
DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

A Genome-wide Association Study of Cocaine Use Disorder Accounting for Phenotypic Heterogeneity and Gene-Environment Interaction

Jiangwen Sun,¹ Henry R. Kranzler², Joel Gelernter³, Jinbo Bi,¹

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

**Figure S1.** Dot plot of the top three principal components in GWAS sample

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

**Figure S2.** Dot plot of the top three principal components in ES sample

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

---

**Figure S3.** Manhattan plot showing results from GWAS that tests the genetic effect on DSM5 diagnostic criteria count for cocaine use disorder in AA sample, modulated by the childhood environmental factor, change in residence.

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

---

**Figure S4.** Quantile-quantile plot showing the observed distribution of P-values comparing to the expected distribution for GWAS that tests the genetic effect on DSM5 diagnostic criteria count for cocaine use disorder in AA sample, modulated by the childhood environmental factor, change in residence. The genomic inflation factor (\( \lambda \)) is the ratio of the median of the observed P-value to its expected value.

\[ \lambda = 1.015 \]

DOI: 10.1503/jpn.180098

© Joule Inc.

Online appendices are unedited and posted as supplied by the authors.

Figure S5. Manhattan plot showing results from GWAS that tests the genetic effect on DSM5 diagnostic criteria count for cocaine use disorder in AA sample, modulated by the childhood environmental factor, household drinking and illicit drug use.

DOI: 10.1503/jpn.180098

© Joule Inc.

Online appendices are unedited and posted as supplied by the authors.

Figure S6. Quantile-quantile plot showing the observed distribution of P-values comparing to the expected distribution for GWAS that tests the genetic effect on DSM5 diagnostic criteria count for cocaine use disorder in AA sample, modulated by the childhood environmental factor, “household drinking and illicit drug use. The genomic inflation factor ($\lambda$) is the ratio of the median of the observed P-value to its expected value.

DOI: 10.1503/jpn.180098

© Joule Inc.

Online appendices are unedited and posted as supplied by the authors.

**Figure S7.** Manhattan plot showing results from GWAS that tests the genetic effect on the membership score of Subtype 5 in AA sample, modulated by the childhood environmental factor, non-traditional parental care.

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

**Figure S8.** Quantile-quantile plot showing the observed distribution of P-values comparing to the expected distribution for GWAS that tests the genetic effect on the membership score of Subtype 5 in AA sample, modulated by the childhood environmental factor, non-traditional parental care. The genomic inflation factor ($\lambda$) is the ratio of the median of the observed P-value to its expected value.

$\lambda = 1.048$

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

**Figure S9.** Manhattan plot showing results from GWAS that tests the genetic effect on the membership score of Subtype 5 in AA sample, modulated by the childhood environmental factor, change in residence.

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

**Figure S10.** Quantile-quantile plot showing the observed distribution of P-values comparing to the expected distribution for GWAS that tests the genetic effect on the membership score of Subtype 5 in AA sample, modulated by the childhood environmental factor, change in residence. The genomic inflation factor ($\lambda$) is the ratio of the median of the observed P-value to its expected value.

\[ \lambda = 1.076 \]

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

**Figure S11.** Manhattan plot showing results from GWAS that tests the genetic effect on the membership score of Subtype 5 in AA sample, modulated by the childhood environmental factor, household drinking and illicit drug use.

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

**Figure S12.** Quantile-quantile plot showing the observed distribution of P-values comparing to the expected distribution for GWAS that tests the genetic effect on the membership score of Subtype 5 in AA sample, modulated by the childhood environmental factor, household drinking and illicit drug use. The genomic inflation factor ($\lambda$) is the ratio of the median of the observed P-value to its expected value.

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

**Figure S13.** Manhattan plot showing results from GWAS that tests the genetic effect on the membership score of Subtype 4 in EA sample, modulated by the childhood environmental factor, experience of violence and/or abuse.

**Figure S14.** Quantile-quantile plot showing the observed distribution of P-values comparing to the expected distribution for GWAS that tests the genetic effect on the membership score of Subtype 5 in EA sample, modulated by the childhood environmental factor, experience of violence and/or abuse. The genomic inflation factor ($\lambda$) is the ratio of the median of the observed P-value to its expected value.

$$\lambda = 1.009$$

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

---

**Figure S15.** Manhattan plot showing results from GWAS that tests the genetic effect on the membership score of Subtype 4 in EA sample, modulated by the childhood environmental factor, household drinking and illicit drug use.

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

**Figure S16.** Quantile-quantile plot showing the observed distribution of P-values comparing to the expected distribution for GWAS that tests the genetic effect on the membership score of Subtype 4 in EA sample, modulated by the childhood environmental factor, household drinking and illicit drug use. The genomic inflation factor ($\lambda$) is the ratio of the median of the observed P-value to its expected value.

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

![Gene expression distribution (RPKM, Reads per Kilobase Million) of LINC01411 across human tissues.](image)

**Figure S17.** Gene expression distribution (RPKM, Reads per Kilobase Million) of *LINC01411* across human tissues.

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

**Figure S18.** Gene expression distribution (RPKM, Reads per Kilobase Million) of *TEME51* across human tissues.

DOI: 10.1503/jpn.180098

© Joule Inc.

Online appendices are unedited and posted as supplied by the authors.

**Figure S19.** Gene expression distribution (RPKM, Reads per Kilobase Million) of SYN1 across human tissues.

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

---

**Figure S20.** Gene expression distribution (RPKM, Reads per Kilobase Million) of *TENM3* across human tissues.
Figure S21. Correlations among the three CUD related traits. DCC: DSM-5 diagnostic criterion count, S4ML: subtype 4 membership likelihood, S5ML: subtype 5 membership likelihood. Correlations were calculated using the combined sample set (i.e., discover samples plus replication samples.)

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

Figure S22. Correlation among the four childhood environmental factors. NTPC: Non-traditional parent care, CIR: Change in residence, TE: Traumatic experience, HDDU: Household drinking and illicit drug use, HTU: Household tobacco use. Correlations were calculated using the combined sample set (i.e., discover samples plus replication samples.)

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

Table S1: Self-reported and predicted race (by running k-means clustering with the first PC)

<table>
<thead>
<tr>
<th>Self-reported race</th>
<th>GWAS data</th>
<th>EMA data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Predicted AA</td>
<td>Predicted EA</td>
</tr>
<tr>
<td>AA Hispanic Origin</td>
<td>104</td>
<td>64</td>
</tr>
<tr>
<td>AA None Hispanic Origin</td>
<td>3,195</td>
<td>14</td>
</tr>
<tr>
<td>EA Hispanic Origin</td>
<td>4</td>
<td>306</td>
</tr>
<tr>
<td>EA None Hispanic Origin</td>
<td>16</td>
<td>1,832</td>
</tr>
<tr>
<td>Other</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>3,324</td>
<td>2,216</td>
</tr>
</tbody>
</table>

Note that correction for the self-reported race was done on the full GWAS (total 5,540 subjects) and EMA data (total 3,675 subjects)

DOI: 10.1503/jpn.180098

© Joule Inc.

Online appendices are unedited and posted as supplied by the authors.

Table S2: Lifetime prevalence of substance use and psychiatric disorders by subtype

<table>
<thead>
<tr>
<th>Disorders</th>
<th>Subtype 1 (n, %)</th>
<th>Subtype 2 (n, %)</th>
<th>Subtype 3 (n, %)</th>
<th>Subtype 4 (n, %)</th>
<th>Subtype 5 (n, %)</th>
<th>χ² (df=4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substance use disorders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cocaine dependence</td>
<td>0(0.0)</td>
<td>770(62.1)</td>
<td>156(86.7)</td>
<td>3207(98.4)</td>
<td>1907(99.5)</td>
<td>788.4</td>
</tr>
<tr>
<td>Nicotine dependence</td>
<td>663(19.7)</td>
<td>707(57.0)</td>
<td>112(62.2)</td>
<td>2331(71.6)</td>
<td>1568(81.8)</td>
<td>2164.8</td>
</tr>
<tr>
<td>Alcohol dependence</td>
<td>765(22.7)</td>
<td>537(43.3)</td>
<td>105(58.3)</td>
<td>2042(62.7)</td>
<td>1288(67.2)</td>
<td>1361.6</td>
</tr>
<tr>
<td>Opioid dependence</td>
<td>274(8.1)</td>
<td>559(45.0)</td>
<td>65(36.1)</td>
<td>946(29.0)</td>
<td>1389(72.5)</td>
<td>1723.1</td>
</tr>
<tr>
<td>Sedative dependence</td>
<td>32(1.0)</td>
<td>64(5.2)</td>
<td>12(6.7)</td>
<td>190(5.8)</td>
<td>301(15.7)</td>
<td>313.5</td>
</tr>
<tr>
<td>Stimulant dependence</td>
<td>15(0.5)</td>
<td>43(3.5)</td>
<td>13(7.2)</td>
<td>167(5.1)</td>
<td>216(11.3)</td>
<td>203.5</td>
</tr>
<tr>
<td>Other substance dependence</td>
<td>66(2.0)</td>
<td>160(12.9)</td>
<td>18(10.0)</td>
<td>340(10.4)</td>
<td>747(39.0)</td>
<td>1019.6</td>
</tr>
<tr>
<td>Psychiatric disorders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASPD</td>
<td>177(5.3)</td>
<td>135(10.9)</td>
<td>9(5.0)</td>
<td>484(14.9)</td>
<td>423(22.1)</td>
<td>316.9</td>
</tr>
<tr>
<td>MDE</td>
<td>400(11.9)</td>
<td>148(11.9)</td>
<td>37(20.6)</td>
<td>560(17.2)</td>
<td>395(20.6)</td>
<td>92.2</td>
</tr>
<tr>
<td>PTSD</td>
<td>195(5.8)</td>
<td>115(9.3)</td>
<td>30(16.7)</td>
<td>576(17.7)</td>
<td>413(21.6)</td>
<td>323.0</td>
</tr>
<tr>
<td>OCD</td>
<td>35(1.0)</td>
<td>18(1.5)</td>
<td>3(1.7)</td>
<td>93(2.9)</td>
<td>74(3.9)</td>
<td>49.5</td>
</tr>
<tr>
<td>Social Phobia</td>
<td>89(2.6)</td>
<td>42(3.4)</td>
<td>12(6.7)</td>
<td>148(4.5)</td>
<td>130(6.8)</td>
<td>54.9</td>
</tr>
<tr>
<td>Agoraphobia</td>
<td>87(2.6)</td>
<td>41(3.3)</td>
<td>15(8.3)</td>
<td>195(6.0)</td>
<td>161(8.4)</td>
<td>97.7</td>
</tr>
<tr>
<td>Panic disorder</td>
<td>86(2.6)</td>
<td>45(3.6)</td>
<td>16(8.9)</td>
<td>183(5.6)</td>
<td>201(10.5)</td>
<td>148.4</td>
</tr>
<tr>
<td>Compulsive gambling</td>
<td>95(2.8)</td>
<td>72(5.8)</td>
<td>19(10.6)</td>
<td>364(11.2)</td>
<td>236(12.3)</td>
<td>198.2</td>
</tr>
</tbody>
</table>

Note that χ² values were computed by performing Wald tests for coefficients in fitted generalized linear models, where cluster assignments served as independent variables to predict with/without the disorders. The coefficients were estimated with generalized estimation equation to account for dependence among members from same family.

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

Table S3: Cocaine use characteristics, cocaine-related effects, and cocaine treatment history by subtype [N(%)]

<table>
<thead>
<tr>
<th>Behaviors</th>
<th>Subtype 2 1241(12.5)</th>
<th>Subtype 3 180(1.81)</th>
<th>Subtype 4 3258(32.7)</th>
<th>Subtype 5 1916(19.2)</th>
<th>χ²(df=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age of first cocaine use in year</td>
<td>21.1(5.7)</td>
<td>36.5(11.8)</td>
<td>21.4(5.2)</td>
<td>17.9(4.3)</td>
<td>1057.8</td>
</tr>
<tr>
<td>Mean age of onset of heaviest cocaine use in year</td>
<td>27.2(8.2)</td>
<td>45.1(5.0)</td>
<td>29.3(7.4)</td>
<td>25.8(8.4)</td>
<td>2158.0</td>
</tr>
<tr>
<td>Injected cocaine intravenously</td>
<td>861(69.4)</td>
<td>157(87.2)</td>
<td>3026(92.9)</td>
<td>1857(96.9)</td>
<td>540.8</td>
</tr>
<tr>
<td>Stayed high from cocaine for a whole day or more</td>
<td>388(31.3)</td>
<td>44(24.4)</td>
<td>129(3.9)</td>
<td>1417(74.0)</td>
<td>1804.4</td>
</tr>
<tr>
<td>Strong desire for cocaine made it hard to think of anything else</td>
<td>594(47.9)</td>
<td>131(72.8)</td>
<td>2628(80.7)</td>
<td>1623(84.7)</td>
<td>576.2</td>
</tr>
<tr>
<td>Family members, friends, doctor, clergy, boss, or people at work or school objected to cocaine use</td>
<td>404(32.6)</td>
<td>118(65.6)</td>
<td>2532(77.7)</td>
<td>1619(84.5)</td>
<td>933.9</td>
</tr>
<tr>
<td>Been arrested or had trouble with the police because of cocaine use</td>
<td>318(25.6)</td>
<td>111(61.7)</td>
<td>2475(76.0)</td>
<td>1650(86.1)</td>
<td>1132.4</td>
</tr>
<tr>
<td>Give up or greatly reduced important activities due to cocaine use</td>
<td>419(33.8)</td>
<td>131(72.8)</td>
<td>2853(87.6)</td>
<td>1775(92.6)</td>
<td>1418.3</td>
</tr>
<tr>
<td>Ever treated for an cocaine-related problem</td>
<td>432(34.8)</td>
<td>109(60.6)</td>
<td>2372(72.8)</td>
<td>1421(74.2)</td>
<td>585.6</td>
</tr>
<tr>
<td>Ever attended self-help group for cocaine use</td>
<td>363(29.3)</td>
<td>90(50.0)</td>
<td>2186(67.1)</td>
<td>1369(71.5)</td>
<td>601.2</td>
</tr>
</tbody>
</table>

Note that χ² values were computed by performing Wald tests for coefficients in fitted generalized linear models, where cluster assignments served as independent variables to predict cocaine use behaviors. The coefficients were estimated with generalized estimation equation to account for dependence among members from same family.

DOI: 10.1503/jpn.180098

© Joule Inc.

*Online appendices are unedited and posted as supplied by the authors.*

Table S4: Demographic characteristics by subtype [N(%)]

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Subtype 1 3370(33.8)</th>
<th>Subtype 2 1241(12.5)</th>
<th>Subtype 3 180(1.81)</th>
<th>Subtype 4 3258(32.7)</th>
<th>Subtype 5 1916(19.2)</th>
<th>(\chi^2) (df=4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>39.5(14.2)</td>
<td>39.71(10.0)</td>
<td>52.39(6.4)</td>
<td>40.89(8.1)</td>
<td>39.20(9.5)</td>
<td>591.8(4)</td>
</tr>
<tr>
<td>Sex[N(%)]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>262.7(4)</td>
</tr>
<tr>
<td>Men</td>
<td>1466(43.5)</td>
<td>750(60.4)</td>
<td>104(57.8)</td>
<td>1816(55.7)</td>
<td>1226(64.0)</td>
<td>913.8(4)</td>
</tr>
<tr>
<td>Women</td>
<td>1904(56.5)</td>
<td>491(39.6)</td>
<td>76(42.2)</td>
<td>1442(44.3)</td>
<td>690(36.0)</td>
<td>322.3(4)</td>
</tr>
<tr>
<td>Self-reported Race[N(%)]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>638.7(12)</td>
</tr>
<tr>
<td>AA</td>
<td>1303(38.7)</td>
<td>549(44.2)</td>
<td>94(52.2)</td>
<td>1969(60.4)</td>
<td>580(30.3)</td>
<td>913.8(4)</td>
</tr>
<tr>
<td>EA</td>
<td>1612(47.8)</td>
<td>490(39.5)</td>
<td>66(36.7)</td>
<td>793(24.3)</td>
<td>897(46.8)</td>
<td>322.3(4)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>201(6.0)</td>
<td>107(8.6)</td>
<td>7(3.9)</td>
<td>243(7.5)</td>
<td>240(12.5)</td>
<td>1495.4(12)</td>
</tr>
<tr>
<td>Other</td>
<td>253(7.5)</td>
<td>94(7.6)</td>
<td>13(7.2)</td>
<td>253(7.8)</td>
<td>199(10.4)</td>
<td>634.5(8)</td>
</tr>
<tr>
<td>Education[N(%)]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1495.4(12)</td>
</tr>
<tr>
<td>No HS</td>
<td>60(1.8)</td>
<td>73(5.9)</td>
<td>5(2.8)</td>
<td>138(4.2)</td>
<td>168(8.8)</td>
<td></td>
</tr>
<tr>
<td>Some HS</td>
<td>382(11.3)</td>
<td>332(34.8)</td>
<td>51(28.3)</td>
<td>1172(36.0)</td>
<td>832(43.4)</td>
<td></td>
</tr>
<tr>
<td>HS graduate</td>
<td>656(19.5)</td>
<td>397(32.0)</td>
<td>63(35.0)</td>
<td>950(29.2)</td>
<td>472(24.6)</td>
<td></td>
</tr>
<tr>
<td>Beyond HS</td>
<td>2271(67.4)</td>
<td>338(27.2)</td>
<td>61(33.9)</td>
<td>997(30.6)</td>
<td>442(23.1)</td>
<td></td>
</tr>
<tr>
<td>Marital status[N(%)]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>634.5(8)</td>
</tr>
<tr>
<td>Never married</td>
<td>1731(51.4)</td>
<td>789(63.6)</td>
<td>47(26.1)</td>
<td>1920(58.9)</td>
<td>1135(59.2)</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>1025(30.4)</td>
<td>144(11.6)</td>
<td>20(11.1)</td>
<td>398(12.2)</td>
<td>210(11.0)</td>
<td></td>
</tr>
<tr>
<td>Div/Sep/Wid</td>
<td>614(18.2)</td>
<td>308(24.9)</td>
<td>113(62.8)</td>
<td>940(28.9)</td>
<td>571(29.8)</td>
<td></td>
</tr>
</tbody>
</table>

Note that \(\chi^2\) values were computed by performing Wald tests for coefficients in fitted generalized linear models, where cluster assignments served as independent variables to predict the demographic variables. The coefficients were estimated with generalized estimation equation to account for dependence among members from same family.


