AddictGene: An integrated knowledge base for differentially expressed genes associated with addictive substance

Leisheng Shi\(^{a,b,1}\), Yan Wang\(^{a,1}\), Chong Li\(^{a,b}\), Kunlin Zhang\(^{a}\), Quansheng Du\(^{c}\), Mei Zhao\(^{a,b,*}\)

\(^{a}\) CAS Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China
\(^{b}\) University of Chinese Academy of Sciences, Beijing 101408, China
\(^{c}\) Department of Life Sciences, National Natural Science Foundation of China, Beijing 100085, China

**A R T I C L E   I N F O**

Article history:
Received 29 October 2020
Received in revised form 8 April 2021
Accepted 9 April 2021
Available online 19 April 2021

Keywords:
Drug dependence
Gene expression
Drug abuse
Database
Addiction

**A B S T R A C T**

Addiction, a disorder of maladaptive brain plasticity, is associated with changes in numerous gene expressions. Nowadays, high-throughput sequencing data on addictive substance-induced gene expression have become widely available. A resource for comprehensive annotation of genes that show differential expression in response to commonly abused substances is necessary. So, we developed AddictGene by integrating gene expression, gene-gene interaction, gene-drug interaction and epigenetic regulatory annotation for over 70,156 items of differentially expressed genes associated with 7 commonly abused substances, including alcohol, nicotine, cocaine, morphine, heroin, methamphetamine, and amphetamine, across three species (human, mouse, rat). We also collected 1,141 addiction-related experimentally validated genes by techniques such as RT-PCR, northern blot and in situ hybridization. The easy-to-use web interface of AddictGene (http://159.226.67.237/sun/addictgedb/) allows users to search and browse multidimensional data on DEGs of their interest: 1) detailed gene-specific information extracted from the original studies; 2) basic information about the specific gene extracted from NCBI; 3) SNP associated with substance dependence and other psychiatry disorders; 4) expression alteration of specific gene in other psychiatric disorders; 5) expression patterns of interested gene across 31 primary and 54 secondary human tissues; 6) functional annotation of interested gene; 7) epigenetic regulators involved in the alteration of specific genes, including histone modifications and DNA methylation; 8) protein–protein interaction for functional linkage with interested gene; 9) drug-gene interaction for potential druggability. AddictGene offers a valuable repository for researchers to study the molecular mechanisms underlying addiction, and might provide valuable insights into potential therapies for drug abuse and relapse.

\(\triangleright\) 2021 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Addiction, a chronic, relapsing brain disease, is characterized by loss of control over harmful substances intake and compulsive substance consumption despite serious adverse consequences [1]. Alteration of gene expression has been linked to dependence, withdrawal, and relapse of addictive substance-dependent individuals in both animal and human studies [2–4]. Earlier studies revealed numerous expression-altered genes related to addictive substance-dependence including immediate-early genes, transcription factors, and various neurotransmitter genes [5]. For example, immediate-early genes include members of the Fos family (Fos, FosB), the Jun family (c-Jun, JunB, and JunD) and Zif268 (Egr1) overexpressed transiently or permanently in response to a wide range of addictive substances. Transcription factors (CREB, NF-kb) could exert a crucial role in addiction development by influencing the expression of numerous genes simultaneously [2]. Additionally, receptors of dopamine and NMDA, neurotrophic factors, and genes involved in the circadian rhythm were also demonstrated to contribute to the development of addiction [5]. Therefore, identifying the genes responding to addictive substance exposure and uncovering their regulators will not only improve the understanding of the mechanisms underlying addiction, but will also provide valuable insights into potential therapies for addiction and relapse.

Following the development of high-throughput transcriptome profiling technologies, gene expression data relating to changes...
induced by addictive substances have become widely available. However, a resource for comprehensive deciphering and annotation of genes that show differential expression in response to several commonly abused dependence substances is still lacking. To the best of our knowledge, one such database, ERGR database [6], was built for depositing genes related to just one kind of addictive substance, ethanol. It was developed nearly 10 years ago and hosts ethanol-related genes curated from multiple public data sets including linkage, association and microarray-based gene expression. During the preceding decade, gene expression data relating to the use of addictive substances have accumulated rapidly. Thus, there is a growing need to build a platform to collect and integrate the current multi-omic data from multiple studies for convenient retrieval by researchers and clinicians.

To fulfill this need, here, we have developed AddictGene (http://159.226.67.237/sun/addictgedb/), which integrated gene expression, gene–gene interaction, gene–drug interaction and regulatory annotation for over 33,821 items of differentially expressed genes associated with 7 commonly abused substances across three species (human, mouse, rat) from 205 publications. The database presents an easy-to-use web interface that allows users to search, visualize and account for addiction-related genes. In addition, it provides researchers and clinicians substantial convenience to explore the underlying regulatory mechanism for alteration of specific addiction-related genes and drug–gene interaction for potential druggability.

2. Methods and analysis:

2.1. Data collection and curation

To create a list of DEGs (differentially expressed genes) relevant to addiction, a comprehensive search was performed for drug-related expression studies. We searched the PubMed database (http://www.ncbi.nlm.nih.gov/pubmed) with the following query terms: addiction [Title/Abstract] AND the name of a specific addictive substances such as heroin [Title/Abstract] OR Psychoactive substances[Title/Abstract] OR Drug[Title/Abstract] AND (addiction [Title/Abstract] OR dependence[Title/Abstract] OR abuse[Title/Abstract] OR use[Title/Abstract]) AND gene [Title/Abstract] OR gene expression [Title/Abstract] OR PCR [Title/Abstract] OR RNA-seq [Title/Abstract] OR microarray [Title/Abstract] AND (associate* [Title/Abstract]). Consequently, more than 500 publications dating from 2000 to 2020 were obtained. Abstracts of these publications were manually screened on the basis of the following inclusion criteria: 1) original manuscript written in English; 2) gene expression studies conducted in humans or animal models related to addiction; 3) DEGs identified by PCR, northern blot, in situ hybridization, microarray or RNA-seq; 4) unambiguous phenotype (e.g. addiction or dependence, craving, drug-seeking, withdrawal etc.) assessed on specific scales by trained interviewers, psychologists or researchers and drug usage is not limited to one kind of drug used concurrently; 5) results satisfying the significance threshold of p-value < 0.05.

The full text of each eligible article was read carefully, and detailed information pertaining to each gene investigated in the study was extracted manually. For each gene described in a study, detailed information including the method employed, the drug administered, the model organism, the phenotype, and the publication. Consequently, the AddictGene database contains 70,156 items of dependence substances-associated DEGs corresponding to seven kinds of substances from three species and 341 eligible publications. From the literature, we collected 1,141 addiction-related genes, whose alterations were validated by techniques such as RT-PCR (Reverse transcription PCR), northern blot and in situ hybridization.

2.2. Gene annotation and analysis

To provide an informative ‘gene card’ for each DEG relevant to addiction, its annotation was designed to include basic information, gene–related disease, tissue expression patterns, functional and regulatory annotation, protein–protein interaction and drug–gene interaction. We downloaded the gene annotation files from the NCBI’s reference sequence (RefSeq) database (ftp://ftp.ncbi.nlm.nih.gov/refseq/) and extracted the annotation information for our database. For each gene, we retrieved basic information such as gene symbol, its aliases, and full name. Similarly, we extracted gene ontology (GO) and pathway annotations from the gene2go files downloaded from the GO website (http://www.geneontology.org/) and BioSystems (https://www.ncbi.nlm.nih.gov/biosystems), respectively.

To facilitate a comprehensive understanding of identified DEGs, we downloaded gene expression data of other psychiatric disorders from Gene Expression Omnibus of NCBI (https://www.ncbi.nlm.nih.gov/geo/), and tissue expression pattern of each gene from Genotype-Tissue Expression (GTEx) (http://www.gtexportal.org/). In addition, we extracted information about potential drug targets of gene of interest and their protein–protein interaction network from DGIdb (http://www.dgidb.org/) and InBio Map (https://www.intomics.com/inbio/map.html), respectively. To provide insight into the relationship between certain gene and genetic vulnerability to other psychiatric and neurodegenerative disease, we integrated genome-wide association study (GWAS) data of other psychiatric disorders and neurodegenerative disease.

2.3. Analysis of epigenomic regulation

To identify the potential epigenetic regulators involved in the expression of DEGs, we integrated epigenomic regulatory information into our current database. Literature-based methylation alterations of specific genes associated with drug dependent were retrieved from PubMed database (http://www.ncbi.nlm.nih.gov/pubmed), and methylation levels of specific gene’s promoter were also retrieved from MethBank (http://bigd.big.ac.cn/methbank). In addition, raw data of 180 embryonic chromatin immunoprecipitation sequencing (ChiP-seq) (Supplementary Table S1) datasets were downloaded from GEO of NCBI. After removing sequencing adapters and low-quality sequences, all ChiP-seq data were mapped to the hg38 genome for humans or mm10 genome for mouse using SpeedSeq (v0.1.2) [7]. Then, the bigwig files for JBrowse visualization were generated from BAM files using ‘bamCoverage’ from deepTools [8] with parameters ‘–ignoreDuplication cates=normalizeUsingRPKM –skipNonCoveredRegions –binSize 25 –ignoreForNormalization chrX chrY’.

2.4. Database construction and content organization

AddictGene was developed to have a user-friendly web interface supporting versatile browsing and search functionalities. All data were stored and managed in the MySQL relational database. The users may access the data or perform advanced search freely through the web interface (http://159.226.67.237/sun/addictgedb/). The flow charts detailing the search, screening and mining of eligible studies have been summarized in Fig. 1. After searching the literature comprehensively and careful screening, we obtained 205 eligible papers for further data extraction.
2.5. Database usage and access

AddictGene provides two approaches for searching the data. First, there is a quick search box on the homepage for conducting a search by entering gene symbols or the names of addictive substances. Second, AddictGene provides an advanced search page, on which users may combine different search terms (e.g. gene symbol, addictive substance, species, techniques, expression fold change and adjusted \( p \)-value) for a more defined search. The search results page includes detailed information (e.g. addictive substance, species, techniques, treatment details, tissue/cell type, expression fold change, \( p \)-values) extracted from the original studies (Fig. 2).

In addition, AddictGene provides researchers with a powerful search engine and a user-friendly interface to access different data types and links (Fig. 2). The search results consist of 9 categories: 1) detailed gene-specific information extracted from the original studies; 2) basic information about the specific gene extracted from NCBI; 3) SNP associated with substance dependence and other psychiatry disorders; 4) expression alteration of specific gene in other psychiatric disorders; 5) expression patterns of interested gene across 31 primary and 54 secondary human tissues; 6) functional annotation of interested gene; 7) two kinds of potential epigenetic regulators involved in the alteration of specific genes, including histone modifications and DNA methylation; 8) protein–protein interaction for functional linkage with interested gene; 9) drug-gene interaction for potential druggability. Thus, AddictGene provides a range of data query tools that users can employ to investigate the detailed annotation of the identified DEGs.

3. Results

3.1. Data integration and analyses

It is suggested that common biological mechanisms underlie addiction to many harmful substances since they all cause certain shared effects upon acute or chronic exposure. In the current study, based on our curated data, we defined the orthologous DEGs of the three species shared by multiple dependence substances as “common genes”. Upon analyzing data from chronic substance treatment and recovery programs or person with substance use disorder, we found that 61 DEGs were shared by at least four kinds of substances (Fig. 3A). On analyzing data of acute drug treatment, we found that 32 DEGs were shared by at least two kinds of drugs. On analyzing drug withdrawal data, we found that 46 DEGs were shared by at least two kinds of drugs. Pathway analysis of DEGs responding to different drug treatment was performed and top10 was as listed (Table 1). DEG responding to chronic drug treatment were enriched in harmful substances addiction related pathways, cAMP, MAPK, neurotrophin, oxytocin, and GnRH signaling pathways. DEGs responding to acute drug treatment were enriched in amphetamine addiction and immune related pathways such as TNF and IL-17 signaling pathway. And DEGs responding to drug withdrawal were enriched in Estrogen, Ras, Toll-like receptor, JAK-STAT, ErbB signaling pathway, and Nervous system pathways such as Huntington disease, Cholinergic synapse, Long-term potentiation and et al. Besides we did pathway enrichment analyses of DEGs shared by multiple drugs. The top ten enriched pathways mainly are Substance dependence pathways and Nervous system pathways (Fig. 3B). Protein-protein interaction networks of common genes revealed several well-recognized addiction-associated genes including FOS, JUN, EGR1 and BDNF play crucial role in addiction-related behaviors (Fig. 3C).

3.2. Examples of implication

Although different addictive substances has distinct chemical structures and elicit different pharmacological effects, they may share common neural and molecular pathways. Searching for such common mechanisms could provide insights into the development of generalized treatment for drug addiction [9]. This is especially important since persons with substance use disorder frequently take more than one kind of drug concurrently. Alcohol abuse is
especially co-morbid with the abuse of many other drugs [10]. Thus, the discovery of common mechanisms will benefit drug abusers who take multiple addictive substances. Based on our curated data including addicts take one or more than one kind of drug samples, we found that a few DEGs occur frequently among data relating to chronic abuse, acute abuse, or drug withdrawal. These common genes included the ones that have been studied well in drug addiction such as immediate-early genes (FOS, ARC, JUN, FOSB, EGR1, EGR2) [11–16], CARTPT [17–20], PDYN [21–23], TH [24–28], BDNF [29–34] and NPY [35–39]. Among these common genes, CARTPT is the most noteworthy since it expresses differentially upon the chronic abuse of nicotine, morphine, alcohol, methamphetamine, heroin, or cocaine. GWAS result revealed the significant contribution of CARTPT to suicide attempts in bipolar disorder (Fig. 3D) [40], and amphetamine was reported to be an agonist of CARTPT retrieved from the drug-gene interaction of DGIdb (Fig. 3E) [41]. Histone H3K4me3 modification activities, a defining mark of gene activation [42,43], were observed around CARTPT during embryonic development process, which provides a potential epigenetic regulator involved in the alteration of specific gene during the development of addiction. In addition, the differential expression of PDYN, TH, BDNF, and NTRK2 was induced by chronic abuse of five different kinds of drugs. Also, the differential expression of FOS, EGR1, EGR2, NPY and CREB was induced by withdrawal of three different kinds of substance. The differential expression of FOS, EGR2, EGR4, SGK1, PLIN4, ARROCD2 and CDKN1A was induced by acute treatment of three or four different kinds of substance. Numerous genetic evidences showed DRD2 and COMT, another two common gene we identified, were associated with various addictive substances [44–50]. However, the underlying mechanisms have not been studied well. Furthermore, although gene expression changes may depend on the length of exposure or withdrawal from the addictive substances [51], we found that a few genes including FOS, ARC, EGR1 and FOSB differentially expressed irrespective of the stages of addiction. Thus, these common genes might underlie a unitary basis for neurobiological dysfunction of addiction.

3.3. Future plan

We will continuously update and improve AddictGene in the next few years. DEG and histone modification data will be increased items by searching from PubMed every year. Other data will be extended when the public database we used is updated. In addition, to better understand the linkage between addiction and other psychiatric disorders, we will collect more information including gene expression, epigenetic regulation data of depression, bipolar disorder, and schizophrenia. This will allow AddictGene as a hub for the users to integrate the genes.

4. Discussion

In the past few years, ERGR provided gene information for scientists to understand pathogenesis of drug addiction. Differentially expressed genes and the associated pathways are important to
know their molecular mechanism and function linked to the physiological and behavioral changes. AddictGene we developed offered genetic and transcriptomic functional information on addiction-related DEGs. Users can also discover potential druggability targets to the genes of interest and observe their normal expression profiles in primary tissues. Considering the comorbidity between addiction and other psychiatric disorders/neurodegenerative disease, multiple psychiatric disorders/neurodegenerative disease-related GWAS and transcriptomic data were also integrated into AddictGene. To further provide insight into the potential epigenetic regulators involved in the expression of specific DEGs, we integrated epigenomic regulatory information including histone modifications and DNA methylation into our current database.

It was proposed that different drugs of abuse share similar addictive actions and rewarding [52]. In current study, common pathways of Substance dependence pathways and Nervous system pathways and common differentially expressed genes of FOS, ARC, EGR1 and FOSB shared by multiple drugs were identified irrespective of the stages of addiction. FOS, FOSB and ARC are all located in Amphetamine addiction pathway. Thereinto, FOSB was reported to initiate a multitude of molecular pathways via the formation of AP-1 complexes by interaction with JUN and FOS gene [53]. But molecular roles of FOSB of the functional purpose acting on DNA and histones is still unclear. Elucidation of mechanisms of common genes and associated molecular pathways underlying shared rewarding and addictive actions may help the development of effective treatments for a wide range of addictive disorders, especially persons with substance use disorder take more than one kind of drug concurrently.

As a useful repository, AddictGene will help researchers to study the molecular mechanisms underlying addiction, and provide valuable insights into potential therapies for drug harmful use and relapse.

5. Authors’ contributions

SLS and WY carried out database construction, and data integration and analyses; LC collected and curated data; ZKL provided the conception and design of the study; DQS revised article critically for important intellectual content; ZM designed study, interpreted...
data and drafted the article. All authors read and approved the final manuscript.

CRediT authorship contribution statement

Leisheng Shi: Data curation, Visualization. Yan Wang: Funding acquisition, Writing - original draft, Writing - review & editing. Chong Li: Formal analysis, Investigation. Kunlin Zhang: Formal analysis, Methodology. Quansheng Du: Supervision, Mei Zhao: Conceptualization, Supervision, Writing - original draft, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 91132728, 31741062, 61401459); and Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences.

Appendix A Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.04.027.

References

[1] Pol Bodetto S, Carouge D, Fonteneau M, Dietrich JB, Zwiller J, et al. Cocaine represses protein phosphatase-1-beta through DNA methylation and Methyl-CpG Binding Protein-2 recruitment in adult rat brain. Neuropharmacology 2013;73:31–40.
[2] Chao J, Nestler EJ. Molecular neurobiology of drug addiction. Annu Rev Med 2004;55:113–32.
[3] Dymarska SE, Malecka RC, Nestler EJ. Neural mechanisms of addiction: the role of reward-related learning and memory. Annu Rev Neurosci 2006;29:565–98.
[4] Nestler EJ. Molecular mechanisms of drug addiction. Neuropharmacology 2004;47(Suppl 1):24–32.
[5] Shahed JS, Crabbé EC. Gene expression induced by drugs of abuse. Curr Opin Pharmacol 2005;5:26–33.
[6] Guo AY, Webb BT, Miles MF, Zimmerman MP, Kendler KS, et al. ERGR: An ethanol-related gene resource. Nucleic Acids Res 2009;37:D840–845.
[7] Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, et al. SpeedSeq: ultra-fast platform for exploring deep-sequencing data. Nucleic Acids Res 2014;42:W187–191.
[8] Nestler EJ. Is there a common molecular pathway for addiction.pdf. Nat Neurosci 2010;13:946–74.

Acknowledgements

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.04.027.

References

[1] Pol Bodetto S, Carouge D, Fonteneau M, Dietrich JB, Zwiller J, et al. Cocaine represses protein phosphatase-1-beta through DNA methylation and Methyl-CpG Binding Protein-2 recruitment in adult rat brain. Neuropharmacology 2013;73:31–40.
[2] Chao J, Nestler EJ. Molecular neurobiology of drug addiction. Annu Rev Med 2004;55:113–32.
[3] Dymarska SE, Malecka RC, Nestler EJ. Neural mechanisms of addiction: the role of reward-related learning and memory. Annu Rev Neurosci 2006;29:565–98.
[4] Nestler EJ. Molecular mechanisms of drug addiction. Neuropharmacology 2004;47(Suppl 1):24–32.
[5] Shahed JS, Crabbé EC. Gene expression induced by drugs of abuse. Curr Opin Pharmacol 2005;5:26–33.
[6] Guo AY, Webb BT, Miles MF, Zimmerman MP, Kendler KS, et al. ERGR: An ethanol-related gene resource. Nucleic Acids Res 2009;37:D840–845.
[7] Chiang C, Layer RM, Faust GG, Lindberg MR, Rose DB, et al. SpeedSeq: ultra-fast platform for exploring deep-sequencing data. Nucleic Acids Res 2014;42:W187–191.
[8] Nestler EJ. Is there a common molecular pathway for addiction.pdf. Nat Neurosci 2010;13:946–74.

Acknowledgements

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.04.027.

References
[46] De Ruyck K, Nackaerts K, Beels I, Werbrouck J, De Volder A, et al. Genetic variation in three candidate genes and nicotine dependence, withdrawal and smoking cessation in hospitalized patients. Pharmacogenomics 2010;11:1053–63.

[47] Gao X, Wang Y, Lang M, Yuan L, Reece AS, et al. Contribution of Genetic Polymorphisms and Haplotypes in DRD2, BDNF, and Opioid Receptors to Heroin Dependence and Endophenotypes Among the Han Chinese. Hum Psychopharmacol 2017;21:404–12.

[48] Horowitz R, Kotler M, Shufman E, Aharoni S, Kremer I, et al. Confirmation of an excess of the high enzyme activity COMT val allele in heroin addicts in a family-based haplotype relative risk study. Am J Med Genet 2000;96:599–603.

[49] Lohoff FW, Weller AE, Bloch PJ, Nall AH, Ferraro TN, et al. Association between the catechol-O-methyltransferase Val158Met polymorphism and cocaine dependence. Neuropsychopharmacology 2008;33:3678–84.

[50] Malecka I, Jasiewicz A, Suchanecka A, Samochowiec J, Grzywacz A. Association and family studies of DRD2 gene polymorphisms in alcohol dependence syndrome. Postepy Hig Med Dosw (Online) 2014;68:1257–63.

[51] Zhou Z, Enoch MA, Goldman D. Gene expression in the addicted brain. Int Rev Neurobiol 2014;116:251–73.

[52] Nestler EJ. Is there a common molecular pathway for addiction?. Nat Neurosci 2005;8:1445–9.

[53] Ruffle JK. Molecular neurobiology of addiction: what’s all the (Δ)FosB about?. Am J Drug Alcohol Abuse 2014;40:428–37.