Vector Machine. *Evol Bioinform Online* 2019;15:1176934319844522.

Aon, M.A., Bhatt, N. and Cortassa, S.C. Mitochondrial and cellular mechanisms for managing lipid excess. *Front Physiol* 2014;5:282.

Ashburn, T.T. and Thor, K.B. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* 2004;3(8):673-683.

Babalonis, S., Lofwall, M.R., Nuzzo, P.A. and Walsh, S.L. Pharmacodynamic effects of oral oxymorphone: abuse liability, analgesic profile and direct physiologic effects in humans. *Addict Biol* 2016;21(1):146-158.

Bahi, A. and Dreyer, J.L. Cocaine-induced expression changes of axon guidance molecules in the adult rat brain. *Mol Cell Neurosci* 2005;28(2):275-291.

Bahrke, M.S. and Yesalis, C.E. Abuse of anabolic androgenic steroids and related substances in sport and exercise. *Curr Opin Pharmacol* 2004;4(6):614-620.

Bai, J.P.F., Earp, J.C. and Pillai, V.C. Translational Quantitative Systems Pharmacology in Drug Development: from Current Landscape to Good Practices. *AAPS J* 2019;21(4):72.

Bailey, J., Ma, D. and Szumlinski, K.K. Rapamycin attenuates the expression of cocaine-induced place preference and behavioral sensitization. *Addict Biol* 2012;17(2):248-258.

Baillie, G.S., Scott, J.D. and Houslay, M.D. Compartmentalisation of phosphodiesterases and protein kinase A: opposites attract. *FEBS Lett* 2005;579(15):3264-3270.

Bajusz, D., Racz, A. and Heberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J Cheminform* 2015;7:20.

Bandyopadhyay, S. and Mallick, K. A New Feature Vector Based on Gene Ontology Terms for Protein-Protein Interaction Prediction. *IEEE/ACM Trans Comput Biol Bioinform* 2017;14(4):762-770.

Barone, J.A. Domperidone: a peripherally acting dopamine2-receptor antagonist. *Ann Pharmacother* 1999;33(4):429-440.

Bartol, T.M., Bromer, C., Kinney, J., Chirillo, M.A., Bourne, J.N., Harris, K.M. and Sejnowski, T.J. Nanoconnectomic upper bound on the variability of synaptic plasticity. *Elife* 2015;4:e10778.

Basaranoglu, M., Basaranoglu, G., Sabuncu, T. and Senturk, H. Fructose as a key player in the development of fatty liver disease. *World J Gastroenterol* 2013;19(8):1166-1172.

Belousov, A.B. and Fontes, J.D. Neuronal gap junctions: making and breaking connections during development and injury. *Trends Neurosci* 2013;36(4):227-236.

Benarroch, E.E. Endogenous opioid systems: current concepts and clinical correlations. *Neurology* 2012;79(8):807-814.

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N. and Golani, I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* 2001;125(1-2):279-284.

Benztropine, C.-. https://www.ebi.ac.uk/chembl. In.; 2016.

Berg, J.M., Rogers, M.E. and Lyster, P.M. Systems biology and pharmacology. *Clin Pharmacol Ther* 2010;88(1):17-19.

Biernacka, J.M., Geske, J., Jenkins, G.D., Colby, C., Rider, D.N., Karpyak, V.M., Choi, D.S. and Fridley, B.L. Genome-wide gene-set analysis for identification of pathways associated with alcohol dependence. *Int J Neuropsychopharmacol* 2013;16(2):271-278.

Blackett, P.R. and Sanghera, D.K. Genetic determinants of cardiometabolic risk: a proposed model for phenotype association and interaction. *Journal of clinical lipidology* 2013;7(1):65-81.

Bleakley, K. and Yamanishi, Y. Supervised prediction of drug-target interactions using bipartite local models. *Bioinformatics* 2009;25(18):2397-2403.

Bliss, C.I. The toxicity of poisons applied jointly 1. *Annals of applied biology* 1939;26(3):585-615.

Blum, K., Sheridan, P.J., Wood, R.C., Braverman, E.R., Chen, T.J., Cull, J.G. and Comings, D.E. The D2 dopamine receptor gene as a determinant of reward deficiency syndrome. *J R Soc Med* 1996;89(7):396-400.

Bodas, M., Patel, N., Silverberg, D., Walworth, K. and Vij, N. Master Autophagy Regulator Transcription Factor EB Regulates Cigarette Smoke-Induced Autophagy Impairment and Chronic Obstructive Pulmonary Disease-Emphysema Pathogenesis. *Antioxid Redox Signal* 2017;27(3):150-167.

Bodas, M., Silverberg, D., Walworth, K., Brucia, K. and Vij, N. Augmentation of S-Nitrosoglutathione Controls Cigarette Smoke-Induced Inflammatory-Oxidative Stress and Chronic Obstructive Pulmonary Disease-Emphysema Pathogenesis by Restoring Cystic Fibrosis Transmembrane Conductance Regulator Function. *Antioxid Redox Signal* 2017;27(7):433-451.

Bodas, M., Van Westphal, C., Carpenter-Thompson, R., D, K.M. and Vij, N. Nicotine exposure induces bronchial epithelial cell apoptosis and senescence via ROS mediated autophagy-impairment. *Free Radic Biol Med* 2016;97:441-453.

Bolin, B.L., Stoops, W.W., Sites, J.P. and Rush, C.R. Abuse Potential of Oral Phendimetrazine in Cocaine-dependent Individuals: Implications for Agonist-like Replacement Therapy. *J Addict Med* 2016;10(3):156-165.

Bouchecareilh, M., Hutt, D.M., Szajner, P., Flotte, T.R. and Balch, W.E. Histone deacetylase inhibitor (HDACi) suberoylanilide hydroxamic acid (SAHA)-mediated correction of alpha1-antitrypsin deficiency. *J Biol Chem* 2012;287(45):38265-38278.

Boyer, E.W. Dextromethorphan abuse. *Pediatr Emerg Care* 2004;20(12):858-863.

Bradshaw, E.L., Spilker, M.E., Zang, R., Bansal, L., He, H., Jones, R.D.O., Le, K., Penney, M., Schuck, E., Topp, B., Tsai, A., Xu, C., Nijsen, M. and Chan, J.R. Applications of Quantitative Systems Pharmacology in Model-Informed Drug Discovery: Perspective on Impact and Opportunities. *CPT Pharmacometrics Syst Pharmacol* 2019;8(11):777-791.

Braunersreuther, V., Viviani, G.L., Mach, F. and Montecucco, F. Role of cytokines and chemokines in non-alcoholic fatty liver disease. *World J Gastroenterol* 2012;18(8):727-735.

Bray, N.L., Pimentel, H., Melsted, P. and Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* 2016;34:525.

Brunt, E.M., Wong, V.W.S., Nobili, V., Day, C.P., Sookoian, S., Maher, J.J., Bugianesi, E., Sirlin, C.B., Neuschwander-Tetri, B.A. and Rinella, M.E. Nonalcoholic fatty liver disease. *Nature Reviews Disease Primers* 2015;1:15080 EP.

Burrows, J.A., Willis, L.K. and Perlmutter, D.H. Chemical chaperones mediate increased secretion of mutant alpha 1-antitrypsin (alpha 1-AT) Z: A potential pharmacological strategy for prevention of liver injury and emphysema in alpha 1-AT deficiency. *Proc Natl Acad Sci U S A* 2000;97(4):1796-1801.

Buttner, A., Mall, G., Penning, R. and Weis, S. The neuropathology of heroin abuse. *Forensic Sci Int* 2000;113(1-3):435-442.

Cahill, M.E., Bagot, R.C., Gancarz, A.M., Walker, D.M., Sun, H., Wang, Z.J., Heller, E.A., Feng, J., Kennedy, P.J., Koo, J.W., Cates, H.M., Neve, R.L., Shen, L., Dietz, D.M. and Nestler, E.J. Bidirectional Synaptic Structural Plasticity after Chronic Cocaine Administration Occurs through Rap1 Small GTPase Signaling. *Neuron* 2016;89(3):566-582.

Cai, W., Song, J.M., Zhang, B., Sun, Y.P., Yao, H. and Zhang, Y.X. The prevalence of nonalcoholic fatty liver disease and relationship with serum uric acid level in Uyghur population. *ScientificWorldJournal* 2014;2014:393628.

Calhoun, S.R., Galloway, G.P. and Smith, D.E. Abuse potential of dronabinol (Marinol). *J Psychoactive Drugs* 1998;30(2):187-196.

Cao, D.S., Zhang, L.X., Tan, G.S., Xiang, Z., Zeng, W.B., Xu, Q.S. and Chen, A.F. Computational Prediction of DrugTarget Interactions Using Chemical, Biological, and Network Features. *Mol Inform* 2014;33(10):669-681.

Chakravarti, A., Clark, A.G. and Mootha, V.K. Distilling pathophysiology from complex disease genetics. *Cell* 2013;155(1):21-26.

Chambel, S.S., Santos-Goncalves, A. and Duarte, T.L. The Dual Role of Nrf2 in Nonalcoholic Fatty Liver Disease: Regulation of Antioxidant Defenses and Hepatic Lipid Metabolism. *Biomed Res Int* 2015;2015:597134.

Chen, J., Liu, D., Bai, Q., Song, J., Guan, J., Gao, J., Liu, B., Ma, X. and Du, Y. Celecoxib attenuates liver steatosis and inflammation in non-alcoholic steatohepatitis induced by high-fat diet in rats. *Mol Med Rep* 2011;4(5):811-816.

Chen, J.Y., Ren, Y., Yan, P., Belina, M.E., Chung, R.T. and Butt, A.A. Tricyclic antidepressant use and the risk of fibrosis progression in hepatitis C-infected persons: Results from ERCHIVES. *J Viral Hepat* 2018;25(7):825-833.

Chen, K.H., Wang, T.F. and Hu, Y.J. Protein-protein interaction prediction using a hybrid feature representation and a stacked generalization scheme. *BMC Bioinformatics* 2019;20(1):308.

Chen, M., Ju, C.J., Zhou, G., Chen, X., Zhang, T., Chang, K.W., Zaniolo, C. and Wang, W. Multifaceted protein-protein interaction prediction based on Siamese residual RCNN. *Bioinformatics* 2019;35(14):i305-i314.

Chen, M., Ona, V.O., Li, M., Ferrante, R.J., Fink, K.B., Zhu, S., Bian, J., Guo, L., Farrell, L.A., Hersch, S.M., Hobbs, W., Vonsattel, J.P., Cha, J.H. and Friedlander, R.M. Minocycline inhibits

caspase-1 and caspase-3 expression and delays mortality in a transgenic mouse model of Huntington disease. *Nat Med* 2000;6(7):797-801.

Chen, X., Yan, C.C., Zhang, X., Zhang, X., Dai, F., Yin, J. and Zhang, Y. Drug-target interaction prediction: databases, web servers and computational models. *Brief Bioinform* 2016;17(4):696-712.

Cheng, F. In Silico Oncology Drug Repositioning and Polypharmacology. *Methods Mol Biol* 2019;1878:243-261.

Chijiwa, T., Mishima, A., Hagiwara, M., Sano, M., Hayashi, K., Inoue, T., Naito, K., Toshioka, T. and Hidaka, H. Inhibition of forskolin-induced neurite outgrowth and protein phosphorylation by a newly synthesized selective inhibitor of cyclic AMP-dependent protein kinase, N-[2-(p-bromocinnamylamino)ethyl]-5-isoquinolinesulfonamide (H-89), of PC12D pheochromocytoma cells. *J Biol Chem* 1990;265(9):5267-5272.

Choi, W., Namkung, J., Hwang, I., Kim, H., Lim, A., Park, H.J., Lee, H.W., Han, K.-H., Park, S., Jeong, J.-S., Bang, G., Kim, Y.H., Yadav, V.K., Karsenty, G., Ju, Y.S., Choi, C., Suh, J.M., Park, J.Y., Park, S. and Kim, H. Serotonin signals through a gut-liver axis to regulate hepatic steatosis. *Nature Communications* 2018;9(1):4824.

Chou, T.C. and Talalay, P. Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. *Adv Enzyme Regul* 1984;22:27-55.

Cicero, T.J. and Inciardi, J.A. Diversion and abuse of methadone prescribed for pain management. *JAMA* 2005;293(3):297-298.

Clabough, E.B. Huntington's disease: the past, present, and future search for disease modifiers. *Yale J Biol Med* 2013;86(2):217-233.

Coates, K.M. and Flood, P. Ketamine and its preservative, benzethonium chloride, both inhibit human recombinant alpha7 and alpha4beta2 neuronal nicotinic acetylcholine receptors in Xenopus oocytes. *Br J Pharmacol* 2001;134(4):871-879.

Cobanoglu, M.C., Liu, C., Hu, F., Oltvai, Z.N. and Bahar, I. Predicting drug-target interactions using probabilistic matrix factorization. *J Chem Inf Model* 2013;53(12):3399-3409.

Cobanoglu, M.C., Oltvai, Z.N., Taylor, D.L. and Bahar, I. BalestraWeb: efficient online evaluation of drug-target interactions. *Bioinformatics* 2015;31(1):131-133.

Collingridge, G.L., Kehl, S.J. and McLennan, H. Excitatory amino acids in synaptic transmission in the Schaffer collateral-commissural pathway of the rat hippocampus. *J Physiol* 1983;334:33-46.

Collins, S.R., Kemmeren, P., Zhao, X.C., Greenblatt, J.F., Spencer, F., Holstege, F.C., Weissman, J.S. and Krogan, N.J. Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. *Mol Cell Proteomics* 2007;6(3):439-450.

Cosmin Sebastian Voican, M.D., Ph.D. ,, Emmanuelle Corruble, M.D., Ph.D. ,, Sylvie Naveau, M.D., Ph.D. , and and Gabriel Perlemuter, M.D., Ph.D. Antidepressant-Induced Liver Injury: A Review for Clinicians. *American Journal of Psychiatry* 2014;171(4):404-415.

Cui, Y., Zhang, X.Q., Cui, Y., Xin, W.J., Jing, J. and Liu, X.G. Activation of phosphatidylinositol 3-kinase/Akt-mammalian target of Rapamycin signaling pathway in the hippocampus is essential for the acquisition of morphine-induced place preference in rats. *Neuroscience* 2010;171(1):134-143.

Cui, Z., Gao, Y.L., Liu, J.X., Dai, L.Y. and Yuan, S.S. L2,1-GRMF: an improved graph regularized matrix factorization method to predict drug-target interactions. *BMC Bioinformatics* 2019;20(Suppl 8):287.

Daina, A., Michielin, O. and Zoete, V. SwissTargetPrediction: updated data and new features for efficient prediction of protein targets of small molecules. *Nucleic Acids Res* 2019;47(W1):W357-W364.

Davies, S.P., Reddy, H., Caivano, M. and Cohen, P. Specificity and mechanism of action of some commonly used protein kinase inhibitors. *Biochem J* 2000;351(Pt 1):95-105.

Davis, E.F., Klosterhaus, S.L. and Stapleton, H.M. Measurement of flame retardants and triclosan in municipal sewage sludge and biosolids. *Environ Int* 2012;40:1-7.

Dayas, C.V., Smith, D.W. and Dunkley, P.R. An emerging role for the Mammalian target of rapamycin in "pathological" protein translation: relevance to cocaine addiction. *Front Pharmacol* 2012;3:13.

de Mol, P., Krabbe, H.G., de Vries, S.T., Fokkert, M.J., Dikkeschei, B.D., Rienks, R., Bilo, K.M. and Bilo, H.J. Accuracy of handheld blood glucose meters at high altitude. *PLoS One* 2010;5(11):e15485.

De Roo, M., Klauser, P., Garcia, P.M., Poglia, L. and Muller, D. Spine dynamics and synapse remodeling during LTP and memory processes. *Prog Brain Res* 2008;169:199-207.

Demir, M., Lang, S. and Steffen, H.-M. Nonalcoholic fatty liver disease - current status and future directions. *J Dig Dis* 2015;16(10):541--557.

Diaz, J. How Drugs Influence Behavior: A Neuro-Behavioral Approach. *Upper Saddle River, NJ: Prentice Hall* 1997.

DiMasi, J.A., Grabowski, H.G. and Hansen, R.W. Innovation in the pharmaceutical industry: New estimates of R&D costs. *J Health Econ* 2016;47:20-33.

Ding, Y., Tang, J. and Guo, F. Predicting protein-protein interactions via multivariate mutual information of protein sequences. *BMC Bioinformatics* 2016;17(1):398.

DiPilato, L.M., Cheng, X. and Zhang, J. Fluorescent indicators of cAMP and Epac activation reveal differential dynamics of cAMP signaling within discrete subcellular compartments. *Proc Natl Acad Sci U S A* 2004;101(47):16513-16518.

Dooley, M. and Plosker, G.L. Zaleplon: a review of its use in the treatment of insomnia. *Drugs* 2000;60(2):413-445.

Dotson, J.W., Ackerman, D.L. and West, L.J. Ketamine abuse. *Journal of Drug Issues* 1995;25(4):751-757.

Druid, H., Holmgren, P. and Ahlner, J. Flunitrazepam: an evaluation of use, abuse and toxicity. *Forensic Sci Int* 2001;122(2-3):136-141.

Du, J., Ji, Y., Qiao, L., Liu, Y. and Lin, J. Cellular endo-lysosomal dysfunction in the pathogenesis of non-alcoholic fatty liver disease. *Liver Int* 2020;40(2):271-280.

Du, X., Cheng, J., Zheng, T., Duan, Z. and Qian, F. A novel feature extraction scheme with ensemble coding for protein-protein interaction prediction. *Int J Mol Sci* 2014;15(7):12731-12749.

Du, X., Sun, S., Hu, C., Yao, Y., Yan, Y. and Zhang, Y. DeepPPI: Boosting Prediction of Protein-Protein Interactions with Deep Neural Networks. *J Chem Inf Model* 2017;57(6):1499-1510.

Dueck, D., Morris, Q.D. and Frey, B.J. Multi-way clustering of microarray data using probabilistic sparse matrix factorization. *Bioinformatics* 2005;21 Suppl 1:i144-151.

Dumas, M.E., Kinross, J. and Nicholson, J.K. Metabolic phenotyping and systems biology approaches to understanding metabolic syndrome and fatty liver disease. *Gastroenterology* 2014;146(1):46-62.

Edmondson, R., Broglie, J.J., Adcock, A.F. and Yang, L. Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors. *Assay Drug Dev Technol* 2014;12(4):207-218.

Edwards, S., Whisler, K.N., Fuller, D.C., Orsulak, P.J. and Self, D.W. Addiction-related alterations in D1 and D2 dopamine receptor behavioral responses following chronic cocaine self-administration. *Neuropsychopharmacology* 2007;32(2):354-366.

Emamjomeh, A., Goliaei, B., Torkamani, A., Ebrahimpour, R., Mohammadi, N. and Parsian, A. Protein-protein interaction prediction by combined analysis of genomic and conservation information. *Genes Genet Syst* 2014;89(6):259-272.

Endo, H., Niioka, M., Kobayashi, N., Tanaka, M. and Watanabe, T. Butyrate-producing probiotics reduce nonalcoholic fatty liver disease progression in rats: new insight into the probiotics for the gut-liver axis. *PLoS One* 2013;8(5):e63388.

Everitt, B.J. and Robbins, T.W. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat Neurosci* 2005;8(11):1481-1489.

Ezzat, A., Zhao, P., Wu, M., Li, X.L. and Kwoh, C.K. Drug-Target Interaction Prediction with Graph Regularized Matrix Factorization. *IEEE/ACM Trans Comput Biol Bioinform* 2017;14(3):646-656.

Fahey, M.E., Bennett, M.J., Mahon, C., Jager, S., Pache, L., Kumar, D., Shapiro, A., Rao, K., Chanda, S.K., Craik, C.S., Frankel, A.D. and Krogan, N.J. GPS-Prot: a web-based visualization platform for integrating host-pathogen interaction data. *BMC Bioinformatics* 2011;12:298.

Feaver, R.E., Cole, B.K., Lawson, M.J., Hoang, S.A., Marukian, S., Blackman, B.R., Figler, R.A., Sanyal, A.J., Wamhoff, B.R. and Dash, A. Development of an in vitro human liver system for interrogating nonalcoholic steatohepatitis. *JCI Insight* 2016;1(20):e90954.

Fields, S. and Song, O. A novel genetic system to detect protein-protein interactions. *Nature* 1989;340(6230):245-246.

Fink-Jensen, A., Fedorova, I., Wortwein, G., Woldbye, D.P., Rasmussen, T., Thomsen, M., Bolwig, T.G., Knitowski, K.M., McKinzie, D.L., Yamada, M., Wess, J. and Basile, A. Role for M5 muscarinic acetylcholine receptors in cocaine addiction. *J Neurosci Res* 2003;74(1):91-96.

Fleming, J.A. Triazolam abuse. *Can Med Assoc J* 1983;129(4):324-325.

Frankish, A., Vullo, A., Zadissa, A., Yates, A., Thormann, A., Parker, A., Gall, A., Moore, B., Walts, B., Aken, B.L., Cummins, C., Girón, C.G., Ong, C.K., Sheppard, D., Staines, D.M., Murphy, D.N., Zerbino, D.R., Ogeh, D., Perry, E., Haskell, E., Martin, F.J., Cunningham, F., Riat, H.S., Schuilenburg, H., Sparrow, H., Lavidas, I., Loveland, J.E., To, J.K., Mudge, J., Bhai, J., Taylor, K., Billis, K., Gil, L., Haggerty, L., Gordon, L., Amode, M R., Ruffier, M., Patricio, M., Laird, M.R., Muffato, M., Nuhn, M., Kostadima, M., Langridge, N., Izuogu, O.G., Achuthan, P., Hunt, S.E., Janacek, S.H., Trevanion, S.J., Hourlier, T., Juettemann, T., Maurel, T., Newman, V., Akanni, W., McLaren, W., Liu, Z., Barrell, D. and Flicek, P. Ensembl 2018. *Nucleic Acids Research* 2017;46(D1):D754-D761.

Friedman, S.L., Neuschwander-Tetri, B.A., Rinella, M. and Sanyal, A.J. Mechanisms of NAFLD development and therapeutic strategies. *Nature Medicine* 2018;24(7):908-922.

Friedman, S.L., Ratziu, V., Harrison, S.A., Abdelmalek, M.F., Aithal, G.P., Caballeria, J., Francque, S., Farrell, G., Kowdley, K.V., Craxi, A., Simon, K., Fischer, L., Melchor-Khan, L., Vest, J., Wiens, B.L., Vig, P., Seyedkazemi, S., Goodman, Z., Wong, V.W., Loomba, R., Tacke, F., Sanyal, A. and Lefebvre, E. A randomized, placebo-controlled trial of cenicriviroc for treatment of nonalcoholic steatohepatitis with fibrosis. *Hepatology* 2018;67(5):1754-1767.

Gaggini, M., Carli, F., Rosso, C., Buzzigoli, E., Marietti, M., Della Latta, V., Ciociaro, D., Abate, M.L., Gambino, R., Cassader, M., Bugianesi, E. and Gastaldelli, A. Altered amino acid concentrations in NAFLD: Impact of obesity and insulin resistance. *Hepatology* 2018;67(1):145-158.

Galloway, G.P., Frederick-Osborne, S.L., Seymour, R., Contini, S.E. and Smith, D.E. Abuse and therapeutic potential of gamma-hydroxybutyric acid. *Alcohol* 2000;20(3):263-269.

Gandhi, T.K., Zhong, J., Mathivanan, S., Karthick, L., Chandrika, K.N., Mohan, S.S., Sharma, S., Pinkert, S., Nagaraju, S., Periaswamy, B., Mishra, G., Nandakumar, K., Shen, B., Deshpande, N., Nayak, R., Sarker, M., Boeke, J.D., Parmigiani, G., Schultz, J., Bader, J.S. and Pandey, A. Analysis of the human protein interactome and comparison with yeast, worm and fly interaction datasets. *Nat Genet* 2006;38(3):285-293.

Gawin, F.H. and Ellinwood, E.H., Jr. Cocaine and other stimulants. Actions, abuse, and treatment. *N Engl J Med* 1988;318(18):1173-1182.

Gerhard, G.S., Legendre, C., Still, C.D., Chu, X., Petrick, A. and DiStefano, J.K. Transcriptomic Profiling of Obesity-Related Nonalcoholic Steatohepatitis Reveals a Core Set of Fibrosis-Specific Genes. *J Endocr Soc* 2018;2(7):710-726.

Gillman, P.K. Tricyclic antidepressant pharmacology and therapeutic drug interactions updated. (0007-1188 (Print)).

Gloyn, A.L., Weedon, M.N., Owen, K.R., Turner, M.J., Knight, B.A., Hitman, G., Walker, M., Levy, J.C., Sampson, M., Halford, S., McCarthy, M.I., Hattersley, A.T. and Frayling, T.M. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes* 2003;52(2):568-572.

Gold, M.S., Melker, R.J., Dennis, D.M., Morey, T.E., Bajpai, L.K., Pomm, R. and Frost-Pineda, K. Fentanyl abuse and dependence: further evidence for second hand exposure hypothesis. *J Addict Dis* 2006;25(1):15-21.

Gong, L., Thorn Cf Fau - Bertagnolli, M.M., Bertagnolli Mm Fau - Grosser, T., Grosser T Fau - Altman, R.B., Altman Rb Fau - Klein, T.E. and Klein, T.E. Celecoxib pathways: pharmacokinetics and pharmacodynamics. (1744-6880 (Electronic)).

Gough, A., Stern, A.M., Maier, J., Lezon, T., Shun, T.Y., Chennubhotla, C., Schurdak, M.E., Haney, S.A. and Taylor, D.L. Biologically Relevant Heterogeneity: Metrics and Practical Insights. *SLAS Discov* 2017;22(3):213-237.

Gough, A.H., Chen, N., Shun, T.Y., Lezon, T.R., Boltz, R.C., Reese, C.E., Wagner, J., Vernetti, L.A., Grandis, J.R., Lee, A.V., Stern, A.M., Schurdak, M.E. and Taylor, D.L. Identifying and quantifying heterogeneity in high content analysis: application of heterogeneity indices to drug discovery. *PLoS One* 2014;9(7):e102678.

Greco, W.R., Bravo, G. and Parsons, J.C. The search for synergy: a critical review from a response surface perspective. *Pharmacol Rev* 1995;47(2):331-385.

Griffiths, R.R., Bigelow, G. and Liebson, I. Human drug self-administration: double-blind comparison of pentobarbital, diazepam, chlorpromazine and placebo. *Journal of Pharmacology and Experimental Therapeutics* 1979;210(2):301-310.

Guarino, M. and Dufour, J.F. Nicotinamide and NAFLD: Is There Nothing New Under the Sun? *Metabolites* 2019;9(9).

Gunaratnam, K., Vidal, C., Gimble, J.M. and Duque, G. Mechanisms of palmitate-induced lipotoxicity in human osteoblasts. *Endocrinology* 2014;155(1):108-116.

Guney, E., Menche, J., Vidal, M. and Barábasi, A.-L. Network-based in silico drug efficacy screening. *Nature Communications* 2016;7:10331 EP.

Guney, E., Menche, J., Vidal, M. and Barabasi, A.L. Network-based in silico drug efficacy screening. *Nat Commun* 2016;7:10331.

Guo, M.L., Liao, K., Periyasamy, P., Yang, L., Cai, Y., Callen, S.E. and Buch, S. Cocaine-mediated microglial activation involves the ER stress-autophagy axis. *Autophagy* 2015;11(7):995-1009.

Guo, Y., Yu, L., Wen, Z. and Li, M. Using support vector machine combined with auto covariance to predict protein–protein interactions from protein sequences. *Nucleic Acids Research* 2008;36(9):3025-3030.

Ha, J., Park, C., Park, C. and Park, S. IMIPMF: Inferring miRNA-disease interactions using probabilistic matrix factorization. *J Biomed Inform* 2020;102:103358.

Hajak, G., Muller, W.E., Wittchen, H.U., Pittrow, D. and Kirch, W. Abuse and dependence potential for the non-benzodiazepine hypnotics zolpidem and zopiclone: a review of case reports and epidemiological data. *Addiction* 2003;98(10):1371-1378.

Hamp, T. and Rost, B. Evolutionary profiles improve protein-protein interaction prediction from sequence. *Bioinformatics* 2015;31(12):1945-1950.

Hanzelmann, S., Castelo, R. and Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013;14:7.

Hao, M., Bryant, S.H. and Wang, Y. Predicting drug-target interactions by dual-network integrated logistic matrix factorization. *Sci Rep* 2017;7:40376.

Hardwick, R.N., Fisher, C.D., Canet, M.J., Scheffer, G.L. and Cherrington, N.J. Variations in ATP-binding cassette transporter regulation during the progression of human nonalcoholic fatty liver disease. *Drug Metab Dispos* 2011;39(12):2395-2402.

Harris, S.C., Perrino, P.J., Smith, I., Shram, M.J., Colucci, S.V., Bartlett, C. and Sellers, E.M. Abuse potential, pharmacokinetics, pharmacodynamics, and safety of intranasally administered crushed oxycodone HCl abuse-deterrent controlled-release tablets in recreational opioid users. *J Clin Pharmacol* 2014;54(4):468-477.

Harrison, S.A., Bashir, M.R., Guy, C.D., Zhou, R., Moylan, C.A., Frias, J.P., Alkhouri, N., Bansal, M.B., Baum, S., Neuschwander-Tetri, B.A., Taub, R. and Moussa, S.E. Resmetirom (MGL-3196) for the treatment of non-alcoholic steatohepatitis: a multicentre, randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet* 2019;394(10213):2012-2024.

Hawkins, P.C., Skillman, A.G. and Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J Med Chem* 2007;50(1):74-82.

Hebbard, L. and George, J. Animal models of nonalcoholic fatty liver disease. *Nat Rev Gastroenterol Hepatol* 2011;8(1):35-44.

Hecker, N., Ahmed, J., von Eichborn, J., Dunkel, M., Macha, K., Eckert, A., Gilson, M.K., Bourne, P.E. and Preissner, R. SuperTarget goes quantitative: update on drug-target interactions. *Nucleic Acids Res* 2012;40(Database issue):D1113-1117.

Heikkila, R.E., Orlansky, H. and Cohen, G. Studies on the distinction between uptake inhibition and release of [3H] dopamine in rat brain tissue slices. *Biochemical pharmacology* 1975;24(8):847-852.

Henkel, J., Coleman, C.D., Schraplau, A., Jöhrens, K., Weiss, T.S., Jonas, W., Schürmann, A. and Püschel, G.P. Augmented liver inflammation in a microsomal prostaglandin E synthase 1 (mPGES-1)-deficient diet-induced mouse NASH model. *Scientific reports* 2018;8(1):16127.

Hermann, T.S., Li, W., Dominguez, H., Ihlemann, N., Rask-Madsen, C., Major-Pedersen, A., Nielsen, D.B., Hansen, K.W., Hawkins, M., Kober, L. and Torp-Pedersen, C. Quinapril treatment increases insulin-stimulated endothelial function and adiponectin gene expression in patients with type 2 diabetes. *J Clin Endocrinol Metab* 2006;91(3):1001-1008.

Hernandez-Guerra, M., Hadjihambi, A. and Jalan, R. Gap junctions in liver disease: Implications for pathogenesis and therapy. *J Hepatol* 2019;70(4):759-772.

Herrera, B., Addante, A. and Sanchez, A. BMP Signalling at the Crossroad of Liver Fibrosis and Regeneration. *Int J Mol Sci* 2017;19(1).

Hevers, W., Hadley, S.H., Luddens, H. and Amin, J. Ketamine, but not phencyclidine, selectively modulates cerebellar GABA(A) receptors containing alpha6 and delta subunits. *J Neurosci* 2008;28(20):5383-5393.

Hidvegi, T., Ewing, M., Hale, P., Dippold, C., Beckett, C., Kemp, C., Maurice, N., Mukherjee, A., Goldbach, C., Watkins, S., Michalopoulos, G. and Perlmutter, D.H. An autophagy-enhancing drug promotes degradation of mutant alpha1-antitrypsin Z and reduces hepatic fibrosis. *Science* 2010;329(5988):229-232.

Hidvegi, T., Stolz, D.B., Alcorn, J.F., Yousem, S.A., Wang, J., Leme, A.S., Houghton, A.M., Hale, P., Ewing, M., Cai, H., Garchar, E.A., Pastore, N., Annunziata, P., Kaminski, N., Pilewski, J., Shapiro, S.D., Pak, S.C., Silverman, G.A., Brunetti-Pierri, N. and Perlmutter, D.H. Enhancing Autophagy with Drugs or Lung-directed Gene Therapy Reverses the Pathological Effects of Respiratory Epithelial Cell Proteinopathy. *J Biol Chem* 2015;290(50):29742-29757.

Hingorani, A.D., Kuan, V., Finan, C., Kruger, F.A., Gaulton, A., Chopade, S., Sofat, R., MacAllister, R.J., Overington, J.P., Hemingway, H., Denaxas, S., Prieto, D. and Casas, J.P. Improving the odds of drug development success through human genomics: modelling study. *Sci Rep* 2019;9(1):18911.

Hirota, K., Hashimoto, Y. and Lambert, D.G. Interaction of intravenous anesthetics with recombinant human M1-M3 muscarinic receptors expressed in chinese hamster ovary cells. *Anesth Analg* 2002;95(6):1607-1610, table of contents.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., Yang, L., Wolting, C., Donaldson, I., Schandorff, S., Shewnarane, J., Vo, M., Taggart, J., Goudreault, M., Muskat, B., Alfarano, C., Dewar, D., Lin, Z., Michalickova, K., Willems, A.R., Sassi, H., Nielsen, P.A., Rasmussen, K.J., Andersen, J.R., Johansen, L.E., Hansen, L.H., Jespersen, H., Podtelejnikov, A., Nielsen, E., Crawford, J., Poulsen, V., Sorensen, B.D., Matthiesen, J., Hendrickson, R.C., Gleeson, F., Pawson, T., Moran, M.F., Durocher, D., Mann, M., Hogue, C.W., Figeys, D. and Tyers, M. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature* 2002;415(6868):180-183.

Hollister, L.E., Motzenbecker, F.P. and Degan, R.O. Withdrawal reactions from chlordiazepoxide ("Librium"). *Psychopharmacologia* 1961;2:63-68.

Howell, L.L. and Cunningham, K.A. Serotonin 5-HT2 receptor interactions with dopamine function: implications for therapeutics in cocaine use disorder. *Pharmacol Rev* 2015;67(1):176-197.

Hu, Y., Fang, Z., Yang, Y., Rohlsen-Neal, D., Cheng, F. and Wang, J. Analyzing the genes related to nicotine addiction or schizophrenia via a pathway and network based approach. *Sci Rep* 2018;8(1):2894.

Huang, A.C.W. Autonomic nervous system and brain circuitry for internet addiction. In, *Internet Addiction*. Springer; 2017. p. 161-180.

Huang, Y.-A., You, Z.-H., Gao, X., Wong, L. and Wang, L. Using Weighted Sparse Representation Model Combined with Discrete Cosine Transformation to Predict Protein-Protein Interactions from Protein Sequence. *BioMed Research International* 2015;2015:902198.

Huntley, R.P., Sawford, T., Mutowo-Meullenet, P., Shypitsyna, A., Bonilla, C., Martin, M.J. and O'Donovan, C. The GOA database: gene Ontology annotation updates for 2015. *Nucleic Acids Res* 2015;43(Database issue):D1057-1063.

Hustveit, O., Maurset, A. and Oye, I. Interaction of the chiral forms of ketamine with opioid, phencyclidine, sigma and muscarinic receptors. *Pharmacol Toxicol* 1995;77(6):355-359.

Irwin, J.J., Sterling, T., Mysinger, M.M., Bolstad, E.S. and Coleman, R.G. ZINC: a free tool to discover chemistry for biology. *J Chem Inf Model* 2012;52(7):1757-1768.

Ishac, E.J., Jiang, L., Lake, K.D., Varga, K., Abood, M.E. and Kunos, G. Inhibition of exocytotic noradrenaline release by presynaptic cannabinoid CB1 receptors on peripheral sympathetic nerves. *Br J Pharmacol* 1996;118(8):2023-2028.

Ismaiel, A. and Dumitrascu, D.L. Cardiovascular Risk in Fatty Liver Disease: The Liver-Heart Axis-Literature Review. *Front Med (Lausanne)* 2019;6:202.

Isoetarine, C.-. https://www.ebi.ac.uk/chembl. In.; 2016.

Izzo, E., Martin-Fardon, R., Koob, G.F., Weiss, F. and Sanna, P.P. Neural plasticity and addiction: PI3-kinase and cocaine behavioral sensitization. *Nat Neurosci* 2002;5(12):1263-1264.

James, M.H., Quinn, R.K., Ong, L.K., Levi, E.M., Charnley, J.L., Smith, D.W., Dickson, P.W. and Dayas, C.V. mTORC1 inhibition in the nucleus accumbens 'protects' against the expression of drug seeking and 'relapse' and is associated with reductions in GluA1 AMPAR and CAMKIIalpha levels. *Neuropsychopharmacology* 2014;39(7):1694-1702.

Jazayeri-Tehrani, S.A., Rezayat, S.M., Mansouri, S., Qorbani, M., Alavian, S.M., Daneshi-Maskooni, M. and Hosseinzadeh-Attar, M.J. Nano-curcumin improves glucose indices, lipids, inflammation, and Nesfatin in overweight and obese patients with non-alcoholic fatty liver disease (NAFLD): a double-blind randomized placebo-controlled clinical trial. *Nutr Metab (Lond)* 2019;16:8.

Jegatheesan, P. and De Bandt, J.P. Fructose and NAFLD: The Multifaceted Aspects of Fructose Metabolism. *Nutrients* 2017;9(3).

Jensen, T., Abdelmalek, M.F., Sullivan, S., Nadeau, K.J., Green, M., Roncal, C., Nakagawa, T., Kuwabara, M., Sato, Y., Kang, D.H., Tolan, D.R., Sanchez-Lozada, L.G., Rosen, H.R., Lanaspa, M.A., Diehl, A.M. and Johnson, R.J. Fructose and sugar: A major mediator of non-alcoholic fatty liver disease. *J Hepatol* 2018;68(5):1063-1075.

Jia, Y., French, B., Tillman, B. and French, S. Different roles of FAT10, FOXO1, and ADRA2A in hepatocellular carcinoma tumorigenesis in patients with alcoholic steatohepatitis (ASH) vs non-alcoholic steatohepatitis (NASH). *Exp Mol Pathol* 2018;105(1):144-149.

Jin, E.S., Lee, M.H., Murphy, R.E. and Malloy, C.R. Pentose phosphate pathway activity parallels lipogenesis but not antioxidant processes in rat liver. *Am J Physiol Endocrinol Metab* 2018;314(6):E543-e551.

Johansson-Åkhe, I., Mirabello, C. and Wallner, B. Predicting protein-peptide interaction sites using distant protein complexes as structural templates. *Scientific Reports* 2019;9(1):4267.

Johnson, C.C. Logistic matrix factorization for implicit feedback data. *Advances in Neural Information Processing Systems* 2014(27).

Jones, A.H. and Mayberry, J.F. Chronic glutethimide abuse. *Br J Clin Pract* 1986;40(5):213.

Jones, S. and Bonci, A. Synaptic plasticity and drug addiction. *Curr Opin Pharmacol* 2005;5(1):20-25.

Joranson, D.E., Ryan, K.M., Gilson, A.M. and Dahl, J.L. Trends in medical use and abuse of opioid analgesics. *Jama* 2000;283(13):1710-1714.

Kalivas, P.W. and Volkow, N.D. New medications for drug addiction hiding in glutamatergic neuroplasticity. *Mol Psychiatry* 2011;16(10):974-986.

Kamada, M., Sakuma, Y., Hayashida, M. and Akutsu, T. Prediction of protein-protein interaction strength using domain features with supervised regression. *ScientificWorldJournal* 2014;2014:240673.

Kanda, T., Matsuoka, S., Yamazaki, M., Shibata, T., Nirei, K., Takahashi, H., Kaneko, T., Fujisawa, M., Higuchi, T., Nakamura, H., Matsumoto, N., Yamagami, H., Ogawa, M., Imazu, H., Kuroda, K. and Moriyama, M. Apoptosis and non-alcoholic fatty liver diseases. *World J Gastroenterol* 2018;24(25):2661-2672.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;45(D1):D353-D361.

Kapur, S. and Seeman, P. NMDA receptor antagonists ketamine and PCP have direct effects on the dopamine D(2) and serotonin 5-HT(2)receptors-implications for models of schizophrenia. *Mol Psychiatry* 2002;7(8):837-844.

Kasiske, B.L., de Mattos, A., Flechner, S.M., Gallon, L., Meier-Kriesche, H.U., Weir, M.R. and Wilkinson, A. Mammalian target of rapamycin inhibitor dyslipidemia in kidney transplant recipients. *Am J Transplant* 2008;8(7):1384-1392.

Kathiramalainathan, K., Kaplan, H.L., Romach, M.K., Busto, U.E., Li, N.Y., Sawe, J., Tyndale, R.F. and Sellers, E.M. Inhibition of cytochrome P450 2D6 modifies codeine abuse liability. *J Clin Psychopharmacol* 2000;20(4):435-444.

Keiser, M.J., Roth, B.L., Armbruster, B.N., Ernsberger, P., Irwin, J.J. and Shoichet, B.K. Relating protein pharmacology by ligand chemistry. *Nat Biotechnol* 2007;25(2):197-206.

Kell, D.B., Dobson, P.D., Bilsland, E. and Oliver, S.G. The promiscuous binding of pharmaceutical drugs and their transporter-mediated uptake into cells: what we (need to) know and how we can do so. *Drug Discov Today* 2013;18(5-6):218-239.

Kerrien, S., Alam-Faruque, Y., Aranda, B., Bancarz, I., Bridge, A., Derow, C., Dimmer, E., Feuermann, M., Friedrichsen, A., Huntley, R., Kohler, C., Khadake, J., Leroy, C., Liban, A., Lieftink, C., Montecchi-Palazzi, L., Orchard, S., Risse, J., Robbe, K., Roechert, B., Thorneycroft, D., Zhang, Y., Apweiler, R. and Hermjakob, H. IntAct--open source resource for molecular interaction data. *Nucleic Acids Res* 2007;35(Database issue):D561-565.

Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A., Balakrishnan, L., Marimuthu, A., Banerjee, S., Somanathan, D.S., Sebastian, A., Rani, S., Ray, S., Harrys Kishore, C.J., Kanth, S., Ahmed, M., Kashyap, M.K., Mohmood, R., Ramachandra, Y.L., Krishna, V., Rahiman, B.A., Mohan, S., Ranganathan, P., Ramabadran, S., Chaerkady, R. and Pandey, A. Human Protein Reference Database--2009 update. *Nucleic Acids Res* 2009;37(Database issue):D767-772.

Khan, B.V., Sola, S., Lauten, W.B., Natarajan, R., Hooper, W.C., Menon, R.G., Lerakis, S. and Helmy, T. Quinapril, an ACE inhibitor, reduces markers of oxidative stress in the metabolic syndrome. *Diabetes Care* 2004;27(7):1712-1715.

Kiagiadaki, F., Kampa, M., Voumvouraki, A., Castanas, E., Kouroumalis, E. and Notas, G. Activin-A causes Hepatic stellate cell activation via the induction of TNFα and TGFβ in Kupffer cells. *Biochimica et biophysica acta. Molecular basis of disease* 2018;1864(3):891-899.

Kim, D.J., Yoon, S., Ji, S.C., Yang, J., Kim, Y.K., Lee, S., Yu, K.S., Jang, I.J., Chung, J.Y. and Cho, J.Y. Ursodeoxycholic acid improves liver function via phenylalanine/tyrosine pathway and microbiome remodelling in patients with liver dysfunction. *Sci Rep* 2018;8(1):11874.

Kingma, D.P. and Ba, J. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv* 2014:1412.6980.

Kiyosawa, N. and Manabe, S. Data-intensive drug development in the information age: applications of Systems Biology/Pharmacology/Toxicology. *J Toxicol Sci* 2016;41(Special):SP15-SP25.

Klein-Schwartz, W. Abuse and toxicity of methylphenidate. *Curr Opin Pediatr* 2002;14(2):219-223.

Koob, G.F. and Le Moal, M. Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacology* 2001;24(2):97-129.

Koob, G.F. and Volkow, N.D. Neurocircuitry of addiction. *Neuropsychopharmacology* 2010;35(1):217-238.

Koob, G.F. and Volkow, N.D. Neurobiology of addiction: a neurocircuitry analysis. *Lancet Psychiatry* 2016;3(8):760-773.

Kostrzewski, T., Maraver, P., Ouro-Gnao, L., Levi, A., Snow, S., Miedzik, A., Rombouts, K. and Hughes, D. A Microphysiological System for Studying Nonalcoholic Steatohepatitis. *Hepatol Commun* 2020;4(1):77-91.

Kotlyar, M., Pastrello, C., Malik, Z. and Jurisica, I. IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species. *Nucleic Acids Res* 2019;47(D1):D581-D589.

Kouvelas, D., Pourzitaki, C., Papazisis, G., Dagklis, T., Dimou, K. and Kraus, M.M. Nandrolone abuse decreases anxiety and impairs memory in rats via central androgenic receptors. *International Journal of Neuropsychopharmacology* 2008;11(7):925-934.

Kouvelas, D., Pourzitaki, C., Papazisis, G., Dagklis, T., Dimou, K. and Kraus, M.M. Nandrolone abuse decreases anxiety and impairs memory in rats via central androgenic receptors. *Int J Neuropsychopharmacol* 2008;11(7):925-934.

Kramer, J.C., Fischman, V.S. and Littlefield, D.C. Amphetamine abuse. Pattern and effects of high doses taken intravenously. *JAMA* 1967;201(5):305-309.

Krichevsky, A., Meyers, B., Vainstein, A., Maliga, P. and Citovsky, V. Autoluminescent plants. *PLoS One* 2010;5(11):e15461.

Kroemer, G., Galluzzi, L., Vandenabeele, P., Abrams, J., Alnemri, E.S., Baehrecke, E.H., Blagosklonny, M.V., El-Deiry, W.S., Golstein, P., Green, D.R., Hengartner, M., Knight, R.A., Kumar, S., Lipton, S.A., Malorni, W., Nunez, G., Peter, M.E., Tschopp, J., Yuan, J., Piacentini, M., Zhivotovsky, B., Melino, G. and Nomenclature Committee on Cell, D. Classification of cell death: recommendations of the Nomenclature Committee on Cell Death 2009. *Cell Death Differ* 2009;16(1):3-11.

Krstic, J., Galhuber, M., Schulz, T.J., Schupp, M. and Prokesch, A. p53 as a Dichotomous Regulator of Liver Disease: The Dose Makes the Medicine. *Int J Mol Sci* 2018;19(3).

Kuzuyama, T. Biosynthetic studies on terpenoids produced by Streptomyces. *J Antibiot (Tokyo)* 2017;70(7):811-818.

Kyrola, A., Blelloch, G. and Guestrin, C. Graphchi: Large-scale graph computation on just a {PC}. In, *10th Symposium on Operating Systems Design and Implementation*. 2012. p. 31-46.

Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S.A., Haggarty, S.J., Clemons, P.A., Wei, R., Carr, S.A., Lander, E.S. and Golub, T.R. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 2006;313(5795):1929-1935.

Lasoff, D.R., Koh, C.H., Corbett, B., Minns, A.B. and Cantrell, F.L. Loperamide Trends in Abuse and Misuse Over 13 Years: 2002-2015. *Pharmacotherapy* 2017;37(2):249-253.

Law, C.W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., Smyth, G.K. and Ritchie, M.E. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Res* 2016;5.

Law, C.W., Chen, Y., Shi, W. and Smyth, G.K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology* 2014;15(2):R29.

Le Dinh, T., Freneaux, E., Labbe, G., Letteron, P., Degott, C., Geneve, J., Berson, A., Larrey, D. and Pessayre, D. Amineptine, a tricyclic antidepressant, inhibits the mitochondrial oxidation of fatty acids and produces microvesicular steatosis of the liver in mice. *The Journal of pharmacology and experimental therapeutics* 1988;247(2):745-750.

Le, T.T., Ziemba, A., Urasaki, Y., Hayes, E., Brotman, S. and Pizzorno, G. Disruption of uridine homeostasis links liver pyrimidine metabolism to lipid accumulation. *J Lipid Res* 2013;54(4):1044-1057.

Lee, H.T., Chang, H.T., Lee, S., Lin, C.H., Fan, J.R., Lin, S.Z., Hsu, C.Y., Hsieh, C.H. and Shyu, W.C. Role of IGF1R(+) MSCs in modulating neuroplasticity via CXCR4 cross-interaction. *Sci Rep* 2016;6:32595.

Lee-Montiel, F.T., George, S.M., Gough, A.H., Sharma, A.D., Wu, J., DeBiasio, R., Vernetti, L.A. and Taylor, D.L. Control of oxygen tension recapitulates zone-specific functions in human liver microphysiology systems. *Experimental biology and medicine (Maywood, N.J.)* 2017;242(16):1617--1632.

Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. and Storey, J.D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012;28(6):882-883.

Leek, J.T. and Storey, J.D. Capturing Heterogeneity in Gene Expression Studies by Surrogate Variable Analysis. *PLOS Genetics* 2007;3(9):e161.

Leung, E.L., Cao, Z.W., Jiang, Z.H., Zhou, H. and Liu, L. Network-based drug discovery by integrating systems biology and computational technologies. *Brief Bioinform* 2013;14(4):491-505.

Li, C.Y., Mao, X. and Wei, L. Genes and (common) pathways underlying drug addiction. *PLoS Comput Biol* 2008;4(1):e2.

Li, H., Gong, X.J., Yu, H. and Zhou, C. Deep Neural Network Based Predictions of Protein Interactions Using Primary Sequences. *Molecules* 2018;23(8).

Li, H., Pei, F., Taylor, D.L. and Bahar, I. QuartataWeb: integrated chemical-protein-pathway mapping for polypharmacology and chemogenomics. *Bioinformatics* 2020.

Li, J., Pak, S.C., O'Reilly, L.P., Benson, J.A., Wang, Y., Hidvegi, T., Hale, P., Dippold, C., Ewing, M., Silverman, G.A. and Perlmutter, D.H. Fluphenazine reduces proteotoxicity in C. elegans and mammalian models of alpha-1-antitrypsin deficiency. *PLoS One* 2014;9(1):e87260.

Li, P., Chen, J., Zhang, W., Fu, B. and Wang, W. Transcriptome inference and systems approaches to polypharmacology and drug discovery in herbal medicine. *J Ethnopharmacol* 2017;195:127-136.

Li, S., Li, C. and Wang, W. Bile acid signaling in renal water regulation. *Am J Physiol Renal Physiol* 2019;317(1):F73-f76.

Li, X., George, S.M., Vernetti, L., Gough, A.H. and Taylor, D.L. A glass-based, continuously zonated and vascularized human liver acinus microphysiological system (vLAMPS) designed for experimental modeling of diseases and ADME/TOX. *Lab Chip* 2018;18(17):2614-2631.

Li, Y. and Ilie, L. SPRINT: ultrafast protein-protein interaction prediction of the entire human interactome. *BMC Bioinformatics* 2017;18(1):485.

Li, Y.H., Yu, C.Y., Li, X.X., Zhang, P., Tang, J., Yang, Q., Fu, T., Zhang, X., Cui, X., Tu, G., Zhang, Y., Li, S., Yang, F., Sun, Q., Qin, C., Zeng, X., Chen, Z., Chen, Y.Z. and Zhu, F. Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res* 2018;46(D1):D1121-D1127.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P. and Mesirov, J.P. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 2011;27(12):1739-1740.

Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardozza, A.P., Santonico, E., Castagnoli, L. and Cesareni, G. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res* 2012;40(Database issue):D857-861.

Lievens, S., Van der Heyden, J., Masschaele, D., De Ceuninck, L., Petta, I., Gupta, S., De Puysseleyr, V., Vauthier, V., Lemmens, I., De Clercq, D.J., Defever, D., Vanderroost, N., De Smet, A.S., Eyckerman, S., Van Calenbergh, S., Martens, L., De Bosscher, K., Libert, C., Hill, D.E., Vidal, M. and Tavernier, J. Proteome-scale Binary Interactomics in Human Cells. *Mol Cell Proteomics* 2016;15(12):3624-3639.

Lin, J., Liu, L., Wen, Q., Zheng, C., Gao, Y., Peng, S., Tan, Y. and Li, Y. Rapamycin prevents drug seeking via disrupting reconsolidation of reward memory in rats. *Int J Neuropsychopharmacol* 2014;17(1):127-136.

Lin, J.T., Chang, W.C., Chen, H.M., Lai, H.L., Chen, C.Y., Tao, M.H. and Chern, Y. Regulation of feedback between protein kinase A and the proteasome system worsens Huntington's disease. *Mol Cell Biol* 2013;33(5):1073-1084.

Lipton, J.O. and Sahin, M. The neurology of mTOR. *Neuron* 2014;84(2):275-291.

Liss, K.H. and Finck, B.N. PPARs and nonalcoholic fatty liver disease. *Biochimie* 2017;136:65-74.

Liu, C., Liu, L., Zhu, H.D., Sheng, J.Q., Wu, X.L., He, X.X., Tian, D.A., Liao, J.Z. and Li, P.Y. Celecoxib alleviates nonalcoholic fatty liver disease by restoring autophagic flux. *Sci Rep* 2018;8(1):4108.

Liu, W., Baker, S.S., Baker, R.D. and Zhu, L. Antioxidant Mechanisms in Nonalcoholic Fatty Liver Disease. *Curr Drug Targets* 2015;16(12):1301-1314.

Loomba, R., Lawitz, E., Mantry, P.S., Jayakumar, S., Caldwell, S.H., Arnold, H., Diehl, A.M., Djedjos, C.S., Han, L., Myers, R.P., Subramanian, G.M., McHutchison, J.G., Goodman, Z.D., Afdhal, N.H., Charlton, M.R. and Investigators, G.-U.-. The ASK1 inhibitor selonsertib in patients with nonalcoholic steatohepatitis: A randomized, phase 2 trial. *Hepatology* 2018;67(2):549-559.

Loomba, R. and Sanyal, A.J. The global NAFLD epidemic. *Nat Rev Gastroenterol Hepatol* 2013;10(11):686-690.

Love, S., Mudasir, M.A., Bhardwaj, S.C., Singh, G. and Tasduq, S.A. Long-term administration of tacrolimus and everolimus prevents high cholesterol-high fructose-induced steatosis in C57BL/6J mice by inhibiting de-novo lipogenesis. *Oncotarget* 2017;8(69):113403-113417.

Lu, B., Al-Ramahi, I., Valencia, A., Wang, Q., Berenshteyn, F., Yang, H., Gallego-Flores, T., Ichcho, S., Lacoste, A., Hild, M., Difiglia, M., Botas, J. and Palacino, J. Identification of NUB1 as a suppressor of mutant Huntington toxicity via enhanced protein clearance. *Nat Neurosci* 2013;16(5):562-570.

Ludwig, A.M. and Levine, J. Patterns of Hallucinogenic Drug Abuse. *JAMA* 1965;191:92-96.

Luo, L. Actin cytoskeleton regulation in neuronal morphogenesis and structural plasticity. *Annu Rev Cell Dev Biol* 2002;18:601-635.

Luo, P., Qin, C., Zhu, L., Fang, C., Zhang, Y., Zhang, H., Pei, F., Tian, S., Zhu, X.Y., Gong, J., Mao, Q., Xiao, C., Su, Y., Zheng, H., Xu, T., Lu, J. and Zhang, J. Ubiquitin-Specific Peptidase 10 (USP10) Inhibits Hepatic Steatosis, Insulin Resistance, and Inflammation Through Sirt6. *Hepatology* 2018;68(5):1786-1803.

Ma, J., Wan, J., Meng, J., Banerjee, S., Ramakrishnan, S. and Roy, S. Methamphetamine induces autophagy as a pro-survival response against apoptotic endothelial cell death through the Kappa opioid receptor. *Cell Death Dis* 2014;5:e1099.

Ma, K., Chen, Y., Liang, X., Miao, J. and Zhao, Q. Inhibition of 5-lipoxygenase inhibitor zileuton in high-fat diet-induced nonalcoholic fatty liver disease progression model. *Iran J Basic Med Sci* 2017;20(11):1207-1212.

Ma, X.M. and Blenis, J. Molecular mechanisms of mTOR-mediated translational control. *Nat Rev Mol Cell Biol* 2009;10(5):307-318.

Ma'ayan, A., Rouillard, A.D., Clark, N.R., Wang, Z., Duan, Q. and Kou, Y. Lean Big Data integration in systems biology and systems pharmacology. *Trends Pharmacol Sci* 2014;35(9):450-460.

Madrak, L.N. and Rosenberg, M. Zolpidem abuse. *Am J Psychiatry* 2001;158(8):1330-1331.

Maiers, J.L. and Malhi, H. Endoplasmic Reticulum Stress in Metabolic Liver Diseases and Hepatic Fibrosis. *Semin Liver Dis* 2019;39(2):235-248.

Mangi, M.A., Rehman, H., Minhas, A.M., Rafique, M., Bansal, V. and Constantin, J. Non-Alcoholic Fatty Liver Disease Association with Cardiac Arrhythmias. *Cureus* 2017;9(4):e1165.

Mann, J.P., Semple, R.K. and Armstrong, M.J. How Useful Are Monogenic Rodent Models for the Study of Human Non-Alcoholic Fatty Liver Disease? *Front Endocrinol (Lausanne)* 2016;7:145.

Mardinoglu, A., Boren, J., Smith, U., Uhlen, M. and Nielsen, J. Systems biology in hepatology: approaches and applications. *Nat Rev Gastroenterol Hepatol* 2018;15(6):365-377.

Marinka Zitnik, R.S., Sagar Maheshwari, and Leskovec, J. BioSNAP Datasets: Stanford Biomedical Network Dataset Collection. In.; 2018.

Martinez-Clemente, M., Ferre, N., Gonzalez-Periz, A., Lopez-Parra, M., Horrillo, R., Titos, E., Moran-Salvador, E., Miquel, R., Arroyo, V., Funk, C.D. and Claria, J. 5-lipoxygenase deficiency reduces hepatic inflammation and tumor necrosis factor alpha-induced hepatocyte damage in hyperlipidemia-prone ApoE-null mice. *Hepatology* 2010;51(3):817-827.

Martz, C.A., Ottina, K.A., Singleton, K.R., Jasper, J.S., Wardell, S.E., Peraza-Penton, A., Anderson, G.R., Winter, P.S., Wang, T., Alley, H.M., Kwong, L.N., Cooper, Z.A., Tetzlaff, M., Chen, P.L., Rathmell, J.C., Flaherty, K.T., Wargo, J.A., McDonnell, D.P., Sabatini, D.M. and Wood, K.C. Systematic identification of signaling pathways with potential to confer anticancer drug resistance. *Sci Signal* 2014;7(357):ra121.

Matsuda, S., Kobayashi, M. and Kitagishi, Y. Roles for PI3K/AKT/PTEN Pathway in Cell Signaling of Nonalcoholic Fatty Liver Disease. *ISRN Endocrinol* 2013;2013:472432.

McPherson, S., Hardy, T., Henderson, E., Burt, A.D., Day, C.P. and Anstee, Q.M. Evidence of NAFLD progression from steatosis to fibrosing-steatohepatitis using paired biopsies: implications for prognosis and clinical management. *J Hepatol* 2015;62(5):1148-1155.

Mechoulam, R. and Parker, L.A. The endocannabinoid system and the brain. *Annu Rev Psychol* 2013;64:21-47.

Mehta, R., Birerdinc, A., Wang, L., Younoszai, Z., Moazzez, A., Elariny, H., Goodman, Z., Chandhoke, V., Baranova, A. and Younossi, Z.M. Expression of energy metabolism related genes in the gastric tissue of obese individuals with non-alcoholic fatty liver disease. *BMC Gastroenterol* 2014;14:72.

Mellar, J. and Hollister, L.E. Phenmetrazine: an obsolete problem drug. *Clin Pharmacol Ther* 1982;32(6):671-675.

Melton, L. Protein arrays: proteomics in multiplex. *Nature* 2004;429(6987):101-107.

Mendelsohn, L.G., Kalra, V., Johnson, B.G. and Kerchner, G.A. Sigma opioid receptor: characterization and co-identity with the phencyclidine receptor. *J Pharmacol Exp Ther* 1985;233(3):597-602.

Mendez, D., Gaulton, A., Bento, A.P., Chambers, J., De Veij, M., Felix, E., Magarinos, M.P., Mosquera, J.F., Mutowo, P., Nowotka, M., Gordillo-Maranon, M., Hunter, F., Junco, L., Mugumbate, G., Rodriguez-Lopez, M., Atkinson, F., Bosc, N., Radoux, C.J., Segura-Cabrera, A., Hersey, A. and Leach, A.R. ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res* 2019;47(D1):D930-D940.

Mewes, H.W., Frishman, D., Mayer, K.F., Munsterkotter, M., Noubibou, O., Pagel, P., Rattei, T., Oesterheld, M., Ruepp, A. and Stumpflen, V. MIPS: analysis and annotation of proteins from whole genomes in 2005. *Nucleic Acids Res* 2006;34(Database issue):D169-172.

Middleton, S.A., Rajpal, N., Cutler, L., Mander, P., Rioja, I., Prinjha, R.K., Rajpal, D., Agarwal, P. and Kumar, V. BET Inhibition Improves NASH and Liver Fibrosis. *Sci Rep* 2018;8(1):17257.

Mikolasevic, I., Filipec-Kanizaj, T., Mijic, M., Jakopcic, I., Milic, S., Hrstic, I., Sobocan, N., Stimac, D. and Burra, P. Nonalcoholic fatty liver disease and liver transplantation - Where do we stand? *World J Gastroenterol* 2018;24(14):1491-1506.

Miller, G.M. The emerging role of trace amine-associated receptor 1 in the functional regulation of monoamine transporters and dopaminergic activity. *J Neurochem* 2011;116(2):164-176.

Moeller, F.G., Dougherty, D.M., Rustin, T., Swann, A.C., Allen, T.J., Shah, N. and Cherek, D.R. Antisocial personality disorder and aggression in recently abstinent cocaine dependent subjects. *Drug Alcohol Depend* 1997;44(2-3):175-182.

Mongia, A. and Majumdar, A. Drug-target interaction prediction using Multi Graph Regularized Nuclear Norm Minimization. *PLoS One* 2020;15(1):e0226484.

Moradi, S., Charkhpour, M., Ghavimi, H., Motahari, R., Ghaderi, M. and Hassanzadeh, K. Gap junction blockers: a potential approach to attenuate morphine withdrawal symptoms. *J Biomed Sci* 2013;20:77.

Morris, D.I., Robbins, J.D., Ruoho, A.E., Sutkowski, E.M. and Seamon, K.B. Forskolin photoaffinity labels with specificity for adenylyl cyclase and the glucose transporter. *J Biol Chem* 1991;266(20):13377-13384.

Mudaliar, S., Henry, R.R., Sanyal, A.J., Morrow, L., Marschall, H.U., Kipnes, M., Adorini, L., Sciacca, C.I., Clopton, P., Castelloe, E., Dillon, P., Pruzanski, M. and Shapiro, D. Efficacy and safety of the farnesoid X receptor agonist obeticholic acid in patients with type 2 diabetes and nonalcoholic fatty liver disease. *Gastroenterology* 2013;145(3):574-582 e571.

Muirhead, R.W. and Monaghan, R.M. A two reservoir model to predict Escherichia coli losses to water from pastures grazed by dairy cows. *Environ Int* 2012;40:8-14.

Murray, A.J. Pharmacological PKA inhibition: all may not be what it seems. *Sci Signal* 2008;1(22):re4.

Myeku, N., Clelland, C.L., Emrani, S., Kukushkin, N.V., Yu, W.H., Goldberg, A.L. and Duff, K.E. Tau-driven 26S proteasome impairment and cognitive dysfunction can be prevented early in disease by activating cAMP-PKA signaling. *Nat Med* 2016;22(1):46-53.

Naik, A., Belic, A., Zanger, U.M. and Rozman, D. Molecular Interactions between NAFLD and Xenobiotic Metabolism. *Front Genet* 2013;4:2.

Narita, M., Akai, H., Kita, T., Nagumo, Y., Narita, M., Sunagawa, N., Hara, C., Hasebe, K., Nagase, H. and Suzuki, T. Involvement of mitogen-stimulated p70-S6 kinase in the development of sensitization to the methamphetamine-induced rewarding effect in rats. *Neuroscience* 2005;132(3):553-560.

Natsvlishvili, N., Goguadze, N., Zhuravliova, E. and Mikeladze, D. Sigma-1 receptor directly interacts with Rac1-GTPase in the brain mitochondria. *BMC Biochem* 2015;16:11.

Neasta, J., Barak, S., Hamida, S.B. and Ron, D. mTOR complex 1: a key player in neuroadaptations induced by drugs of abuse. *J Neurochem* 2014;130(2):172-184.

Neasta, J., Ben Hamida, S., Yowell, Q., Carnicella, S. and Ron, D. Role for mammalian target of rapamycin complex 1 signaling in neuroadaptations underlying alcohol-related disorders. *Proc Natl Acad Sci U S A* 2010;107(46):20093-20098.

Neasta, J., Ben Hamida, S., Yowell, Q.V., Carnicella, S. and Ron, D. AKT signaling pathway in the nucleus accumbens mediates excessive alcohol drinking behaviors. *Biol Psychiatry* 2011;70(6):575-582.

Neiman, J., Haapaniemi, H.M. and Hillbom, M. Neurological complications of drug abuse: pathophysiological mechanisms. *Eur J Neurol* 2000;7(6):595-606.

Nelson, K.M., Dahlin, J.L., Bisson, J., Graham, J., Pauli, G.F. and Walters, M.A. The Essential Medicinal Chemistry of Curcumin. *Journal of medicinal chemistry* 2017;60(5):1620-1637.

Nennig, S.E. and Schank, J.R. The Role of NFkB in Drug Addiction: Beyond Inflammation. *Alcohol Alcohol* 2017;52(2):172-179.

Nestler, E.J. Cellular basis of memory for addiction. *Dialogues Clin Neurosci* 2013;15(4):431-443.

Neuschwander-Tetri, B.A., Loomba, R., Sanyal, A.J., Lavine, J.E., Van Natta, M.L., Abdelmalek, M.F., Chalasani, N., Dasarathy, S., Diehl, A.M., Hameed, B., Kowdley, K.V., McCullough, A., Terrault, N., Clark, J.M., Tonascia, J., Brunt, E.M., Kleiner, D.E., Doo, E. and Network, N.C.R. Farnesoid X nuclear receptor ligand obeticholic acid for non-cirrhotic, non-alcoholic steatohepatitis (FLINT): a multicentre, randomised, placebo-controlled trial. *Lancet* 2015;385(9972):956-965.

Nickel, J., Gohlke, B.O., Erehman, J., Banerjee, P., Rong, W.W., Goede, A., Dunkel, M. and Preissner, R. SuperPred: update on drug classification and target prediction. *Nucleic Acids Res* 2014;42(Web Server issue):W26-31.

Nishimori, I., Minakuchi T Fau - Onishi, S., Onishi S Fau - Vullo, D., Vullo D Fau - Cecchi, A., Cecchi A Fau - Scozzafava, A., Scozzafava A Fau - Supuran, C.T. and Supuran, C.T. Carbonic anhydrase inhibitors: cloning, characterization, and inhibition studies of the cytosolic isozyme III with sulfonamides. (0968-0896 (Print)).

O'Connor, J.J., Moloney, E., Travers, R. and Campbell, A. Buprenorphine abuse among opiate addicts. *Br J Addict* 1988;83(9):1085-1087.

O'Reilly, L.P., Long, O.S., Cobanoglu, M.C., Benson, J.A., Luke, C.J., Miedel, M.T., Hale, P., Perlmutter, D.H., Bahar, I., Silverman, G.A. and Pak, S.C. A genome-wide RNAi screen identifies potential drug targets in a C. elegans model of alpha1-antitrypsin deficiency. *Hum Mol Genet* 2014;23(19):5123-5132.

Oates, J.R., McKell, M.C., Moreno-Fernandez, M.E., Damen, M., Deepe, G.S., Jr., Qualls, J.E. and Divanovic, S. Macrophage Function in the Pathogenesis of Non-alcoholic Fatty Liver Disease: The Mac Attack. *Front Immunol* 2019;10:2893.

Okamoto, T., Minami, K., Uezono, Y., Ogata, J., Shiraishi, M., Shigematsu, A. and Ueta, Y. The inhibitory effects of ketamine and pentobarbital on substance p receptors expressed in Xenopus oocytes. *Anesth Analg* 2003;97(1):104-110, table of contents.

Ona, V.O., Li, M., Vonsattel, J.P., Andrews, L.J., Khan, S.Q., Chung, W.M., Frey, A.S., Menon, A.S., Li, X.J., Stieg, P.E., Yuan, J., Penney, J.B., Young, A.B., Cha, J.H. and Friedlander, R.M. Inhibition of caspase-1 slows disease progression in a mouse model of Huntington's disease. *Nature* 1999;399(6733):263-267.

Orabona, C., Mondanelli, G., Pallotta, M.T., Carvalho, A., Albini, E., Fallarino, F., Vacca, C., Volpi, C., Belladonna, M.L., Berioli, M.G., Ceccarini, G., Esposito, S.M., Scattoni, R., Verrotti, A., Ferretti, A., De Giorgi, G., Toni, S., Cappa, M., Matteoli, M.C., Bianchi, R., Matino, D., Iacono, A., Puccetti, M., Cunha, C., Bicciato, S., Antognelli, C., Talesa, V.N., Chatenoud, L., Fuchs, D., Pilotte, L., Van den Eynde, B., Lemos, M.C., Romani, L., Puccetti, P. and Grohmann, U. Deficiency of immunoregulatory indoleamine 2,3-dioxygenase 1in juvenile diabetes. *JCI Insight* 2018;3(6).

Page, S., Chandhoke, V. and Baranova, A. Melanin and melanogenesis in adipose tissue: possible mechanisms for abating oxidative stress and inflammation? *Obesity reviews : an official journal of the International Association for the Study of Obesity* 2011;12(5):e21-31.

Pamies, D., Barreras, P., Block, K., Makri, G., Kumar, A., Wiersma, D., Smirnova, L., Zang, C., Bressler, J., Christian, K.M., Harris, G., Ming, G.L., Berlinicke, C.J., Kyro, K., Song, H., Pardo, C.A., Hartung, T. and Hogberg, H.T. A human brain microphysiological system derived from induced pluripotent stem cells to study neurological diseases and toxicity. *ALTEX* 2017;34(3):362-376.

Pantziarka, P., Verbaanderd, C., Sukhatme, V., Rica Capistrano, I., Crispino, S., Gyawali, B., Rooman, I., Van Nuffel, A.M., Meheus, L., Sukhatme, V.P. and Bouche, G. ReDO_DB: the repurposing drugs in oncology database. *Ecancermedicalscience* 2018;12:886.

Park, H.W., Park, H., Semple, I.A., Jang, I., Ro, S.H., Kim, M., Cazares, V.A., Stuenkel, E.L., Kim, J.J., Kim, J.S. and Lee, J.H. Pharmacological correction of obesity-induced autophagy arrest using calcium channel blockers. *Nat Commun* 2014;5:4834.

Park, J., Jeon, Y., In, D., Fishel, R., Ban, C. and Lee, J.B. Single-molecule analysis reveals the kinetics and physiological relevance of MutL-ssDNA binding. *PLoS One* 2010;5(11):e15496.

Parolaro, D. and Rubino, T. The role of the endogenous cannabinoid system in drug addiction. *Drug News Perspect* 2008;21(3):149-157.

Pasarin, M., Abraldes, J.G., Liguori, E., Kok, B. and La Mura, V. Intrahepatic vascular changes in non-alcoholic fatty liver disease: Potential role of insulin-resistance and endothelial dysfunction. *World J Gastroenterol* 2017;23(37):6777-6787.

Paschos, P. and Tziomalos, K. Nonalcoholic fatty liver disease and the renin-angiotensin system: Implications for treatment. *World J Hepatol* 2012;4(12):327-331.

Passie, T., Seifert, J., Schneider, U. and Emrich, H.M. The pharmacology of psilocybin. *Addict Biol* 2002;7(4):357-364.

Patel, A.S., Lin, L., Geyer, A., Haspel, J.A., An, C.H., Cao, J., Rosas, I.O. and Morse, D. Autophagy in idiopathic pulmonary fibrosis. *PLoS One* 2012;7(7):e41394.

Paz-Filho, G., Mastronardi, C.A., Parker, B.J., Khan, A., Inserra, A., Matthaei, K.I., Ehrhart-Bornstein, M., Bornstein, S., Wong, M.L. and Licinio, J. Molecular pathways involved in the improvement of non-alcoholic fatty liver disease. *J Mol Endocrinol* 2013;51(1):167-179.

Pei, F., Li, H., Henderson, M.J., Titus, S.A., Jadhav, A., Simeonov, A., Cobanoglu, M.C., Mousavi, S.H., Shun, T., McDermott, L., Iyer, P., Fioravanti, M., Carlisle, D., Friedlander, R.M., Bahar, I., Taylor, D.L., Lezon, T.R., Stern, A.M. and Schurdak, M.E. Connecting Neuronal Cell Protective Pathways and Drug Combinations in a Huntington's Disease Model through the Application of Quantitative Systems Pharmacology. *Sci Rep* 2017;7(1):17803.

Pei, F., Li, H., Liu, B. and Bahar, I. Quantitative Systems Pharmacological Analysis of Drugs of Abuse Reveals the Pleiotropy of Their Targets and the Effector Role of mTORC1. *Front Pharmacol* 2019;10:191.

Peretti, S., Judge, R. and Hindmarch, I. Safety and tolerability considerations: tricyclic antidepressants vs. selective serotonin reuptake inhibitors. *Acta Psychiatr Scand Suppl* 2000;403:17-25.

Perez-Nueno, V.I. Using quantitative systems pharmacology for novel drug discovery. *Expert Opin Drug Discov* 2015;10(12):1315-1331.

Peri, S., Navarro, J.D., Amanchy, R., Kristiansen, T.Z., Jonnalagadda, C.K., Surendranath, V., Niranjan, V., Muthusamy, B., Gandhi, T.K., Gronborg, M., Ibarrola, N., Deshpande, N., Shanker, K., Shivashankar, H.N., Rashmi, B.P., Ramya, M.A., Zhao, Z., Chandrika, K.N., Padma, N., Harsha, H.C., Yatish, A.J., Kavitha, M.P., Menezes, M., Choudhury, D.R., Suresh, S., Ghosh, N., Saravana, R., Chandran, S., Krishna, S., Joy, M., Anand, S.K., Madavan, V., Joseph, A., Wong, G.W., Schiemann, W.P., Constantinescu, S.N., Huang, L., Khosravi-Far, R., Steen, H., Tewari, M., Ghaffari, S., Blobe, G.C., Dang, C.V., Garcia, J.G., Pevsner, J., Jensen, O.N., Roepstorff, P., Deshpande, K.S., Chinnaiyan, A.M., Hamosh, A., Chakravarti, A. and Pandey, A. Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res* 2003;13(10):2363-2371.

Philibin, S.D., Hernandez, A., Self, D.W. and Bibb, J.A. Striatal signal transduction and drug addiction. *Front Neuroanat* 2011;5:60.

Pöch, G. and Kukovetz, W.R. Papaverine-induced inhibition of phosphodiesterase activity in various mammalian tissues. *Life Sciences* 1971;10(3):133-144.

Pockros, P.J., Fuchs, M., Freilich, B., Schiff, E., Kohli, A., Lawitz, E.J., Hellstern, P.A., Owens-Grillo, J., Van Biene, C., Shringarpure, R., MacConell, L., Shapiro, D. and Cohen, D.E. CONTROL: A randomized phase 2 study of obeticholic acid and atorvastatin on lipoproteins in nonalcoholic steatohepatitis patients. *Liver Int* 2019;39(11):2082-2093.

Poklis, J.L., Thompson, C.C., Long, K.A., Lichtman, A.H. and Poklis, A. Disposition of cannabichromene, cannabidiol, and Delta(9)-tetrahydrocannabinol and its metabolites in mouse brain following marijuana inhalation determined by high-performance liquid chromatography-tandem mass spectrometry. *J Anal Toxicol* 2010;34(8):516-520.

Polyzos, S.A., Kang, E.S., Boutari, C., Rhee, E.J. and Mantzoros, C.S. Current and emerging pharmacological options for the treatment of nonalcoholic steatohepatitis. *Metabolism* 2020:154203.

Preston, K.L., Jasinski, D.R. and Testa, M. Abuse potential and pharmacological comparison of tramadol and morphine. *Drug Alcohol Depend* 1991;27(1):7-17.

Prosser, R.A., Stowie, A., Amicarelli, M., Nackenoff, A.G., Blakely, R.D. and Glass, J.D. Cocaine modulates mammalian circadian clock timing by decreasing serotonin transport in the SCN. *Neuroscience* 2014;275:184-193.

Puighermanal, E., Marsicano, G., Busquets-Garcia, A., Lutz, B., Maldonado, R. and Ozaita, A. Cannabinoid modulation of hippocampal long-term memory is mediated by mTOR signaling. *Nat Neurosci* 2009;12(9):1152-1158.

Pytliak, M., Vargova, V., Mechirova, V. and Felsoci, M. Serotonin receptors - from molecular biology to clinical applications. *Physiol Res* 2011;60(1):15-25.

Rabanal-Ruiz, Y., Otten, E.G. and Korolchuk, V.I. mTORC1 as the main gateway to autophagy. *Essays Biochem* 2017;61(6):565-584.

Radziejewska, A., Muzsik, A., Milagro, F.I., Martinez, J.A. and Chmurzynska, A. One-Carbon Metabolism and Nonalcoholic Fatty Liver Disease: The Crosstalk between Nutrients, Microbiota, and Genetics. *Lifestyle Genom* 2019:1-11.

Ratziu, V., Harrison, S.A., Francque, S., Bedossa, P., Lehert, P., Serfaty, L., Romero-Gomez, M., Boursier, J., Abdelmalek, M., Caldwell, S., Drenth, J., Anstee, Q.M., Hum, D., Hanf, R., Roudot, A., Megnien, S., Staels, B., Sanyal, A. and Group, G.-I.S. Elafibranor, an Agonist of the Peroxisome Proliferator-Activated Receptor-alpha and -delta, Induces Resolution of Nonalcoholic Steatohepatitis Without Fibrosis Worsening. *Gastroenterology* 2016;150(5):1147-1159 e1145.

Rinaldi, L., Sepe, M., Donne, R.D. and Feliciello, A. A dynamic interface between ubiquitylation and cAMP signaling. *Front Pharmacol* 2015;6:177.

Riordan, J.D. and Nadeau, J.H. Modeling progressive non-alcoholic fatty liver disease in the laboratory mouse. *Mamm Genome* 2014;25(9-10):473-486.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. and Smyth, G.K. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43(7):e47.

Rix, U. and Superti-Furga, G. Target profiling of small molecules by chemical proteomics. *Nat Chem Biol* 2009;5(9):616-624.

Robson, P. Abuse potential and psychoactive effects of delta-9-tetrahydrocannabinol and cannabidiol oromucosal spray (Sativex), a new cannabinoid medicine. *Expert Opin Drug Saf* 2011;10(5):675-685.

Rocha, B.A., Fumagalli, F., Gainetdinov, R.R., Jones, S.R., Ator, R., Giros, B., Miller, G.W. and Caron, M.G. Cocaine self-administration in dopamine-transporter knockout mice. *Nat Neurosci* 1998;1(2):132-137.

Rogers, D. and Hahn, M. Extended-connectivity fingerprints. *J Chem Inf Model* 2010;50(5):742-754.

Romero-Molina, S., Ruiz-Blanco, Y.B., Harms, M., Munch, J. and Sanchez-Garcia, E. PPI-Detect: A support vector machine model for sequence-based prediction of protein-protein interactions. *J Comput Chem* 2019;40(11):1233-1242.

Ross, A.P., Bruggeman, E.C., Kasumu, A.W., Mielke, J.G. and Parent, M.B. Non-alcoholic fatty liver disease impairs hippocampal-dependent memory in male rats. *Physiol Behav* 2012;106(2):133-141.

Rostami, S. and Parsian, H. Hyaluronic Acid: from biochemical characteristics to its clinical translation in assessment of liver fibrosis. *Hepat Mon* 2013;13(12):e13787.

Rothenfluh, A. and Cowan, C.W. Emerging roles of actin cytoskeleton regulating enzymes in drug addiction: actin or reactin'? *Curr Opin Neurobiol* 2013;23(4):507-512.

Roux, K.J., Kim, D.I., Burke, B. and May, D.G. BioID: A Screen for Protein-Protein Interactions. *Curr Protoc Protein Sci* 2018;91:19 23 11-19 23 15.

Rush, C.R., Higgins, S.T., Bickel, W.K. and Hughes, J.R. Abuse liability of alprazolam relative to other commonly used benzodiazepines: a review. *Neurosci Biobehav Rev* 1993;17(3):277-285.

Russo, S.J., Dietz, D.M., Dumitriu, D., Morrison, J.H., Malenka, R.C. and Nestler, E.J. The addicted synapse: mechanisms of synaptic and structural plasticity in nucleus accumbens. *Trends Neurosci* 2010;33(6):267-276.

Ryaboshapkina, M. and Hammar, M. Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. *Sci Rep* 2017;7(1):12361.

Saadati, S., Sadeghi, A., Mansour, A., Yari, Z., Poustchi, H., Hedayati, M., Hatami, B. and Hekmatdoost, A. Curcumin and inflammation in non-alcoholic fatty liver disease: a randomized, placebo controlled clinical trial. *BMC Gastroenterol* 2019;19(1):133.

Sacca, C., Teso, S., Diligenti, M. and Passerini, A. Improved multi-level protein-protein interaction prediction with semantic-based regularization. *BMC Bioinformatics* 2014;15:103.

Sachs, R.E., Ginsburg, P.B. and Goldman, D.P. Encouraging New Uses for Old Drugs. *JAMA* 2017;318(24):2421-2422.

Sample, V., DiPilato, L.M., Yang, J.H., Ni, Q., Saucerman, J.J. and Zhang, J. Regulation of nuclear PKA revealed by spatiotemporal manipulation of cyclic AMP. *Nat Chem Biol* 2012;8(4):375-382.

Sancho, R., Calzado, M.A., Di Marzo, V., Appendino, G. and Munoz, E. Anandamide inhibits nuclear factor-kappaB activation through a cannabinoid receptor-independent pathway. *Mol Pharmacol* 2003;63(2):429-438.

Sanyal, A.J. Past, present and future perspectives in nonalcoholic fatty liver disease. *Nat Rev Gastroenterol Hepatol* 2019;16(6):377-386.

Satapathy, S.K. and Sanyal, A.J. Epidemiology and Natural History of Nonalcoholic Fatty Liver Disease. *Semin Liver Dis* 2015;35(3):221--235.

Saxton, R.A. and Sabatini, D.M. mTOR Signaling in Growth, Metabolism, and Disease. *Cell* 2017;169(2):361-371.

Schenone, M., Dancik, V., Wagner, B.K. and Clemons, P.A. Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol* 2013;9(4):232-240.

Schwartz, R.H. Adolescent abuse of dextromethorphan. *Clin Pediatr (Phila)* 2005;44(7):565-568.

Schwinghammer, U.A., Melkonyan, M.M., Hunanyan, L., Tremmel, R., Weiskirchen, R., Borkham-Kamphorst, E., Schaeffeler, E., Seferyan, T., Mikulits, W., Yenkoyan, K., Schwab, M. and Danielyan, L. α2-Adrenergic Receptor in Liver Fibrosis: Implications for the Adrenoblocker Mesedin. *Cells* 2020;9(2).

Scott, D.E., Bayly, A.R., Abell, C. and Skidmore, J. Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. *Nat Rev Drug Discov* 2016;15(8):533-550.

Seger, D. Cocaine, metamfetamine, and MDMA abuse: the role and clinical importance of neuroadaptation. *Clin Toxicol (Phila)* 2010;48(7):695-708.

Self, D.W. and Nestler, E.J. Molecular mechanisms of drug reinforcement and addiction. *Annu Rev Neurosci* 1995;18:463-495.

Shen, F., Wang, X.W., Ge, F.F., Li, Y.J. and Cui, C.L. Essential role of the NO signaling pathway in the hippocampal CA1 in morphine-associated memory depends on glutaminergic receptors. *Neuropharmacology* 2016;102:216-228.

Shen, J., Zhang, J., Luo, X., Zhu, W., Yu, K., Chen, K., Li, Y. and Jiang, H. Predicting protein-protein interactions based only on sequences information. *Proc Natl Acad Sci U S A* 2007;104(11):4337-4341.

Shi, J., Jun, W., Zhao, L.Y., Xue, Y.X., Zhang, X.Y., Kosten, T.R. and Lu, L. Effect of rapamycin on cue-induced drug craving in abstinent heroin addicts. *Eur J Pharmacol* 2009;615(1-3):108-112.

Shi, J.Y., Zhang, A.Q., Zhang, S.W., Mao, K.T. and Yiu, S.M. A unified solution for different scenarios of predicting drug-target interactions via triple matrix factorization. *BMC Syst Biol* 2018;12(Suppl 9):136.

Shi, Y., Zhang, X., Liao, X., Lin, G. and Schuurmans, D. Protein-chemical interaction prediction via kernelized sparse learning SVM. *Pac Symp Biocomput* 2013:41-52.

Shimada, S., Kitayama, S., Lin, C.L., Patel, A., Nanthakumar, E., Gregor, P., Kuhar, M. and Uhl, G. Cloning and expression of a cocaine-sensitive dopamine transporter complementary DNA. *Science* 1991;254(5031):576-578.

Shivalingappa, P.C., Hole, R., Westphal, C.V. and Vij, N. Airway Exposure to E-Cigarette Vapors Impairs Autophagy and Induces Aggresome Formation. *Antioxid Redox Signal* 2016;24(4):186-204.

Simoes, E.S.A.C., Miranda, A.S., Rocha, N.P. and Teixeira, A.L. Renin angiotensin system in liver diseases: Friend or foe? *World J Gastroenterol* 2017;23(19):3396-3406.

Simon, T.G., Henson, J., Osganian, S., Masia, R., Chan, A.T., Chung, R.T. and Corey, K.E. Daily Aspirin Use Associated With Reduced Risk For Fibrosis Progression In Patients With Nonalcoholic Fatty Liver Disease. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* 2019;17(13):2776-2784.e2774.

Simpson, D., Braithwaite, R.A., Jarvie, D.R., Stewart, M.J., Walker, S., Watson, I.W. and Widdop, B. Screening for drugs of abuse (II): Cannabinoids, lysergic acid diethylamide, buprenorphine, methadone, barbiturates, benzodiazepines and other drugs. *Ann Clin Biochem* 1997;34 (Pt 5):460-510.

Skrabanek, L., Saini, H.K., Bader, G.D. and Enright, A.J. Computational prediction of protein-protein interactions. *Mol Biotechnol* 2008;38(1):1-17.

Slavney, P.R., Rich, G.B., Pearlson, G.D. and McHugh, P.R. Phencyclidine abuse and symptomatic mania. *Biol Psychiatry* 1977;12(5):697-700.

Smyth, G.K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004;3:Article3.

Softic, S., Cohen, D.E. and Kahn, C.R. Role of Dietary Fructose and Hepatic De Novo Lipogenesis in Fatty Liver Disease. *Dig Dis Sci* 2016;61(5):1282-1293.

Sofuoglu, M. and Sewell, R.A. Norepinephrine and stimulant addiction. *Addict Biol* 2009;14(2):119-129.

Solinas, M., Yasar, S. and Goldberg, S.R. Endocannabinoid system involvement in brain reward processes related to drug abuse. *Pharmacol Res* 2007;56(5):393-405.

Soneson, C., Love, M.I. and Robinson, M.D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* 2015;4:1521.

Sora, I., Wichems, C., Takahashi, N., Li, X.F., Zeng, Z., Revay, R., Lesch, K.P., Murphy, D.L. and Uhl, G.R. Cocaine reward models: conditioned place preference can be established in dopamine- and in serotonin-transporter knockout mice. *Proc Natl Acad Sci U S A* 1998;95(13):7699-7704.

Sorger, P.K., Allerheiligen, S.R.B., Abernethy, D.R., Altman, R.B., Brouwer, K.L.R., Califano, A., D'Argenio, D.Z., Iyengar, R., Jusko, W.J. and Lalonde, R. Quantitative and systems pharmacology in the post-genomic era: new approaches to discovering drugs and understanding therapeutic mechanisms. In.: NIH Bethesda Bethesda, MD; 2011.

Stahl, S.M. and Shayegan, D.K. The psychopharmacology of ziprasidone: receptor-binding properties and real-world psychiatric practice. (0160-6689 (Print)).

Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A. and Tyers, M. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* 2006;34(Database issue):D535-539.

Stern, A.M., Schurdak, M.E., Bahar, I., Berg, J.M. and Taylor, D.L. A Perspective on Implementing a Quantitative Systems Pharmacology Platform for Drug Discovery and the Advancement of Personalized Medicine. *J Biomol Screen* 2016;21(6):521-534.

Su, L.Y., Luo, R., Liu, Q., Su, J.R., Yang, L.X., Ding, Y.Q., Xu, L. and Yao, Y.G. Atg5- and Atg7-dependent autophagy in dopaminergic neurons regulates cellular and behavioral responses to morphine. *Autophagy* 2017;13(9):1496-1511.

Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., Lahr, D.L., Hirschman, J.E., Liu, Z., Donahue, M., Julian, B., Khan, M., Wadden, D., Smith, I.C., Lam, D., Liberzon, A., Toder, C., Bagul, M., Orzechowski, M., Enache, O.M., Piccioni, F., Johnson, S.A., Lyons, N.J., Berger, A.H., Shamji, A.F., Brooks, A.N., Vrcic, A., Flynn, C., Rosains, J., Takeda, D.Y., Hu, R., Davison, D., Lamb, J., Ardlie, K., Hogstrom, L., Greenside, P., Gray, N.S., Clemons, P.A., Silver, S., Wu, X., Zhao, W.-N., Read-Button, W., Wu, X., Haggarty, S.J., Ronco, L.V., Boehm, J.S., Schreiber, S.L., Doench, J.G., Bittker, J.A., Root, D.E., Wong, B. and Golub, T.R. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* 2017;171(6):1437--1452.

Subramanian, A., Narayan, R., Corsello, S.M., Peck, D.D., Natoli, T.E., Lu, X., Gould, J., Davis, J.F., Tubelli, A.A., Asiedu, J.K., Lahr, D.L., Hirschman, J.E., Liu, Z., Donahue, M., Julian, B., Khan, M., Wadden, D., Smith, I.C., Lam, D., Liberzon, A., Toder, C., Bagul, M., Orzechowski, M., Enache, O.M., Piccioni, F., Johnson, S.A., Lyons, N.J., Berger, A.H., Shamji, A.F., Brooks, A.N., Vrcic, A., Flynn, C., Rosains, J., Takeda, D.Y., Hu, R., Davison, D., Lamb, J., Ardlie, K., Hogstrom, L., Greenside, P., Gray, N.S., Clemons, P.A., Silver, S., Wu, X., Zhao, W.N., Read-Button, W., Wu, X., Haggarty, S.J., Ronco, L.V., Boehm, J.S., Schreiber, S.L., Doench, J.G., Bittker, J.A., Root, D.E., Wong, B. and Golub, T.R. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* 2017;171(6):1437-1452 e1417.

Sulaiman, S.A., Muhsin, N.I.A. and Jamal, R. Regulatory Non-coding RNAs Network in Non-alcoholic Fatty Liver Disease. *Front Physiol* 2019;10:279.

Sun, T., Zhou, B., Lai, L. and Pei, J. Sequence-based prediction of protein protein interaction using a deep-learning algorithm. *BMC Bioinformatics* 2017;18(1):277.

Sun, W.L., Quizon, P.M. and Zhu, J. Molecular Mechanism: ERK Signaling, Drug Addiction, and Behavioral Effects. *Prog Mol Biol Transl Sci* 2016;137:1-40.

Supuran, C.T., Scozzafava, A. and Casini, A. Carbonic anhydrase inhibitors. *Med Res Rev* 2003;23(2):146-189.

Syn, W.K., Jung, Y., Omenetti, A., Abdelmalek, M., Guy, C.D., Yang, L., Wang, J., Witek, R.P., Fearing, C.M., Pereira, T.A., Teaberry, V., Choi, S.S., Conde-Vancells, J., Karaca, G.F. and Diehl, A.M. Hedgehog-mediated epithelial-to-mesenchymal transition and fibrogenic repair in nonalcoholic fatty liver disease. *Gastroenterology* 2009;137(4):1478-1488.e1478.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., Jensen, L.J. and Mering, C.V. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;47(D1):D607-D613.

Szklarczyk, D., Santos, A., von Mering, C., Jensen, L.J., Bork, P. and Kuhn, M. STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res* 2016;44(D1):D380-384.

Taipale, T., Seppala, I., Raitoharju, E., Mononen, N., Lyytikainen, L.P., Illig, T., Waldenberger, M., Juonala, M., Hutri-Kahonen, N., Oksala, N., Kahonen, M., Raitakari, O. and Lehtimaki, T. Fatty liver is associated with blood pathways of inflammatory response, immune system activation and prothrombotic state in Young Finns Study. *Sci Rep* 2018;8(1):10358.

Taylor, D.L., Gough, A., Schurdak, M.E., Vernetti, L., Chennubhotla, C.S., Lefever, D., Pei, F., Faeder, J.R., Lezon, T.R., Stern, A.M. and Bahar, I. Harnessing Human Microphysiology Systems as Key Experimental Models for Quantitative Systems Pharmacology. *Handb Exp Pharmacol* 2019;260:327-367.

Taylor, S.B., Lewis, C.R. and Olive, M.F. The neurocircuitry of illicit psychostimulant addiction: acute and chronic effects in humans. *Subst Abuse Rehabil* 2013;4:29-43.

The UniProt, C. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 2017;45(D1):D158-D169.

Tomkins, D.M. and Sellers, E.M. Addiction and the brain: the role of neurotransmitters in the cause and treatment of drug dependence. *CMAJ* 2001;164(6):817-821.

Trettel, F., Rigamonti, D., Hilditch-Maguire, P., Wheeler, V.C., Sharp, A.H., Persichetti, F., Cattaneo, E. and MacDonald, M.E. Dominant phenotypes produced by the HD mutation in STHdh(Q111) striatal cells. *Hum Mol Genet* 2000;9(19):2799-2809.

Tritsch, N.X., Ding, J.B. and Sabatini, B.L. Dopaminergic neurons inhibit striatal output through non-canonical release of GABA. *Nature* 2012;490(7419):262-266.

Troisi 2nd, J.R., Critchfield, T.S. and Griffiths, R.R. Buspirone and lorazepam abuse liability in humans: behavioral effects, subjective effects and choice. *Behav Pharmacol* 1993;4(3):217-230.

Tsay, M.E., Procopio, G., Anderson, B.D. and Klein-Schwartz, W. Abuse and Intentional Misuse of Promethazine Reported to US Poison Centers: 2002 to 2012. *J Addict Med* 2015;9(3):233-237.

Tsvetkov, A.S., Arrasate, M., Barmada, S., Ando, D.M., Sharma, P., Shaby, B.A. and Finkbeiner, S. Proteostasis of polyglutamine varies among neurons and predicts neurodegeneration. *Nat Chem Biol* 2013;9(9):586-592.

Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szigyarto, C.A., Odeberg, J., Djureinovic, D., Takanen, J.O., Hober, S., Alm, T., Edqvist, P.H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J.M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zwahlen, M., von Heijne, G., Nielsen, J. and Ponten, F. Proteomics. Tissue-based map of the human proteome. *Science* 2015;347(6220):1260419.

Valenti, L., Mendoza, R.M., Rametta, R., Maggioni, M., Kitajewski, C., Shawber, C.J. and Pajvani, U.B. Hepatic notch signaling correlates with insulin resistance and nonalcoholic fatty liver disease. *Diabetes* 2013;62(12):4052-4062.

Van Herck, M.A., Weyler, J., Kwanten, W.J., Dirinck, E.L., De Winter, B.Y., Francque, S.M. and Vonghia, L. The Differential Roles of T Cells in Non-alcoholic Fatty Liver Disease and Obesity. *Front Immunol* 2019;10:82.

van Laarhoven, T., Nabuurs, S.B. and Marchiori, E. Gaussian interaction profile kernels for predicting drug-target interaction. *Bioinformatics* 2011;27(21):3036-3043.

Vernetti, L., Gough, A., Baetz, N., Blutt, S., Broughman, J.R., Brown, J.A., Foulke-Abel, J., Hasan, N., In, J., Kelly, E., Kovbasnjuk, O., Repper, J., Senutovitch, N., Stabb, J., Yeung, C., Zachos, N.C., Donowitz, M., Estes, M., Himmelfarb, J., Truskey, G., Wikswo, J.P. and Taylor, D.L. Functional Coupling of Human Microphysiology Systems: Intestine, Liver, Kidney Proximal Tubule, Blood-Brain Barrier and Skeletal Muscle. *Sci Rep* 2017;7:42296.

Vernetti, L.A., Senutovitch, N., Boltz, R., DeBiasio, R., Shun, T.Y., Gough, A. and Taylor, D.L. A human liver microphysiology platform for investigating physiology, drug safety, and disease models. *Exp Biol Med (Maywood)* 2016;241(1):101-114.

Volkow, N.D., Fowler, J.S. and Wang, G.-J. The addicted human brain: insights from imaging studies. *The Journal of clinical investigation* 2003;111(10):1444-1451.

Volkow, N.D., Koob, G.F. and McLellan, A.T. Neurobiologic Advances from the Brain Disease Model of Addiction. *N Engl J Med* 2016;374(4):363-371.

Volkow, N.D. and Morales, M. The Brain on Drugs: From Reward to Addiction. *Cell* 2015;162(4):712-725.

Volkow, N.D., Wang, G.J., Fischman, M.W., Foltin, R.W., Fowler, J.S., Abumrad, N.N., Vitkun, S., Logan, J., Gatley, S.J., Pappas, N., Hitzemann, R. and Shea, C.E. Relationship between subjective effects of cocaine and dopamine transporter occupancy. *Nature* 1997;386(6627):827-830.

Wagner, B.K. and Schreiber, S.L. The Power of Sophisticated Phenotypic Screening and Modern Mechanism-of-Action Methods. *Cell Chem Biol* 2016;23(1):3-9.

Walsh, S.L., Nuzzo, P.A., Lofwall, M.R. and Holtman, J.R., Jr. The relative abuse liability of oral oxycodone, hydrocodone and hydromorphone assessed in prescription opioid abusers. *Drug Alcohol Depend* 2008;98(3):191-202.

Wang, H., Huang, H., Ding, C. and Nie, F. Predicting protein-protein interactions from multimodal biological data sources via nonnegative matrix tri-factorization. *J Comput Biol* 2013;20(4):344-358.

Wang, S., Song, K., Srivastava, R., Dong, C., Go, G.W., Li, N., Iwakiri, Y. and Mani, A. Nonalcoholic fatty liver disease induced by noncanonical Wnt and its rescue by Wnt3a. *FASEB J* 2015;29(8):3436-3445.

Wang, X., Shen, Y., Wang, S., Li, S., Zhang, W., Liu, X., Lai, L., Pei, J. and Li, H. PharmMapper 2017 update: a web server for potential drug target identification with a comprehensive target pharmacophore database. *Nucleic Acids Res* 2017;45(W1):W356-W360.

Wang, X., Zhu, S., Drozda, M., Zhang, W., Stavrovskaya, I.G., Cattaneo, E., Ferrante, R.J., Kristal, B.S. and Friedlander, R.M. Minocycline inhibits caspase-independent and -dependent mitochondrial cell death pathways in models of Huntington's disease. *Proc Natl Acad Sci U S A* 2003;100(18):10483-10487.

Wang, X., Zhu, S., Pei, Z., Drozda, M., Stavrovskaya, I.G., Del Signore, S.J., Cormier, K., Shimony, E.M., Wang, H., Ferrante, R.J., Kristal, B.S. and Friedlander, R.M. Inhibitors of cytochrome c release with therapeutic potential for Huntington's disease. *J Neurosci* 2008;28(38):9473-9485.

Wang, Y., You, Z., Li, X., Chen, X., Jiang, T. and Zhang, J. PCVMZM: Using the Probabilistic Classification Vector Machines Model Combined with a Zernike Moments Descriptor to Predict Protein-Protein Interactions from Protein Sequences. *International journal of molecular sciences* 2017;18(5):1029.

Wang, Y., You, Z.-H., Yang, S., Li, X., Jiang, T.-H. and Zhou, X. A High Efficient Biological Language Model for Predicting Protein–Protein Interactions. *Cells* 2019;8(2):122.

Wang, Y. and Zeng, J. Predicting drug-target interactions using restricted Boltzmann machines. *Bioinformatics* 2013;29(13):i126-134.

Wang, Y.B., You, Z.H., Li, X., Jiang, T.H., Chen, X., Zhou, X. and Wang, L. Predicting protein-protein interactions from protein sequences by a stacked sparse autoencoder deep neural network. *Mol Biosyst* 2017;13(7):1336-1344.

Weber, A., Casini A Fau - Heine, A., Heine A Fau - Kuhn, D., Kuhn D Fau - Supuran, C.T., Supuran Ct Fau - Scozzafava, A., Scozzafava A Fau - Klebe, G. and Klebe, G. Unexpected nanomolar inhibition of carbonic anhydrase by COX-2-selective celecoxib: new pharmacological opportunities due to related binding site recognition. (0022-2623 (Print)).

White, C.M. and Lee, J.Y. The impact of turmeric or its curcumin extract on nonalcoholic fatty liver disease: a systematic review of clinical trials. *Pharm Pract (Granada)* 2019;17(1):1350.

Widimsky, J. and Sirotiakova, J. Efficacy and tolerability of rilmenidine compared with isradipine in hypertensive patients with features of metabolic syndrome. *Curr Med Res Opin* 2006;22(7):1287-1294.

Williams, M.J. and Adinoff, B. The role of acetylcholine in cocaine addiction. *Neuropsychopharmacology* 2008;33(8):1779-1797.

Winslow, B.T., Voorhees, K.I. and Pehl, K.A. Methamphetamine abuse. *Am Fam Physician* 2007;76(8):1169-1174.

Winstock, A.R., Kaar, S. and Borschmann, R. Dimethyltryptamine (DMT): prevalence, user characteristics and abuse liability in a large global sample. *J Psychopharmacol* 2014;28(1):49-54.

Wise, R.A. and Koob, G.F. The development and maintenance of drug addiction. *Neuropsychopharmacology* 2014;39(2):254-262.

Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., Pon, A., Knox, C. and Wilson, M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2017.

Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., Pon, A., Knox, C. and Wilson, M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;46(D1):D1074-D1082.

Wood, J.D. and Galligan, J.J. Function of opioids in the enteric nervous system. *Neurogastroenterol Motil* 2004;16 Suppl 2:17-28.

Wooden, B., Goossens, N., Hoshida, Y. and Friedman, S.L. Using Big Data to Discover Diagnostics and Therapeutics for Gastrointestinal and Liver Diseases. *Gastroenterology* 2017;152(1):53-67 e53.

Woody, G.E., O'Brien, C.P. and Greenstein, R. Misuse and abuse of diazepam: an increasingly common medical problem. *Int J Addict* 1975;10(5):843-848.

Wouters, O.J., McKee, M. and Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA* 2020;323(9):844-853.

Wright, M.H. and Sieber, S.A. Chemical proteomics approaches for identifying the cellular targets of natural products. *Nat Prod Rep* 2016;33(5):681-708.

Xenarios, I., Salwinski, L., Duan, X.J., Higney, P., Kim, S.M. and Eisenberg, D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res* 2002;30(1):303-305.

Xia, L.Y., Yang, Z.Y., Zhang, H. and Liang, Y. Improved Prediction of Drug-Target Interactions Using Self-Paced Learning with Collaborative Matrix Factorization. *J Chem Inf Model* 2019;59(7):3340-3351.

Xia, Z., Wu, L.Y., Zhou, X. and Wong, S.T. Semi-supervised drug-protein interaction prediction from heterogeneous biological spaces. *BMC Syst Biol* 2010;4 Suppl 2:S6.

Xiao, N., Cao, D.S., Zhu, M.F. and Xu, Q.S. protr/ProtrWeb: R package and web server for generating various numerical representation schemes of protein sequences. *Bioinformatics* 2015;31(11):1857-1859.

Xiao, X., Min, J.L., Lin, W.Z., Liu, Z., Cheng, X. and Chou, K.C. iDrug-Target: predicting the interactions between drug compounds and target proteins in cellular networking via benchmark dataset optimization approach. *J Biomol Struct Dyn* 2015;33(10):2221-2233.

Xie, X.-Q., Wang, L., Wang, J., Xie, Z., Yang, P. and Ouyang, Q. In silico chemogenomics knowledgebase and computational system neuropharmacology approach for cannabinoid drug research. In, *Neuropathology of Drug Addictions and Substance Misuse*. Elsevier; 2016. p. 183-195.

Xie, X.Q., Wang, L., Liu, H., Ouyang, Q., Fang, C. and Su, W. Chemogenomics knowledgebased polypharmacology analyses of drug abuse related G-protein coupled receptors and their ligands. *Front Pharmacol* 2014;5:3.

Xu, H. and Zhuang, X. Atypical antipsychotics-induced metabolic syndrome and nonalcoholic fatty liver disease: a critical review. *Neuropsychiatr Dis Treat* 2019;15:2087-2099.

Xu, K.Y., Shameem, R. and Wu, S. Risk of hyperglycemia attributable to everolimus in cancer patients: A meta-analysis. *Acta Oncol* 2016;55(9-10):1196-1203.

Xu, Y., Hu, W., Chang, Z., Duanmu, H., Zhang, S., Li, Z., Li, Z., Yu, L. and Li, X. Prediction of human protein-protein interaction by a mixed Bayesian model and its application to exploring underlying cancer-related pathway crosstalk. *J R Soc Interface* 2011;8(57):555-567.

Xuan, Z., Li, J., Yu, J., Feng, X., Zhao, B. and Wang, L. A Probabilistic Matrix Factorization Method for Identifying lncRNA-disease Associations. *Genes (Basel)* 2019;10(2).

Yamamoto, I., Watanabe, K., Matsunaga, T., Kimura, T., Funahashi, T. and Yoshimura, H. Pharmacology and toxicology of major constituents of marijuana—on the metabolic activation of cannabinoids and its mechanism. *Journal of Toxicology: Toxin Reviews* 2003;22(4):577-589.

Yamamura, T., Ohsaki, Y., Suzuki, M., Shinohara, Y., Tatematsu, T., Cheng, J., Okada, M., Ohmiya, N., Hirooka, Y., Goto, H. and Fujimoto, T. Inhibition of Niemann-Pick-type C1-like1 by ezetimibe activates autophagy in human hepatocytes and reduces mutant alpha1-antitrypsin Z deposition. *Hepatology* 2014;59(4):1591-1599.

Yamanishi, Y., Araki, M., Gutteridge, A., Honda, W. and Kanehisa, M. Prediction of drug-target interaction networks from the integration of chemical and genomic spaces. *Bioinformatics* 2008;24(13):i232-240.

Yamanishi, Y., Kotera, M., Kanehisa, M. and Goto, S. Drug-target interaction prediction from chemical, genomic and pharmacological data in an integrated framework. *Bioinformatics* 2010;26(12):i246-254.

Yamanishi, Y., Kotera, M., Moriya, Y., Sawada, R., Kanehisa, M. and Goto, S. DINIES: drug-target interaction network inference engine based on supervised analysis. *Nucleic Acids Res* 2014;42(Web Server issue):W39-45.

Yan, Z., Miao, X., Zhang, B. and Xie, J. p53 as a double-edged sword in the progression of non-alcoholic fatty liver disease. *Life Sci* 2018;215:64-72.

Yang, J., Li, Z., Fan, X. and Cheng, Y. Drug-disease association and drug-repositioning predictions in complex diseases using causal inference-probabilistic matrix factorization. *J Chem Inf Model* 2014;54(9):2562-2569.

Yang, L., Xia, J.F. and Gui, J. Prediction of protein-protein interactions from protein sequence using local descriptors. *Protein Pept Lett* 2010;17(9):1085-1090.

Yano, H., Baranov, S.V., Baranova, O.V., Kim, J., Pan, Y., Yablonska, S., Carlisle, D.L., Ferrante, R.J., Kim, A.H. and Friedlander, R.M. Inhibition of mitochondrial protein import by mutant huntingtin. *Nat Neurosci* 2014;17(6):822-831.

Yao, J., Ma, C., Gao, W., Liang, J., Liu, C., Yang, H., Yan, Q. and Wen, Q. Fentanyl induces autophagy via activation of the ROS/MAPK pathway and reduces the sensitivity of cisplatin in lung cancer cells. *Oncol Rep* 2016;36(6):3363-3370.

Yao, Y., Du, X., Diao, Y. and Zhu, H. An integration of deep learning with feature embedding for protein-protein interaction prediction. *PeerJ* 2019;7:e7126.

Yee, M.L. and Tan, H.H. Use of everolimus in liver transplantation. *World J Hepatol* 2017;9(23):990-1000.

You, Z.H., Lei, Y.K., Zhu, L., Xia, J. and Wang, B. Prediction of protein-protein interactions from amino acid sequences with ensemble extreme learning machines and principal component analysis. *BMC Bioinformatics* 2013;14 Suppl 8:S10.

Younossi, Z., Anstee, Q.M., Marietti, M., Hardy, T., Henry, L., Eslam, M., George, J. and Bugianesi, E. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol* 2018;15(1):11-20.

Younossi, Z.M., Ratziu, V., Loomba, R., Rinella, M., Anstee, Q.M., Goodman, Z., Bedossa, P., Geier, A., Beckebaum, S., Newsome, P.N., Sheridan, D., Sheikh, M.Y., Trotter, J., Knapple, W., Lawitz, E., Abdelmalek, M.F., Kowdley, K.V., Montano-Loza, A.J., Boursier, J., Mathurin, P., Bugianesi, E., Mazzella, G., Olveira, A., Cortez-Pinto, H., Graupera, I., Orr, D., Gluud, L.L.,

Dufour, J.F., Shapiro, D., Campagna, J., Zaru, L., MacConell, L., Shringarpure, R., Harrison, S., Sanyal, A.J. and Investigators, R.S. Obeticholic acid for the treatment of non-alcoholic steatohepatitis: interim analysis from a multicentre, randomised, placebo-controlled phase 3 trial. *Lancet* 2019;394(10215):2184-2196.

Yousefi, O.S., Wilhelm, T., Maschke-Neuss, K., Kuhny, M., Martin, C., Molderings, G.J., Kratz, F., Hildenbrand, B. and Huber, M. The 1,4-benzodiazepine Ro5-4864 (4-chlorodiazepam) suppresses multiple pro-inflammatory mast cell effector functions. *Cell Commun Signal* 2013;11(1):13.

Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Giron, C.G., Gil, L., Gordon, L., Haggerty, L., Haskell, E., Hourlier, T., Izuogu, O.G., Janacek, S.H., Juettemann, T., To, J.K., Laird, M.R., Lavidas, I., Liu, Z., Loveland, J.E., Maurel, T., McLaren, W., Moore, B., Mudge, J., Murphy, D.N., Newman, V., Nuhn, M., Ogeh, D., Ong, C.K., Parker, A., Patricio, M., Riat, H.S., Schuilenburg, H., Sheppard, D., Sparrow, H., Taylor, K., Thormann, A., Vullo, A., Walts, B., Zadissa, A., Frankish, A., Hunt, S.E., Kostadima, M., Langridge, N., Martin, F.J., Muffato, M., Perry, E., Ruffier, M., Staines, D.M., Trevanion, S.J., Aken, B.L., Cunningham, F., Yates, A. and Flicek, P. Ensembl 2018. *Nucleic Acids Res* 2018;46(D1):D754-D761.

Zhang, Q.C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C.A., Bisikirska, B., Lefebvre, C., Accili, D., Hunter, T., Maniatis, T., Califano, A. and Honig, B. Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature* 2012;490(7421):556-560.

Zhang, S.W., Hao, L.Y. and Zhang, T.H. Prediction of protein-protein interaction with pairwise kernel support vector machine. *Int J Mol Sci* 2014;15(2):3220-3233.

Zhao, N.J., Liao, M.J., Wu, J.J. and Chu, K.X. Curcumin suppresses Notch1 signaling: Improvements in fatty liver and insulin resistance in rats. *Mol Med Rep* 2018;17(1):819-826.

Zhao, X., Kwan, J.Y.Y., Yip, K., Liu, P.P. and Liu, F.F. Targeting metabolic dysregulation for fibrosis therapy. *Nat Rev Drug Discov* 2020;19(1):57-75.

Zhou, H., Gao, M. and Skolnick, J. Comprehensive prediction of drug-protein interactions and side effects for the human proteome. *Sci Rep* 2015;5:11090.

Zhou, Y.Z., Gao, Y. and Zheng, Y.Y. Prediction of Protein-Protein Interactions Using Local Description of Amino Acid Sequence. In: Zhou, M. and Tan, H., editors, *Advances in Computer Science and Education Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 254-262.

Zhu, J., Huang Jw Fau - Tseng, P.-H., Tseng Ph Fau - Yang, Y.-T., Yang Yt Fau - Fowble, J., Fowble J Fau - Shiau, C.-W., Shiau Cw Fau - Shaw, Y.-J., Shaw Yj Fau - Kulp, S.K., Kulp Sk Fau - Chen, C.-S. and Chen, C.S. From the cyclooxygenase-2 inhibitor celecoxib to a novel class of 3-phosphoinositide-dependent protein kinase-1 inhibitors. (0008-5472 (Print)).

Zhu, J., Spencer, T.J., Liu-Chen, L.Y., Biederman, J. and Bhide, P.G. Methylphenidate and mu opioid receptor interactions: a pharmacological target for prevention of stimulant abuse. *Neuropharmacology* 2011;61(1-2):283-292.

Zhu, M., Xu, Y., Wang, H., Shen, Z., Xie, Z., Chen, F., Gao, Y., Chen, X., Zhang, Y., Wu, Q., Li, X., Yu, J., Luo, H. and Wang, K. Heroin Abuse Results in Shifted RNA Expression to Neurodegenerative Diseases and Attenuation of TNFalpha Signaling Pathway. *Sci Rep* 2018;8(1):9231.

Zuccato, C., Valenza, M. and Cattaneo, E. Molecular mechanisms and potential therapeutical targets in Huntington's disease. *Physiol Rev* 2010;90(3):905-981.