

Genetic signatures of heroin addiction

Shaw-Ji Chen, MD^{a,b}, Ding-Lieh Liao, MD^c, Tsu-Wang Shen, PhD^a, Hsin-Chou Yang, PhD^d, Kuang-Chi Chen, PhD^a, Chia-Hsiang Chen, MD, PhD^{e,f,*}

Abstract

Heroin addiction is a complex psychiatric disorder with a chronic course and a high relapse rate, which results from the interaction between genetic and environmental factors. Heroin addiction has a substantial heritability in its etiology; hence, identification of individuals with a high genetic propensity to heroin addiction may help prevent the occurrence and relapse of heroin addiction and its complications. The study aimed to identify a small set of genetic signatures that may reliably predict the individuals with a high genetic propensity to heroin addiction. We first measured the transcript level of 13 genes (RASA1, PRKCB, PDK1, JUN, CEBPG, CD74, CEBPB, AUTS2, ENO2, IMPDH2, HAT1, MBD1, and RGS3) in lymphoblastoid cell lines in a sample of 124 male heroin addicts and 124 male control subjects using real-time quantitative PCR. Seven genes (PRKCB, PDK1, JUN, CEBPG, CEBPB, ENO2, and HAT1) showed significant differential expression between the 2 groups. Further analysis using 3 statistical methods including logistic regression analysis, support vector machine learning analysis, and a computer software BIASLESS revealed that a set of 4 genes (JUN, CEBPB, PRKCB, ENO2, or CEBPG) could predict the diagnosis of heroin addiction with the accuracy rate around 85% in our dataset. Our findings support the idea that it is possible to identify genetic signatures of heroin addiction using a small set of expressed genes. However, the study can only be considered as a proof-of-concept study. As the establishment of lymphoblastoid cell line is a laborious and lengthy process, it would be more practical in clinical settings to identify genetic signatures for heroin addiction directly from peripheral blood cells in the future study.

Abbreviations: AUTS2 = autism susceptibility candidate 2, BIASLESS = Biomarkers Identification and Samples Subdivision, CD74 = CD74 molecule, CEBPB = CCAAT/enhancer binding protein beta, CEBPG = CCAAT/enhancer binding protein gamma, ENO2 = enolase 2 (gamma, neuronal), HAT1 = histone acetyltransferase 1, IMPDH2 = IMP (inosine 5'-monophosphate) dehydrogenase 2, JAK-Stat = Janus kinase-signal transducer, and activator of transcription, JUN = jun proto-oncogene, LCL = lymphoblastoid cell line, MBD1 = methyl-CpG binding domain protein 1, lncRNAs = long noncoding RNA, LR = logistic regression, PDK1 = pyruvate dehydrogenase kinase 1, PRKCB = protein kinase C, beta, RASA1 = RAS p21 protein activator (GTPase activating protein) 1, RGS3 = regulator of G-protein signaling 3, RT-qPCR = real-time quantitative PCR, SVM = support vector machine.

Keywords: biomarker, diagnosis, genetic signatures, heroin addiction

1. Introduction

Heroin was synthesized as a legal drug in 1895. It was found to be a highly addictive drug in the early twentieth century, and became

popularly abused after the mid-twentieth century.^[1,2] Heroin users develop tolerance to heroin quickly and suffer from severe withdrawal symptoms when they stop using heroin. Hence, heroin addicts have a very high relapse rate.^[3,4] Heroin addiction became a serious problem in southeastern and southwestern Asia recently.^[5] Like that in other countries, heroin users not only had a high recidivism rate but also had the highest mortality rate among the addictive substance users in Taiwan.^[6,7]

Heroin addiction is a complex disorder resulting from the interactions between genetic susceptibility and environmental factors.^[8] Previous studies have demonstrated that the estimated heritability of drug addiction ranges from 0.39 to 0.72 in different drugs.^[9,10] In a twin resemblance study for illicit psychoactive substance use, heavy use, abuse, and dependence in a US population, the heritability of illicit psychoactive substance use ranges from 60% to 80%. These findings suggest that genetic factor plays an important role in the pathogenesis of heroin addiction.

Many genes associated with heroin addiction have been reported using different approaches, supporting the high genetic heterogeneity of heroin addiction.^[2,9] However, there is still a gap of applying the genetic knowledge of heroin addiction to help manage the patients with heroin addiction in clinical settings. The study aimed to find a small set of genes that can reliably predict the individuals with a high propensity to heroin addiction. The tests of these genes shall be conveniently implemented in the clinical laboratory, and hopefully, the assay of these genes can be

Editor: Michele Fornaro.

S-JC and D-LL contributed equally to this work.

This work was supported by the intramural grant from the National Health Research Institutes, Taiwan.

All authors critically reviewed content and approved final version for publication.

The authors have no conflicts of interest to disclose.

^aInstitute of Medical Sciences, Tzu Chi University, Hualien, ^bDepartment of Psychiatry, Mackay Memorial Hospital, Taitung Branch, ^cDepartment of Health Executive Yuan, Bali Psychiatric Center, ^dInstitute of Statistical Science, Academia Sinica, Taipei, ^eDepartment of Psychiatry, Chang Gung Memorial Hospital at Linkou, ^fDepartment and Graduate Institute of Biomedical Sciences, Chang Gung University, Taoyuan, Taiwan.

* Correspondence: Chia-Hsiang Chen, Department of Psychiatry, Chang Gung Memorial Hospital at Linkou, No. 5 Fusing Street, Kueishan, Taoyuan, 333 Taiwan (e-mail: cchen3801@gmail.com)

Copyright © 2016 the Author(s). Published by Wolters Kluwer Health, Inc. All rights reserved.

This is an open access article distributed under the Creative Commons Attribution License 4.0 (CCBY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Medicine (2016) 95:31(e4473)

Received: 27 January 2016 / Received in final form: 7 July 2016 / Accepted: 9 July 2016

<http://dx.doi.org/10.1097/MD.0000000000004473>

used in combination with other social, familial programs to help prevent the occurrence and the relapse of heroin addiction in clinical settings.

2. Materials and methods

2.1. Subjects

Male adult patients (≥ 20 years old) met the diagnostic criteria of heroin use disorder according to the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5) was recruited into this study. The diagnosis was made based on medical records and interview by senior psychiatrists with consensus. Patients comorbid with other major psychiatric diagnoses of DSM-5 such as neurodevelopmental disorders, schizophrenia spectrum and other psychotic disorders, bipolar related disorders, depressive disorders, and neurocognitive disorders were excluded. Male adult subjects (≥ 20 years old) who received regular medical checkups at a local medical center were recruited into this study as controls. A senior psychiatrist evaluated their mental statuses and histories of mental illness. Those who had a history or currently had a diagnosis of major psychiatric disorders according to DSM-5 such as substance-related and addictive disorders, neurodevelopmental disorders, schizophrenia spectrum and other psychotic disorders, bipolar-related disorders, depressive disorders, and neurocognitive disorders were excluded. A total of 8 mL venous blood was collected from each subject for the establishment of lymphoblastoid cell lines (LCLs). Total RNA was extracted from the LCL for the measurement of cDNA of selected genes. The study protocol was approved by the Ethical Committee of Bali Psychiatric Center (approval number: IRB970609-03), and written informed consent was obtained after full explanation of the protocol.

2.2. LCL, total RNA, and cDNA preparation

LCL from each subject was established by transforming the lymphocytes with Epstein–Barr virus following the procedures described in our previous paper.^[11] After the establishment of LCL, total RNA from the LCL was extracted using TRIZOL Total RNA Isolation Reagent according to manufacturer's instruction (Invitrogen Life Technologies, Carlsbad, CA). The cDNA was prepared using Superscript II RNase H⁻ Reverse Transcriptase following the instructions provided by the

manufacturer (Invitrogen Life Technologies). The details were also described in our previous report.^[12]

2.3. Measurement of genetic transcript level in LCL

A total of 13 genes (RASA1, PRKCB, PDK1, JUN, CEBPG, CD74, CEBPB, AUTS2, ENO2, IMPDH2, HAT1, MBD1, and RGS3) were measured in this study. The genetic transcript level was measured using real-time quantitative PCR (RT-qPCR) that was performed by SYBR Green method and implemented in StepOnePlus Real-Time PCR System according to the manufacturer's protocol (Applied Biosystems, Forster City, CA). The detailed procedures were described in our previous report.^[12] The relative standard curve method was used for the measurement of the gene transcript level according to the procedure described in User Bulletin #2 ABI PRISM 7700 sequence detection system (Applied Biosystems). In this method, serial dilutions of known amount of RNA from a reference sample (pooled from 40 LCLs of male controls) were used to generate the external standard curve. For each unknown sample, the relative amount was calculated using linear regression analysis from their respective standard curves. The transcript level of measured gene in each subject was normalized by his/her 18S rRNA level. The reference 18S rRNA level was measured using predeveloped TaqMan assay reagents 18S rRNA MGB according to the manufacturer's protocol (Applied Biosystems). All experiments were performed in duplicate. The sequences of all the primer sets and optimal annealing temperature used for RT-qPCR are listed in Table 1.

2.4. Statistical analysis

To cross-validate the findings of genetic signatures for heroin addiction, we used 3 statistical methods to search for the genes that can reliably predict the diagnosis of heroin addiction in this dataset, including multiple logistic regression (LR) analysis, support vector machine (SVM) method, and the online computer program BIASLESS (Biomarkers Identification and Samples Subdivision). LR analysis was implemented using the Statistical Package for the Social Science V18.0 (SPSS Inc., Chicago, IL). The SVM model analysis was implemented using the LIBSVM package by Chang and Lin.^[13] The BIASLESS is a software developed for integrative analysis of single nucleotide polymorphisms (SNP) and gene expression, which is an efficient tool for

Table 1

Sequences of primer sets sequences for the real-time quantitative PCR experiments.

| Gene symbol | Gene ID | F-primer (5'–3') | R-primer (5'–3') | Ta, °C |
|-------------|---------|-----------------------------|------------------------|--------|
| JUN | 3725 | CGAAAAAGGAAGCTGGAGAG | TGTTAAGCTGTGCCACCTG | 60 |
| CEBPB | 1051 | TGCAGAAGAAGGTGGAGCAG | TAGCAGTGGCCGGAGGAGG | 63 |
| CEBPG | 1054 | GGGCTAGAGGAGCAGGTACA | ACTCTCCCTTGCCAAACACAG | 60 |
| PRKCB | 5579 | TACTGAGCCAGGAGGAAGG | TGGAGACAGTGTGGTCGTC | 60 |
| ENO2 | 2026 | AGCTGAGGGATGGAGACAAA | GCTCCACCACAGAGAGACCT | 60 |
| HAT1 | 8520 | TCTAAAGTTGATGAGAACCTTGACTGT | TTGTCTAATTTTGCCCTCAACA | 60 |
| PDK1 | 5163 | TATGGAACACCATGCCAACA | CACCTCCTCGGTCACTCATC | 60 |
| RASA1 | 5921 | ATGACACAGTGGATGGCAAG | AGCTCGTTTTTCGCTTTCAA | 60 |
| CD74 | 972 | AAATGAGCAGGCACTCCTTG | AGATCCTGCTTGGTCACACG | 60 |
| AUTS2 | 26053 | GATTTGCCATGACCAGCTTT | TGAAAGGACAAGCCATTGGT | 60 |
| IMPDH2 | 3615 | GAAGACTTGGTGGTAGCCCC | TGGAGGCTAGTGGGTAGTCC | 60 |
| MBD1 | 4152 | GGTGTGTGAGAACTGTGGA | ACATTCTGTTCGCCGTTG | 60 |
| RGS3 | 5998 | CCAGCACCCCAAGAAAGAG | TGAAGATCCCCAGCTGTTC | 60 |

Ta=optimal annealing temperature.

selecting key SNP and gene expression markers and then building models for sample subdivision.^[14]

3. Results

3.1. Subjects

All the subjects were Han Chinese from Taiwan. A total of 124 male adults (37.5 ± 9.4 years) fulfilling the diagnostic criteria for heroin use disorders as defined by the DSM-5 were recruited into this study. Also, a total of 124 male adults (42.9 ± 14.6 years) without history or current diagnosis of substance use disorders and other major psychiatric diagnoses were recruited as the control groups. Their LCLs were established, cDNAs were prepared for the genetic assessment.

3.2. Genetic assessment

A total of 13 genetic transcript levels were measured for each subject and normalized by the total amount of their 18S rRNA. The mean transcript levels of each of the 13 genes in the patient and control groups are listed in Table 2. Among these 13 genes, 7 genes had a differential expression that reached statistical significance between the 2 groups, including JUN, CEBP, and CEBPG genes that were upregulated and ENO2, PRKCB, HAT1, and PDK1 genes that downregulated in heroin addicts compared with the control group. The mean transcript levels of RASA1, CD47, and IMPDH2 were nominally downregulated in heroin addicts compared with control subjects; however, they did not reach statistical significance after correction for multiple tests. No differences of in the mean transcript levels of AUTS2, MBD1, and RGS3 genes were observed between the 2 groups.

3.3. LR analysis

In LR analysis, subjects with and without the diagnosis of heroin addiction were regressed with a linear combination of the gene expression levels. The accuracy rates of different models containing 13, 7, 5, and 4 genes are listed in Table 3. The odds ratios of the 7 genes that showed significant differential expressions between the 2 groups are listed in Table 4. In this analysis, we found that the 4-gene model that contained JUN, CEBPB, ENO2, and PRKCB genes had an accuracy rate of 85.5% in the prediction of heroin addiction in this dataset.

3.4. SVM analysis

In SVM analysis, we conducted the kernel function as radial basis function as it is the most accepted in a similar analysis. We used leave-one-out cross-validation on the patient gene expression dataset, and all subjects were built-in for both training and testing using the same formula. Individuals in the patient and control groups were randomly partitioned into 10 subsets for a cross-validation. Nine subsets formed a training set and the remaining 1 subset function a testing set. The training set was used to classify individuals in a testing dataset and calculated a testing accuracy. The accuracy rates of different gene combinations are summarized in Table 5. We found that the combination of 4 genes (JUN, ENO2, CEBPG, and PRKCB) had the highest accuracy rate around 84%.

3.5. BIASLESS analysis

In the analysis using BIALESS software, the data were run with the default parameter values. For each gene, gene expression level was normalized by subtracting the sample mean and dividing by the sample standard deviation of all 248 individuals (from 6 batches). Individuals in the patient and control groups were randomly partitioned into 10 subsets for a cross-validation. Nine subsets formed a training set, and the remaining 1 subset was a testing set. A flexible discrimination analysis was applied to the training set first. The genes with the highest increment of training accuracy were included into the classification model sequentially until the training accuracy reached 1.0 or its increment was less than a threshold of 0.001. Next, the classification model in the training set was used to classify individuals in a testing dataset and calculated a testing accuracy. The previous procedures were repeated until each of the 10 subsets had been analyzed as a testing dataset. Finally, among the 10 candidate classification models, the 1 with the highest cross-validation consistency was selected as the best classification model in this study. A leave-one-out testing accuracy of the best model was calculated.

Our classification analysis identified JUN, CEBPG, ENO2, and PRKCB as the key gene expression signatures for a subdivision of heroin dependence and controls. The suggested classification model with 4 gene signatures had a cross-validation consistency of 4/10 in our 10-fold cross-validation analysis. The average training accuracy rate was 86.7%, and the leave-one-out testing accuracy was 85.5%. Distributions of gene expression levels in

Table 2
Summary of expression levels of 13 genes in this study.

| Gene symbol | Gene ID | Heroin addicts (n=124) | Controls (n=124) | P |
|-------------|---------|------------------------|------------------|--------|
| JUN | 3725 | 6.11 ± 4.64 | 3.26 ± 3.00 | <0.001 |
| CEBPB | 1051 | 2.38 ± 3.19 | 0.69 ± 0.57 | <0.001 |
| CEBPG | 1054 | 6.72 ± 4.76 | 4.24 ± 2.72 | <0.001 |
| PRKCB | 5579 | 3.45 ± 3.48 | 5.85 ± 4.18 | <0.001 |
| ENO2 | 2026 | 6.26 ± 3.81 | 11.26 ± 9.06 | <0.001 |
| HAT1 | 8520 | 5.26 ± 4.75 | 8.38 ± 10.56 | 0.003 |
| PDK1 | 5163 | 8.67 ± 6.57 | 12.00 ± 10.93 | 0.004 |
| RASA1 | 5921 | 7.02 ± 8.55 | 9.24 ± 5.70 | 0.02 |
| CD47 | 972 | 3.96 ± 4.60 | 5.88 ± 9.61 | 0.05 |
| AUTS2 | 26053 | 4.04 ± 4.91 | 5.2 ± 7.25 | 0.13 |
| IMPDH2 | 3615 | 5.20 ± 3.48 | 6.74 ± 6.07 | 0.015 |
| MBD1 | 4152 | 15.48 ± 12.00 | 19.03 ± 24.00 | 0.14 |
| RGS3 | 5998 | 1.20 ± 1.21 | 2.49 ± 2.98 | 1.13 |

Table 3**Logistic regression of genetic signature of heroin dependence.**

| Gene number | Gene list | Accuracy, % |
|-------------|--|-------------|
| 13 genes | RASA1, PRKCB, PDK1, JUN, CEBPG, CD74, CEBPB, AUTS2, ENO2, IMPDH2, HAT1, MBD1, RGS3 | 87.9 |
| 7 genes | PRKCB, JUN, CEBPG, CEBPB, ENO2, HAT1, PDK1 | 88.3 |
| 5 genes | PRKCB, JUN, CEBPG, CEBPB, ENO2 | 85.9 |
| 4 genes | JUN, CEBPB, ENO2, PRKCB | 85.5 |

the heroin dependence and control groups are displayed in box-whisker plots (Fig. 1).

4. Discussion

Heroin addiction, like the other substance use disorders, is a complex disorder resulting from the interplays between environment and genetic predisposition.^[8] Hence, identification of the predisposing genes of heroin addiction is crucial to understand the neurobiology and the pathogenesis of heroin addiction. Microarray-based gene expression profiling analysis allows simultaneous measurement of multiple genes expression in cells or tissues,^[15] which had been used in discovering genetic pathways related to drug addiction. For example, a set of differentially expressed genes involved in the regulation of transcription, chromatin, and dopamine cell phenotype were detected in the postmortem human brains in chronic cocaine abusers compared with control subjects by microarray and quantitative PCR.^[16] Also, altered gene expression in pathways related to T-cell receptor, Janus kinase-signal transducer, and activator of transcription (JAK-Stat) signaling were found among in these 3 groups of alcohol drinkers (alcohol dependence, heavy drinkers, and moderate drinkers) using expression array technology.^[17] Our group also conducted a comparative gene expression profiling analysis of LCLs in a small sample of male heroin addicts and male control subjects previously. We detected 924 differentially expressed gene transcripts between these 2 groups, including 279 upregulated and 645 downregulated gene transcripts in individuals with heroin dependence compared with control subjects.^[11] These findings support the complex polygenic nature of heroin addiction.

Results from microarray-based gene expression studies usually contain dozens of genes, which are difficult to be applied in clinical settings. Hence, several groups have attempted to use a small set of genes as signatures to represent the findings from microarray-based studies. For example, a 21-gene assay was found helpful to assist in adjuvant treatment decision in hormone receptor-positive breast cancer patients.^[18] An 8-gene expression signature could help to predict the survival and time to the treatment of chronic lymphocytic leukemia,^[19] and a 6-gene

signature was found to be able to distinguish 4 subgroups of neuroblastoma.^[20] In drug abuse, no such study was reported yet in the literature to our knowledge. Hence, the study aimed to test the hypothesis whether a small set of signature genes can reliably differentiate heroin addicts from control subjects.

The study is an extension of our previous study of comparative gene expression profiling analysis of LCLs in a small sample of 20 male heroin addicts and 20 male control subjects.^[11] We originally intended to validate the differentially expressed genes found in expression microarray using RT-qPCR. Then, we came up with the idea that if we could find a small set of genes that could reliably predict the individuals with a high genetic propensity to heroin addiction, which would be useful in preventing the occurrence and relapse of heroin addiction. To test this hypothesis, we randomly selected 13 genes based on our preliminary study and measured the transcript levels of 13 genes in 124 male heroin addicts and 124 male control subjects using RT-qPCR. Among these 13 genes, we found 7 genes were differentially expressed that reached statistical significance after correcting for multiple testing. These 7 genes were further analyzed for building genetic signatures using 3 different statistical methods. We finally obtained 4-gene signatures that can predict the diagnosis of heroin addiction with the accuracy rate of around 85% in our dataset.

These 4 genes include JUN, CEBPB, ENO2, and PRKCB in LR modeling and SVM analysis while in the analysis using BIALESS computer program, the CEBPB gene was replaced by the CEBPG. The discrepancy between these methods can be attributed to the different algorithms of these methods. Nevertheless, the identification of these genes supports the idea that a small set of gene transcripts can function as genetic signatures of heroin addiction.

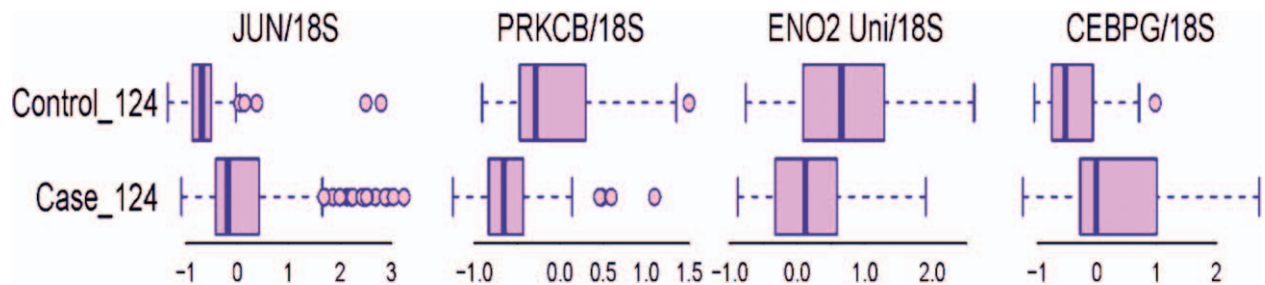
The interpretation of the biological significance of these signature genes of heroin addiction can be intriguing. JUN is proto-oncogene and a transcription factor that regulates the expression of its target genes. A recent paper reported that increased c-Jun expression in the cerebellum of rats under long-term heroin administration, which was attributed to be one of the mechanisms underlying the heroin-induced cerebellum neuronal damage.^[21] However, no other studies reported elevated intrinsic JUN in heroin addicts to our knowledge. Both CEBPB and CEBPG are members of CEPB family that contain a leucine zipper domain to form homodimer and heterodimer with each other. These 2 transcription factors are involved in many physiological and pathological conditions in humans.^[22–25] PRKCB is a member of the protein kinase C (PRKC) family that are threonine-specific protein kinases that play an important role in signal transduction, regulation of gene expression, and control of cell division and differentiation. Studies reported that PRKCB played an essential role in dendritic cell differentiation^[26] and was involved in the pathogenesis of autism.^[27,28] However, the relation between PRKCB and substance use disorders remains to

Table 4**Odds ratio of 7 genes in logistic regression analysis.**

| | Crude odds ratio | Adjusted odds ratio |
|-------|------------------|---------------------|
| PRKCB | 0.83 (0.76–0.90) | 0.82 (0.69–0.98) |
| JUN | 1.28 (1.16–1.41) | 1.65 (1.33–2.06) |
| CEBPB | 2.68 (1.79–4.01) | 2.08 (1.11–3.92) |
| CEBPG | 1.21 (1.11–1.31) | 1.26 (1.04–1.52) |
| ENO2 | 0.86 (0.81–0.91) | 0.84 (0.76–0.94) |
| PDK1 | 0.96 (0.93–0.99) | 0.95 (0.87–1.04) |
| HAT1 | 0.93 (0.88–0.98) | 0.89 (0.78–1.01) |

Table 5**Support vector machines' analysis of genetic signature of heroin addiction.**

| Gene number | Gene list | Accuracy, % |
|-------------|--|-------------|
| 13 genes | RASA1, PRKCB, PDK1, JUN, CEBPG, CD74, CEBPB, APTS2, ENO2, IMPDH2, HAT1, MBD1, RGS3 | 73.4 |
| 7 genes | PRKCB, JUN, CEBPG, CEBPB, ENO2, HAT1, PDK1 | 74.2 |
| 5 genes | PRKCB, JUN, CEBPG, CEBPB, ENO2 | 80.7 |
| 4 genes | JUN, CEBPB, ENO2, PRKCB | 83.9 |

**Figure 1.** Box-whisker plots of transcript levels of the 4 selected gene expression signatures in the case and control groups.

be studied. Enolase is involved in the glycolysis and gluconeogenesis pathways. Human enolase is a homodimeric protein complex consisted of 2 subunits from 3 isoforms of enolase that are encoded by ENO1, ENO3, and ENO2, respectively. Tannu et al reported increased gamma enolase that was encoded by the ENO2 in the nucleus accumbens of cocaine overdose victims compared with control subjects,^[29] suggesting the involvement of the ENO2 in the pathophysiology of cocaine dependence. Our previous study also indicated that ENO2 was associated with heroin dependence.^[11] These signature genes may not directly link to heroin addiction at first look; they may represent the complex genetic pathways underlying the genetic disposition of heroin addiction. Hence, a combination of the expression of these signature genes can reliably distinct heroin addiction from control subjects with high accuracy rate as shown in this study.

LR is a powerful modeling tool for predicting the outcome; it assumes a linear relationship between predicting variables and outcome.^[30] SVM is a supervised machine learning algorithm that is used for both linear and nonlinear classification.^[31] Interestingly, both LR and SVM analyses obtain the same 4 genes that had the accuracy rate of approximately 85.5% and 83.9%, respectively, supporting the robustness of the 4-gene signatures of heroin addiction. In the analysis using BIASLESS, the CEBPB was replaced by the CEBPG. Still, the 4-gene model had the accuracy rate of 85.5% in predicting the diagnosis of heroin addiction.

The study has several limitations. Due to limited resources, we were not able to measure more gene transcripts to find better genetic signatures with higher accuracy rate than the current findings in the prediction of heroin addiction. Also, the relatively small sample size can also affect the predicting power of our model in this study. Furthermore, the gene signatures found in this study may not be specific for heroin addiction, as substance use disorders share some common genetic susceptibility.^[31] Hence, the specificity of our findings in this study should be tested in another kind of substance use disorders. Also, the present study only measured the gene expression at mRNA level, which may not represent the gene

expression at the protein level. It would be interesting to examine the changes of these signature genes at protein level in the postmortem brains of heroin addicts or the brains of rodents under long-term heroin administration in the future study.

Heroin addiction is a complicated result from interactions between genetic susceptibility and environmental factors. To reduce the confounding environment factors, we used LCLs as our experimental materials instead of peripheral blood cells, because the LCL was cultured in the medium for a long period, the influence of environmental factors was controlled to the minimum. Hence, the genes identified in this study can be considered as trait markers than as state markers. Also, the establishment of LCLs is a laborious and lengthy process, which also limits its clinical use. Hence, this study can only be considered as a proof-of-concept study. We suggest that it would be more useful in clinical settings if the future study can identify a small set of signature genes of heroin addiction from peripheral blood cells rather than from LCL.

In conclusion, the present study suggests the feasibility of detecting a small set of genetic signatures from peripheral tissues that can classify the diagnosis of heroin addiction with reasonable accuracy rate. The approach in this study shall facilitate the search for trait genetic markers and biomarkers of heroin addiction, and promote the translational research of heroin addiction in clinical settings.

References

- [1] Tarabar AF, Nelson LS. The resurgence and abuse of heroin by children in the United States. *Curr Opin Pediatr* 2003;15:210–5.
- [2] Bailey A, Yufarov V, Bendor J, et al. Immediate withdrawal from chronic “binge” cocaine administration increases mu-opioid receptor mRNA levels in rat frontal cortex. *Mol Brain Res* 2005;137:258–62.
- [3] Bell J, Zador D. A risk-benefit analysis of methadone maintenance treatment. *Drug Safety* 2000;22:179–90.
- [4] Trujols J, Guardia J, Pero M, et al. Multi-episode survival analysis: an application modelling readmission rates of heroin dependents at an inpatient detoxification unit. *Addict Behav* 2007;32:2391–7.

- [5] Kulsudjarit K. Drug problem in southeast and southwest Asia. *Ann N Y Acad Sci* 2004;1025:446–57.
- [6] Chiang SC, Chan HY, Chen CH, et al. Recidivism among male subjects incarcerated for illicit drug use in Taiwan. *Psychiatry Clin Neurosci* 2006;60:444–51.
- [7] Chen CY, Wu PN, Su LW, et al. Three-year mortality and predictors after release: a longitudinal study of the first-time drug offenders in Taiwan. *Addiction* 2010;105:920–7.
- [8] Reder LM, Weber K, Shang J, et al. The adaptive character of the attentional system: statistical sensitivity in a target localization task. *J Exp Psychol Hum Percept Perform* 2003;29:631–49.
- [9] Goldman D, Oroszi G, Ducci F. The genetics of addictions: uncovering the genes. *Nat Rev Genet* 2005;6:521–32.
- [10] Uhl GR, Drgon T, Johnson C, et al. “Higher order” addiction molecular genetics: convergent data from genome-wide association in humans and mice. *Biochem Pharmacol* 2008;75:98–111.
- [11] Liao DL, Cheng MC, Lai CH, et al. Comparative gene expression profiling analysis of lymphoblastoid cells reveals neuron-specific enolase gene (ENO2) as a susceptibility gene of heroin dependence. *Addict Biol* 2014;19:102–10.
- [12] Huang CH, Chen ML, Tsai YL, et al. Elevated adrenomedullin mRNA in lymphoblastoid cells from schizophrenic patients. *Neuroreport* 2004;15:1443–6.
- [13] Chang CC, Lin CJ. LIBSVM: a library for support vector machines. *ACM T Intel Syst Tec* 2011;2: Article 27.
- [14] Yang HC, Wang PL, Lin CW, et al. Integrative analysis of single nucleotide polymorphisms and gene expression efficiently distinguishes samples from closely related ethnic populations. *BMC Genomics* 2012;13:346.
- [15] Schena M, Shalon D, Heller R, et al. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA* 1996;93:10614–9.
- [16] Bannon MJ, Johnson MM, Michelhaugh SK, et al. A molecular profile of cocaine abuse includes the differential expression of genes that regulate transcription, chromatin, and dopamine cell phenotype. *Neuropsychopharmacology* 2014;39:2191–9.
- [17] Beech RD, Qu J, Leffert JJ, et al. Altered expression of cytokine signaling pathway genes in peripheral blood cells of alcohol dependent subjects: preliminary findings. *Alcohol Clin Exp Res* 2012;36:1487–96.
- [18] Kelly CM, Warner E, Tsoi DT, et al. Review of the clinical studies using the 21-gene assay. *Oncologist* 2010;15:447–56.
- [19] Herold T, Jurinovic V, Metzeler KH, et al. An eight-gene expression signature for the prediction of survival and time to treatment in chronic lymphocytic leukemia. *Leukemia* 2011;25:1639–45.
- [20] Abel F, Dalevi D, Nethander M, et al. A 6-gene signature identifies four molecular subgroups of neuroblastoma. *Cancer Cell Int* 2011;11:9.
- [21] Pu H, Wang X, Zhang J, et al. Cerebellar neuronal apoptosis in heroin-addicted rats and its molecular mechanism. *Int J Clin Exp Pathol* 2015;8:8260–7.
- [22] van der Krieken SE, Popeijus HE, Mensink RP, et al. CCAAT/enhancer binding protein beta in relation to ER stress, inflammation, and metabolic disturbances. *BioMed Res Int* 2015;2015:324815.
- [23] Guo L, Li X, Tang QQ. Transcriptional regulation of adipocyte differentiation: a central role for CCAAT/enhancer-binding protein (C/EBP) beta. *J Biol Chem* 2015;290:755–61.
- [24] Huggins CJ, Malik R, Lee S, et al. C/EBPgamma suppresses senescence and inflammatory gene expression by heterodimerizing with C/EBPbeta. *Mol Cell Biol* 2013;33:3242–58.
- [25] Melchionna R, Bellavia G, Romani M, et al. C/EBPgamma regulates wound repair and EGF receptor signaling. *J Invest Dermatol* 2012;132:1908–17.
- [26] Cejas PJ, Carlson LM, Zhang J, et al. Protein kinase C betaII plays an essential role in dendritic cell differentiation and autoregulates its own expression. *J Biol Chem* 2005;280:28412–23.
- [27] Lintas C, Sacco R, Garbett K, et al. Involvement of the PRKCB1 gene in autistic disorder: significant genetic association and reduced neocortical gene expression. *Mol Psychiatry* 2009;14:705–18.
- [28] Philippi A, Roschmann E, Tores F, et al. Haplotypes in the gene encoding protein kinase C-beta (PRKCB1) on chromosome 16 are associated with autism. *Mol Psychiatry* 2005;10:950–60.
- [29] Tannu N, Mash DC, Hemby SE. Cytosolic proteomic alterations in the nucleus accumbens of cocaine overdose victims. *Mol Psychiatry* 2007;12:55–73.
- [30] Delen D, Walker G, Kadam A. Predicting breast cancer survivability: a comparison of three data mining methods. *Artif Intell Med* 2005; 34:113–27.
- [31] Agrawal A, Verweij KJ, Gillespie NA, et al. The genetics of addiction—a translational perspective. *Transl Psychiatry* 2012;2:e140.