

MIT Open Access Articles

Convergence of case-specific epigenetic alterations identify a confluence of genetic vulnerabilities tied to opioid overdose

The MIT Faculty has made this article openly available. ***Please share*** how this access benefits you. Your story matters.

Citation: Corradin, Olivia, Sallari, Richard, Hoang, An T, Kassim, Bibi S, Ben Hutta, Gabriella et al. 2022. "Convergence of case-specific epigenetic alterations identify a confluence of genetic vulnerabilities tied to opioid overdose." *Molecular Psychiatry*, 27 (4).

As Published: 10.1038/S41380-022-01477-Y

Publisher: Springer Science and Business Media LLC

Persistent URL: <https://hdl.handle.net/1721.1/146782>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike





HHS Public Access

Author manuscript

Mol Psychiatry. Author manuscript; available in PMC 2022 September 17.

Published in final edited form as:

Mol Psychiatry. 2022 April ; 27(4): 2158–2170. doi:10.1038/s41380-022-01477-y.

Convergence of case-specific epigenetic alterations identify a confluence of genetic vulnerabilities tied to opioid overdose

Olivia Corradin^{1,2}, Richard Sallari³, An T. Hoang¹, Bibi S Kassim⁴, Gabriella Ben Hutta⁴, Lizette Cuoto⁴, Bryan C. Quach⁵, Katreya Lovrenert⁶, Cameron Hays⁷, Berkley E. Gryder⁶, Marina Iskhakova⁴, Hannah Cates⁴, Yanwei Song¹, Cynthia F. Bartels⁶, Dana B. Hancock⁵, Deborah C. Mash⁸, Eric O. Johnson^{5,9}, Schahram Akbarian⁴, Peter C. Scacheri^{6,10,*}

¹Whitehead Institute Biomedical Research

²Department of Biology, Massachusetts Institute of Technology

³Axiotl Inc. Cleveland OH

⁴Department of Psychiatry and Department of Neuroscience, Friedman Brain Institute, Icahn School of Medicine at Mount Sinai

⁵GenOmics, Bioinformatics, and Translational Research Center, Biostatistics and Epidemiology Division, RTI International, Research Triangle Park, NC 27709

⁶Department of Genetics and Genome Sciences, Case Western Reserve University

⁷College of Health and Human Services, Western Michigan University

⁸Dr. Kiran C. Patel College of Allopathic Medicine, Nova South Eastern University

⁹Fellow Program, RTI International, Research Triangle Park, NC 27709, USA.

¹⁰Case Comprehensive Cancer Center, Case Western Reserve University

Abstract

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: <https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms>

*correspondence to: peter.scacheri@case.edu.

Author Contributions

This work was conceptualized and written by P.C.S., S.A., E.O.J, D.C.M., D.B.H, R.S. and O.C. ChIP-seq and RNA-seq experiments were completed by S.A., B.S.K., G.B.H., L.C., K.L., M.I., H.C., and C.F.B. Data analysis was performed by, O.C., R.S., A.T.H., B.C.Q, K.L., and Y.S. Discussion and feedback was provided by C.H. and B.E.G.

Conflict of Interest

Peter Scacheri received compensation as a consultant for Kronos Bio.

Code Availability

Code for the machine learning and VEL analyses are available through: <https://github.com/corradin-lab/DLPFC-opioid-overdose>.

Convergence analysis code is available through: <https://axiotl.com/> and <https://convergence.axiotl.com/>

Data Access

ChIP and RNA seq data are available through dbgap accession number: phs002724.v1.p1

Human Subjects Research

The epigenomic analyses performed herein are Category 4 research, on previously collected, de-identified tissue specimens. No human subject identifiers were included. Tissues were previously collected by the University of Miami Brain Endowment Bank under direction of Deborah C. Mash.

Opioid use disorder is a highly heterogeneous disease driven by a variety of genetic and environmental risk factors which have yet to be fully elucidated. Opioid overdose, the most severe outcome of opioid use disorder, remains the leading cause of accidental death in the United States. We interrogated the effects of opioid overdose on the brain using ChIP-seq to quantify patterns of H3K27 acetylation in dorsolateral prefrontal cortical neurons isolated from 51 opioid-overdose cases and 51 accidental death controls. Among opioid cases, we observed global hypoacetylation and identified 388 putative enhancers consistently depleted for H3K27ac. Machine learning on H3K27ac patterns predicted case-control status with high accuracy. We focused on case-specific regulatory alterations, revealing 81,399 hypoacetylation events, uncovering vast inter-patient heterogeneity. We developed a strategy to decode this heterogeneity based on convergence analysis, which leveraged promoter-capture Hi-C to identify five genes over-burdened by alterations in their regulatory network or “plexus”: *ASTN2*, *KCNMA1*, *DUSP4*, *GABBR2*, *ENOX1*. These convergent loci are enriched for opioid use disorder risk genes and heritability for generalized anxiety, number of sexual partners, and years of education. Overall, our multi-pronged approach uncovers neurobiological aspects of opioid use disorder and captures genetic and environmental factors perpetuating the opioid epidemic.

Introduction

Opioid use disorder (OUD) and the rise of opioid overdose deaths have taken a devastating emotional and economic toll on society. This public health crisis has only been exacerbated during the COVID19 pandemic with reports of a 16 to 25% increase in opioid-related overdose deaths in the US (1). Uncovering the neurobiological basis of OUD is urgently needed to drive prevention and treatment interventions. However, current GWAS, with over 10,000 cases, have only replicably identified the *OPRM1* gene (2)(3), making it an especially challenging phenotype to study. Conservative estimates suggest that hundreds of thousands of samples will be needed to uncover a majority of risk loci (4). Animal models have produced valuable insights into the biology of disease, but the extent to which these will translate to the complexity of human behavioral traits remains unclear. In lieu of larger GWAS sample sizes that are not currently available from human studies, we present a methodological innovation to shortcut discovery of genes involved in the neurobiology of OUD. We further identify novel genes and pathways that point us towards biology relevant for understanding OUD and opioid overdose.

Here, we reason that a successful approach would encompass three key features: (1) the accurate measurement of epigenomic activity and connectivity to genes, (2) the importance of making these measurements in the appropriate cellular context, and (3) the need to harness heterogeneous, case-specific alterations. We pioneer such an analysis in OUD through the analysis of histone acetylation landscapes in opioid overdose cases vs. accidental sudden death (physiologically normal) controls. We focus on neurons from dorsolateral prefrontal cortex (DLPFC) as a cell type broadly involved in the neural circuitry of OUD and addiction more broadly, as well as, the frequently co-occurring disorders of anxiety and impulsivity. Moreover, animal and human studies have drawn a firm link between long-term vulnerability to addiction and cellular and molecular adaptations in the brain’s reward-processing circuitry (5), including changes in gene expression programs in

the prefrontal cortex and other target areas of the mesocorticolimbic dopamine system (6). These include the dorsolateral PFC, which by functional neuroimaging in subjects diagnosed with substance abuse, shows abnormal activation profiles in context of drug exposures, motivational drive and inhibitory control over drug-related behaviors (including intoxication and binge) that critically depend on the balanced interaction between DLPFC and the more ventral portions of the PFC (7). We therefore reasoned that mapping the epigenome of DLPFC in opioid overdose death cases vs. controls provides an important context to discover regulatory loci important in mediating addictive behaviors.

Tissue-specific regulatory elements consistently show a high correlation with GWAS catalog risk loci in phenotypically relevant tissue types (8)(9)(10). We focus our epigenomic assays on H3K27ac, the canonical mark of active regulatory elements. We interrogate neuronal H3K27ac epigenomes of DLPFC from opioid overdose cases compared with controls. Moreover, by analyzing H3K27ac epigenomes in the context of promoter-bound chromosomal conformations, using a promoter-capture Hi-C reference set from the DLPFC (11), we are able to connect individual H3K27ac peaks to promoter targets with high confidence.

Previous H3K27ac epigenomic profiling of colon cancer, metastatic osteosarcoma and numerous other types of cancer (12)(13)(14) has revealed stark tumor-specific alterations in enhancer activity. We termed these VELs (Variable or Variant Enhancer Loci). Unlike these observations from rapidly dividing cells undergoing a constant evolution from normal to malignant, here we study post-mitotic neurons in the human brain. Thus, we do not anticipate dramatic changes in enhancer activities in this setting. This requires a rethinking of our theoretical framework; instead of single clear changes with potent effects on gene expression, we transition to dispersed, subtle effects that cumulatively tweak gene expression. We hypothesize that direct ascertainment of subtle yet converging heterogeneous variants present in the brain epigenome of fatal opioid overdose cases will illuminate the biological and genetic underpinnings of the traits involved in OUD.

RESULTS

Identification of homogeneous H3K27ac hypoacetylation events

We generated H3K27ac ChIP-seq data from neuronal nuclei sorted from postmortem DLPFC dissected from 60 confirmed opioid overdose death cases and 58 age-matched, sudden accidental death controls. Opioid cases had a documented history of opioid abuse, positive toxicology report for opioids, and forensic determination of opioids as the cause of death. Controls were age-matched from sudden accidental or cardiac deaths with negative urine screens for all common drugs and no history of licit or illicit drug use. Samples with known history of psychiatric disorders, debilitating chronic pain or death by suicide were excluded from both groups. A total of 51 cases and 51 controls passed stringent quality control (**Methods**, Table S1) and were selected for further analysis (Figure 1A, S1). These samples were sequenced to a mean read depth per sample of 112,215,946 (s.d. = 33,175,526), with no difference in read depth between case and control groups (Welsh unpaired two-sample t-test, two-sided $P=0.605$, $n=102$ biologically independent samples, read count mean difference \pm SEM: 3418882 ± 6593642 ; 95% confidence

interval (CI): -9684412-16522176; $t=0.5185$, $df=88.09$) (Figure S1). To our knowledge, this represents one of the most extensive analyses of H3K27ac in human neuronal DLPFC performed to date, and the only one focused on opioid overdose. Across the cohort, we detected 655,505 H3K27ac peaks. We utilized a linear model to regress out covariates, including age, postmortem interval, ancestry and sex, and used residuals in subsequent T tests (**Methods**, Figure S1D,E). We observed modest but widespread hypoacetylation of H3K27 in cases versus controls, with 388 peaks showing statistically significant differences ($P_{\text{Bonf.}} < 1 \times 10^{-7}$) (Figure 1B,C, Table S2). An exemplar locus linked to target gene dual-specificity phosphatase 4 (*DUSP4*) in DLPFC promoter capture Hi-C data, involved in the addiction-relevant GPCR and MAPK signaling pathways (15,16), is shown in Figure 1D. Strikingly, no significant hyperacetylated peaks in cases were found. Of the 388 significant hypoacetylated peaks, 95.9% were located distal (> 2 kb) to transcription start sites and are putative active regulatory elements. Motifs corresponding to AP1 transcription factors, including FOS, ATF3, and JUNB ($P < 1 \times 10^{-15}$) were highly enriched at these regulatory elements (Figure 1E). While opioid exposed promoter capture data was not available, we identified putative gene targets of the differentially acetylated loci using promoter-capture Hi-C data available from control DLPFC (11).

We performed RNA-seq analysis of a subset of the cohort (24 cases, 27 controls, bulk DLPFC, not neuronal nuclei sorted) (Figure 1F, Figure S1F). While bulk tissue profiling limits the ability to detect cell type specific expression changes, we nonetheless identified ten differentially expressed genes, 8 downregulated, 2 upregulated in cases compared to controls (Figure 1G, $P_{\text{Bonf.}} < 2.4 \times 10^{-6}$). The AP1 transcription factor *FOS* was among the significantly reduced genes. Genes assigned by promoter capture Hi-C to the 388 hypoacetylated regulatory elements generally show a modest but statistically significant decrease in gene expression levels in cases compared to controls ($P < 0.05$) (Figure 1H). Overlap of the significant epigenome and transcriptome results highlights 3 genes: *DUSP4*, *DUSP6*, *ETV5* (Figure 1G). *DUSP4* and *DUSP6* regulate MAPK-ERK signaling. Consistent with this known biological function, the MAP kinase pathway was the most enriched GO molecular function term associated with the set of differentially expressed genes in opioid cases versus controls ($P_{\text{adj}} < 0.01$) (Figure 1I) (17,18). Both the input transcription factors to the 388 loci (FOS AP1 factors) and output MAPK genes are consistent with known addiction pathways (19)(15)(6)(16)(20).

Seven loci can accurately predict case/control status

The 388 hypoacetylated loci identified above represent only a small proportion of the hundreds of thousands of ChIP-seq peaks evaluated. We therefore applied machine learning (ML) as an agnostic strategy to re-evaluate the entirety of the data. We used Gradient Boosting classification algorithm with ChIP-seq peaks and covariates as potential features to distinguish case/control status (**Methods**) (21). Remarkably, seven ChIP-seq peaks can differentiate opioid cases from controls with high accuracy (5-fold cross-validated mean AUC = 0.972, s.d.=0.037) (Figure 2A). This outperformed models that used only covariates, or the 388 hypoacetylated peaks ($P < 0.001$) (Figure 2B). Given the global hypoacetylation observed in this dataset, we reasoned that these 7 peaks may not be the only features capable of distinguishing opioid cases. To identify additional peak sets with similar predictive

accuracy, we developed a stepwise strategy which iteratively removed peaks used by previous models and then identified new models from the remaining peaks (Figure 2C). In total, we identified 108 independent models ranging from 4 to 43 peaks with median AUCs >0.9 (Supplemental Figure 2A,B, Table S3). Altogether, these 108 models incorporated 3200 peaks. Using unbiased Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction (22), we found that these features could easily distinguish opioid cases from controls (Figure 2D) and that their separation was not observed when utilizing all ChIP-seq peaks (Figure 2E). Using DLPFC promoter-capture Hi-C data, we identified 830 gene targets of model peaks. While individual peaks were unique to each model, we found 167 (20%) target genes that were incorporated across multiple models. Genes found by three or more models include *DUSP4* and the potassium channel *KCNMA1* (Figure 2F). Using Functional Mapping and Annotation (FUMA) (23), we found that gene targets of model peaks were enriched for genes involved in six different pathways. This included several pathways broadly involved in synapse function and behavior such as regulating GABAergic transmission, G protein signaling, and neuronal phosphodiesterases and sex steroid (incl. androgen) receptor pathways (Figure 2G). Collectively, these results demonstrate that while individual changes in regulatory elements are small, they are sufficient to distinguish opioid overdose death cases from controls. More importantly, the repeated appearance of synaptic gene targets in the 108 models, and the use of peaks beyond the 388 in these models, suggests the heterogeneity amongst individual cases is converging on shared underlying biology.

Identification of heterogeneous case-specific alterations

In order to uncover the epigenetic heterogeneity of opioid overdose, we used an information-theoretic approach to identify changes that distinguish individual cases from the full set of controls (Figure 3A, **Methods**). A similar strategy has been used to pinpoint key regulatory differences among heterogeneous cancers (12)(13)(14). We refer to these individual-specific changes as Variable or Variant Enhancer Loci, or VELs. The vast majority of differentially acetylated regions identified through VEL-mapping were hypoacetylation events, consistent with our prior findings. Specifically, we identified 81,399 VELs with reduced H3K27ac, or “lost VELs”, in at least one opioid overdose death case relative to controls (FDR<5%), and 976 gained VELs (Table S4). Of the 388 hypoacetylated regions, all but two were detected via the VEL-mapping approach. On average, the magnitude of the effect was far greater for H3K27ac differences found through VEL-mapping (Figure 3B) than for hypoacetylation analysis in Figure 1, suggesting that case-specific changes may be larger contributors to phenotype than the hypoacetylation events revealed through case/control analysis. VEL gene targets showed changes in expression consistent with the accumulation of gains and losses, albeit asymmetrically, with gains showing a more immediate effect (Figure 3C). The number of VELs shared across cases is far greater for losses than gains (Figure S3).

Case-specific lost VELs showed striking convergence at the gene level. Upon linking VELs to their putative gene targets using DLPFC promoter-capture Hi-C data, we found that while only 10% of lost VELs occurred in two or more overdose cases, 60% of target genes were physically linked to VELs in two or more overdose cases (Figure 3D). As we looked for VEL peaks and VEL target genes that were found in a larger number of cases, the

percentage of genes and peaks decreased exponentially (Figure 3D). However, the ratio of VEL target genes to VEL peaks (convergence ratio) increased exponentially, indicating that heterogeneous case-specific losses accumulated at the gene level in a large number of cases. In other words, despite vast heterogeneity at the level of individual enhancers, the signal appeared to “yield” at the gene level, becoming increasingly homogenous. We next sought out to capture this homogeneity and resolve it into concrete genes at which these alterations converged through the development of a novel statistical framework which we refer to as “convergence analysis.”

Regulatory losses converge on synaptic genes

Convergence analysis is designed to resolve genomically dispersed, low-frequency variation in either the DNA sequence or epigenomic signals into high-frequency convergent events. We use the term “plexus” to refer to the collection of regulatory elements that physically contact the gene promoter (24)(25). Plexi are derived from DLPFC promoter-capture Hi-C (11). Here, we used convergence analysis to identify genetic loci and target genes that have an overabundance of opioid case VELs. The strategy identifies single elements in a gene plexus altered across multiple cases, multiple elements altered within a single case, multiple elements altered across multiple cases, and intermediate patterns (Figure 4A). This approach takes into account the differing number of elements contained in each gene’s plexus and their activity to identify genes with the most significant accumulation of VELs.

Five convergent plexi were statistically significant for accumulation of lost VELs: *ASTN2*, *KCNMA1*, *DUSP4*, *GABBR2*, and *ENOX1* (P_{FDR} : 1×10^{-11} , 1×10^{-7} , 1×10^{-7} , 1×10^{-6} , 1×10^{-6} ; Figure 4B,C, Table S5). We did not identify significant plexi for the accumulation of gained VELs. The *ASTN2* plexus, the most significant result, is composed of 106 putative regulatory elements contained in loci that interact with the *ASTN2* promoter in DLPFC; 65.1% of its elements and 21.6% of the cases have at least one VEL (plexus and cohort convergence, respectively). *KCNMA1* showed the highest single-patient burden of losses: 36.3% (41 out of 113 elements). *DUSP4* showed the highest level of cohort convergence: 37.3% (19 of the 51 patients); however, in this plexus, the highest frequency of losses at any single element is 9.8% (5 of the 51 cases), highlighting the ability of the convergence framework to decode case heterogeneity. *GABBR2* showed the highest plexus convergence: 84.6% (88 out of 104 elements). The plexi of these five genes do not overlap, indicating that the significance of the results was mutually independent.

Here, we were able to distill 81,399 case-specific VEL alterations into five genes where global hypoacetylation was concentrated beyond what is expected by chance, showing clear erosion of regulatory activity across each gene’s plexus. Conventional strategies seek to identify robust changes that distinguish all cases from controls. The VEL convergence strategy offers a novel opportunity to take full advantage of the richness of information contained in the network of subtle and heterogenous regulatory changes.

Convergent losses account for heritability of anxiety and overlap OUD risk genes

We hypothesized that direct ascertainment of converging heterogeneous effects of opioid exposure on the human brain’s epigenome will illuminate the genetic underpinnings of the

traits involved in OUD. To evaluate this hypothesis, we compared the top 5 convergent plexi to the heritability of a panel of 27 traits (10) (Figure 5A,B, Table S6). This included summary statistics from the MVP study of OUD with 10,544 cases and 72,163 controls with one replicable finding (*OPRM1*) (2). We did not observe heritability enrichment for OUD, potentially owing to low power of the OUD GWAS; alternatively, as *OPRM1* enhancers were not differentially acetylated, it is possible that opioid overdose death is epigenetically heritable at loci distinct from the genetically heritable loci. Similar to our findings, recent postmortem studies have reported that *OPRM1* transcript levels in the anterior insula (a cortical node in reward circuitry heavily interconnected with DLPFC) as well as in the DLPFC are not significantly affected by opioid use or SUD diagnosis (26,27).

In evaluating other phenotypes, the top 5 gene's plexi account for a significant proportion of heritability of generalized anxiety disorder (N=200K), number of sexual partners (N=370K) and educational attainment (N=126K) (Figure 5C). Given that regulatory regions active in the CNS have been previously shown to explain a substantial proportion of the heritability of neuropsychiatric disorders and behavioral phenotypes, we compared these results to control sets of cortical enhancers with similar H3K27ac patterns as the VELs. VELs associated with the top 5 plexus genes were substantially more enriched for heritability of these three phenotypes, with generalized anxiety disorder showing nearly 4-fold more heritability enrichment than activity-matched cortical enhancer regions (Figure 5C). These traits have been shown to genetically correlate with OUD and generalized anxiety disorder is also frequently co-morbid with OUD (28)(2)(29). These results suggest that previously identified genetic vulnerabilities to traits linked with OUD co-localize with the convergent epigenetic changes we identified in our opioid overdose cohort. Notably, amongst all identified lost VELs, we observed heritability enrichment for additional traits including alcohol use disorder, drinks per week, smoking initiation, smoking cessation and measures of risk tolerance. However, in contrast to the top 5 plexi, heritability enrichment for VEL regions was equal to the enrichment found in other regulatory elements active in the neurons (Supplemental Figure 4A,B).

We expanded our analysis to the entire list of plexi ranked genes to further investigate overlap with OUD GWAS (2). We directly assessed whether loci with greater VEL convergence were enriched among the strongest OUD-associated SNPs as sorted by GWAS p-values. We identified genes nearest the top 1%, 0.5%, and 0.05% of OUD GWAS SNP associations. Using GSEA (30,31), we found these genes to be significantly enriched amongst genes with the greatest VEL convergence (Figure 5D). Of the top 40 genes with the greatest VEL convergence, 23 (57.5%) were the nearest gene to a SNP in the top 1% of the OUD GWAS results (Figure 5E). Interestingly, of these 40 target genes, 8 were also identified through our earlier hypoacetylation analysis (Figure 1) and 21 through ML analysis (Figure 2). These 40 also include multiple genes from pathways enriched in previous analyses (Figure 5E). These results suggest that our multi-pronged approach may be effective at revealing loci that predispose to OUD and its related traits.

Discussion

Neuropsychiatric traits like OUD and its most severe outcome, opioid overdose death, are complex and are best addressed by multiple complementary epigenomic analyses. Our new approach embraces the deep heterogeneity of complex traits, by employing linear regression, T tests, ML, VEL mapping and convergence analysis to identify novel genes associated with opioid overdose. Our novel VEL-convergence approach recovers the rich, case-specific variation that would be discarded in a conventional study, either because their effect sizes would be too small in aggregate across subjects, scattered across the genome, or case-specific. The approach we use here leverages the topology of the genome to decode heterogeneous variation, resolving low-frequency patterns of variation across all regulatory elements connected to a gene into more detectable high-frequency alterations at the gene level. *DUSP4* provides a good example of the value of our ensemble approach: it is connected to an element identified by T tests, and it shows a transcriptional change; however, it also harbors a preponderance of case-specific alterations, observed in 37.3% of the cases. Without the convergence approach, this profuse effect would have gone unnoticed. Most importantly, the convergence framework reveals additional genes with even more extensive burden of alterations that would have been completely missed, as they don't fall near any of the 388 LR-identified elements.

We use three different approaches to characterize epigenetic events in opioid overdose cases. Across all three approaches, we identified *KCNMA1*, *DUSP4*, and *GABBR2* as critical target genes. Our analyses implicate MAPK signaling, GABA signaling, and potassium channel genes. Each of these three pathways have been broadly linked to addiction. For example, *KCNMA1* is a large conductance potassium channel previously linked to alcohol dependence (32) and considered a hub gene for alcohol-induced transcriptomic alterations in the prefrontal cortex (33). *DUSP4* encodes a MAP kinase broadly relevant for synaptic plasticity and memory functions in the brain (34). *GABBR2* encodes a metabotropic, G-protein coupled GABA receptor and potential OUD risk gene (35), with altered gene expression in the prefrontal cortex in an animal model (36). Finally, common polymorphisms in Astrotactin 2 (*ASTN2*), a membrane-bound regulator of the surface expression of many synaptic proteins (37), exert a surprisingly strong effect on postoperative pain phenotypes, including patient-controlled dosing of opioid-based analgesics (38).

The approach pioneered here is able to uncover some of the biology of OUD with as few as 51 cases and 51 controls. Expanding to larger cohorts and other addiction-relevant brain regions beyond DLPFC may yield many more genes for further investigation as either biomarkers of opioid exposure, potential druggable targets, or causal genes that underlie risk of OUD, similarly to how one would functionally characterize GWAS hits. Coupling this approach to single cell strategies could further inform specific neuronal or non-neuronal cell subtypes that are especially affected by opioid abuse. Our approach does not replace GWAS and animal studies, as both are critical in advancing our understanding of complex traits, particularly for traits that disproportionately impact hard-to-reach populations. The hybrid experimental and computational strategy that we present here is designed to augment or even accelerate the mapping of complex genetic architectures.

What have we uncovered with this new methodology, and what is the relevance to the current opioid crisis? We believe our study is capturing shards of the genetic and environmental factors perpetuating the opioid epidemic. In fact, our top 5 plexus-genes are enriched among the top 1% of OUD GWAS. While we did not observe global enrichment for OUD heritability, this is not unexpected given the incompleteness of the OUD genetic architecture, as the largest OUD GWAS captures <12% of the expected heritability. Alternatively, the most severe outcome of OUD, opioid overdose death, may be driven by genetic and/or environmental factors that are either entirely distinct from OUD, or are comprised of a particular subset of OUD risk factors. This further highlights the need for approaches which simultaneously capture the genetic and epigenetic influences of the disorder.

The three traits uncovered in the top 5 gene plexi - anxiety, years of education and number of sexual partners - have been associated with early childhood adversity (39)(40)(41). All of these traits as well as early childhood adversity are major risk factors for OUD (42)(43)(44). This observation points to a potential shared pleiotropic risk that may be revealed as the risk factors for all of these related traits are uncovered. Early life adversity (ELA) and resource scarcity has been shown to have lasting impacts on the epigenome in both rodent and human studies (45)(46), and thus has the potential to play a causal role in some of the epigenetic changes we identified. ELA is an increasingly recognized contributor to many public health crises (47). We expect that genetic and environmental factors, such as ELA and drug exposure, may cooperatively contribute to the epigenetic changes we have identified. Risk factors in adulthood for opioid overdose death, such as lack of social support, exposure to opioids, and lacking economic opportunity have been rising steadily over the past few decades, driven by shifting socioeconomic factors and changes in opioid availability (48)(49). Angus Deaton and colleagues called these changes “the ongoing erosion of working-class life for those born after 1950” (50). In their landmark study, they describe a generational increase in self-reported pain in the United States, which they tie to the opioid epidemic. Intriguingly, two variants in our top result, *ASTN2*, are associated with sensitivity to opioids in a study of pain-related phenotypes of fentanyl use after surgery (38), suggesting we may also be identifying variation in pain response. Furthermore, genetic and environmental risk factors may influence the overlapping sets of target genes through distinct mechanisms. Epigenomic studies like ours are uniquely poised to identify the loci most critically affected by this milieu of genetic and environmental influences.

Our study identifies novel elements of the OUD exposure and risk architecture, as well as phenotypes related to early life adversity, expanding the field’s current limited view of the underlying influences that contribute to the opioid epidemic. Altogether our results indicate that we are seeing just the tip of the iceberg of the intertwined causes, consequences of opioid exposure in the brain, and related traits that ultimately lead to opioid related fatalities. The epigenome is a palimpsest where inherited epigenetic and genetic patterns are repeatedly overwritten by environmental exposures throughout our lives. The ensemble of methods we present here allow us to reveal previously-hidden genes. Further studies are needed to distinguish the nature of the factors leading to their dysregulation. Expanding both epigenetic and genetic studies of OUD and related phenotypes will enable us to further elucidate how diverse contributors manifest in the brain and drive this devastating epidemic.

METHODS

Patient cohort

Cases (N= 51) were selected from an opportunistic sample of opioid-related deaths and unaffected controls that came to autopsy. The use of de-identified cadaver specimens is not defined as human subjects research and exempt from regulation 45 CFR pt 46 (NIH SF424 guide Part II: Human Subjects). The cause of death was determined by forensic pathologists following medico-legal investigations evaluating the circumstances of death including medical records, police reports, autopsy findings and forensic toxicology analysis. Common drugs of abuse and alcohol and positive urine screens were confirmed by quantitative analysis of blood and brain. The detection of 6-acetyl morphine (6-AM) was taken as definitive evidence of acute heroin exposure. Retrospective chart reviews were conducted to confirm history of opioid abuse, methadone or addiction treatment, drug related arrests or drug paraphernalia found at the scene. Supplemental brain toxicology was done on select cases for comparison to blood levels at the time of death. Case inclusion in the opioid group was based on a documented history of opioid abuse, toxicology report positive for opioids, and forensic determination of opioids as cause of death. Drug-free control subjects (N=51) were age-matched from sudden accidental (motor vehicle accidents or trauma) or cardiac deaths with negative urine screens for all common drugs and no history of licit or illicit drug use reported by next-of-kin informants prior to death. Exclusion criteria for either group included death by suicide, a known history of psychiatric disorder, or debilitating chronic illness. The subject groups were matched as closely as possible for age, gender, and postmortem interval (PMI). Neurons were isolated via sorting of NeuN+ cells (Catalog number: MAB377X, vendor: Sigma Aldrich, manufacturer: Chemicon) following ChIP-seq for H3K27ac (Catalog number: 39133, vendor: Active Motif) as previously described (Kundakovic et al. 2017).

ChIP-Seq data processing

Cutadapt v1.9.1 (51) was used to remove paired-end adapter sequences and discard reads with a length less than 20bp. All fastqs were aligned to the hg19 genome assembly (retrieved from hgdownload.cse.ucsc.edu/goldenPath/hg19/chromosomes) using BWA-MEM v0.7.17-r1188 (52) with default parameters in paired-end mode. Output SAM files were converted to binary (BAM) format, sorted, indexed, and PCR duplicates were removed using SAMtools v1.10 (53). Peaks were detected with MACS v2.1.2 (54) with the --broad flag set. DeepTools v3.2.0 (55) was used to generate RPGC-normalized bigWig tracks with 50 bp bin sizes of the final sample BAM files and visualized on the Integrative Genomics Viewer (56) in order to eliminate samples with pronounced track irregularities or low signal-to-noise ratio. The Bioconductor (57) package ChIPQC (58) was used to generate quality metrics for each processed library – those libraries which contained very few mapped reads, a RelativeCC enrichment score (RelCC) < 1, percentage of reads in peak regions (RiP%) < 2%, or number of detected peaks < 10,000 were excluded from additional analysis. Peak lists were filtered to remove all peaks overlapping ENCODE blacklisted regions (<https://sites.google.com/site/anshulkundaje/projects/blacklists>).

Normalization

We use the Inherent Normalization algorithm as part of the Axiotl Convergence Platform to process and normalize H3K27ac BigWig tracks. This normalization approach is designed specifically for epigenomes and uses a robust and reproducible reference found within the sample itself. In order to find this “inherent unit” we first focus on the most stable epigenomic elements in the genome: promoters. We use the set of 81,232 reference promoters from the Roadmap project [Roadmap2015]. This set of promoters contains a subset of constitutive elements (23,374), active in virtually all cell types, and a subset of highly cell-type-specific elements (16,807), active in only one reference cell type. We refer to the first list as the “on” promoters and to the second as the “off” promoters. By using 111 reference epigenomes, the group of “off” promoters will overwhelmingly be inactive in any cell type. Given the large number of elements in each list, we can compute very precise averages. The distance between the averages of the “on” and “off” collections constitutes our “inherent unit,” hence the name of the algorithm. The values of the “on” and “off” averages in each sample are then used to align the data, such that all the “off” averages across samples are at zero and all “on” averages are at one. Data is log₂ transformed so as to make the data normally distributed. The approach is described in more detail in Chakraborty et al. (In preparation), where the authors show consistent reduction of noise when reprocessing the ENCODE 3 reference epigenomes. Noise reduction translates to greater statistical power to discover new results in the data. The approach also makes the data biologically interpretable and universally meaningful, with a value of one representing the standard activity of a constitutive promoter.

ChIP-seq case/control analysis

Power Analysis: We used the package Web power (<https://rdrr.io/cran/WebPower/man/wp.regression.html>) with the following parameters to calculate power, $p_1 = 5$ (4 covariates (Age, Race, Gender, PMI) + 1 peak), $p_2 = 4$ (4 covariates), $\alpha = 0.05$, power = NULL. Our sample size of 102 achieves 80% power with an effective size (proportion of variance explained to variance unaccounted) > 0.08 .

Covariate and differential acetylation Analysis: We tested and corrected inherent normalized H3K27ac data for covariate effects using the following statistical procedures. Levene’s tests (R package “car”) showed homogeneity of variances for continuous covariates Age ($P = 0.1558$) and PMI ($P = 0.07561$). Chi-squared tests (R package “stats”) showed no difference for discrete covariates Race ($P = 0.8203$) and Gender ($P = 1$). T tests (R package “stats”) revealed differences in means for PMI ($P = 4.355e-06$) and not Age ($P = 0.3483$). We use a linear model (R package “stats”) to regress out all covariates. We performed T tests between cases and controls at each H3K27ac element on residuals from the linear model, correcting for all four covariates (Age, Race, Gender, PMI). The global hypoacetylation we report in opioid-exposed neurons naturally skews the p-value distribution obtained from the T tests, nevertheless we use uncorrected p-values as a heuristic measure to identify the most pronounced hyper- and hypoacetylated elements.

RNA-seq data processing and quality control

Postmortem, dorsolateral prefrontal cortex (Brodmann area 9) were dissected from cryopreserved coronal slices of brain to obtain bulk tissue RNA samples. 150 samples were prepared for sequencing using Takara Bio SMARTer Stranded RNA-seq Kit and 150 bp paired-end sequencing was performed on an Illumina. *Trimmomatic v0.39* (59) was used for read trimming and quality filtering with the following settings: *ILLUMINACLIP seedMismatches=2, ILLUMINACLIP palindromeClipThreshold=30, ILLUMINACLIP simpleClipThreshold=10, ILLUMINACLIP minAdapterLength=4, ILLUMINACLIP keepBothReads=true, LEADING quality=3, TRAILING quality=3, SLIDINGWINDOW windowSize=4, SLIDINGWINDOW requiredQuality=10, MINLEN length=75, AVGQUAL quality=20*. Read pairs with both reads passing quality filtering were mapped using *Salmon v1.1.0* (60) in selective alignment mode (61) with an index comprised of a GENCODE v30 transcriptome (comprehensive gene annotation) and the full GRCh38 primary assembly as a decoy. The options *--gcBias* and *--seqBias* were enabled to account for potential transcript quantification biases (62). Salmon transcript quantifications were aggregated to the gene level using *tximport v1.12.3* (63).

Samples were excluded based on the following criteria: >40% of reads removed by Trimmomatic, mean read GC content <40%, effective sequencing depth <10 million read pairs, transcriptomic mapping rate <0.3, RIN score <5, mitochondrial mapping rate >0.5, gene expression principal component analysis outliers (along principal components 1 and 2), age outlier (age at death >60 years), and self-reported race not African-American or White. Several RNA samples were derived from the same individuals (i.e., replicates). For each of these replicates sets, the sample with the highest effective sequencing depth was retained among the samples that passed quality control. Following all sample exclusions, 24 opioid intoxication cases and 27 controls were used for subsequent analysis.

Differential gene expression analysis

Genes were tested for differential expression by opioid overdose death case/control status using *DESeq2 v1.26.0* (64) to fit a generalized linear model with normalized gene counts as the outcome. Opioid intoxication status, ancestry, sex, age at death, PMI, RIN, RNA-seq sequencing batch, and 11 surrogate variables (SVs) were included in the model as explanatory variables. The 11 SVs were estimated using *SVA v3.36.0* (65) with variance stabilized gene counts and the aforementioned explanatory variables as null model variables. To reduce the multiple testing burden, lowly expressed genes, defined as having <10 counts in 40% of samples (i.e., the approximate proportion of cases), were excluded, leaving 20,855 genes in our analysis.

Variable Enhancer Loci (VEL) mapping approach

VELs were identified as previously described (12). Briefly, each opioid case was compared independently to the full panel of controls. Shannon entropy scoring was performed on inherent normalized H3K27ac ChIP-seq data (described above) to quantify peak specificity for each opioid case. Case/control status was then randomly permuted and Shannon entropy scoring was repeated. These permutations were performed at least 70,000 times to generate a unique null expectation of peak specificity scores for each individual peak region. These

permutations reveal the expected distribution of specificity scores due to inter-sample variation at each peak region. The true value was compared to these distributions for each peak to define the p-value. As each peak was tested twice to identify cases with significantly higher signal than controls (gained VELs) and significantly lower (lost) VELs, p-values were first multiplied by 2. We then applied Benjamini-Hochberg FDR correction to identify lost and gained VELs at an FDR 0.05%.

Machine learning models

We utilized inherent normalized (see above) data for all 652,055 H3K27ac peaks, as well as a list of covariates as input features. The covariates include: sample descriptors such as, sex, age, ancestry and PMI; and ChIP-seq data descriptors such as number of peaks per ChIP, number sequencing reads generated, number of aligned reads after duplicate removal, alignment rate, fraction reads in peaks, and batch number. We selected Gradient Boosting models based on overall performance on tabular data and interpretability via SHAP method. Additionally, we aimed to evaluate the cohort with a non-linear approach. Gradient boosting models are well suited to capturing complex interactions among features in datasets of small sample size. We used XGBoost's (21) default hyper-parameters to generate models to predict case control status of ChIP samples and only varied the input feature list in order to compare the models.

Step-wise strategy for peak set removal.—We utilized a step-wise strategy to identify distinct models capable of distinguishing cases and controls at high accuracy. The first model receives all input features (see definition above). For each subsequent model, the features utilized by the previous models are removed from the input feature list. We repeated this strategy until we identified a model where the AUC median (described below) fell below 0.9, resulting in a total of 108 models. To find peak sets that are predictive and robust to the choice of train/test set, we utilized two stage assessment of new models: 1) peak set discovery and 2) peak set evaluation.

Peak set discovery.—We split the data into five folds, with 4 folds of 10 samples and 1 fold of 11 samples. We train 5 models, each using a different combination of four folds as training set (80% of samples) and the remaining fold as test set (20% of samples). By training 5 models, each fold is used as test set exactly once. Early stopping was used during training to obtain the minimal number of features that achieves good performance. After training, we examined each model to extract information about which features the model deemed most relevant to the classification task. We used the SHAP package to calculate SHAP values for each feature per sample (66). In order to find features that are important in distinguishing cases from controls, we used the mean of absolute SHAP values as the overall importance of each feature across all samples (a non-negative value). We define the “peak set” of a model as peaks that had non-zero overall importance.

Peak set evaluation.—The predictive performance of a peak set in stage 1 might be specific to how the samples are split into folds. If there is heterogeneity in the data, the specific samples in the test fold can influence model performance. Thus, to obtain unbiased metric for model performance, we evaluated the peak sets on multiple different 5-fold splits.

For each of the 5 peak set discovered in step 1, we trained 1000 5-fold cross-validated models (5000 models in total per peak set). To select the final peak set, we examine the AUC distribution across the aforementioned models. The chosen peak set for downstream analyses is the one that ranks highest when sorted by decreasing AUC mean and AUC minimum.

Convergence analysis

Interactions between gene promoters and non-promoter loci are intersected with merged, non-overlapping elements called from the 51 case and 51 control H3K27ac tracks (652,055 elements).

Convergence analysis is run on two binary matrices of gained and lost variant H3K27ac elements (VEL) identified on 51 opioid-overdose individuals using our entropy-based approach using the Axiotl Convergence Platform (convergence.axiotl.com). The analysis consists of a permutation test in which a convergence statistic in the observed plexus is compared to a null distribution obtained from a collection of null plexi sampled from the binary matrix so as to match the activity and connectivity properties of the original plexus. Heterogeneity in global patient activity is also preserved. The algorithm is described in detail in Sallari et al. (manuscript in preparation).

LD score regression

LDSR was used to estimate heritability enrichment across significant plexi. Annotated SNPs lists were generated for all SNPs within VEL regions (± 5 kb from VEL peak centers) associated with the top 5 significant plexi. European LD scores and weights were downloaded from the LDSC (67). Summary statistics for 27 traits were obtained from previously published studies listed in Supplemental Table S6. LD scores were calculated as recommended <https://github.com/bulik/ldsc/wiki/> using HapMap3 SNPs and 1000 genomes phase 3 reference files (68). Heritability enrichment calculations were performed using the recommended ‘baseline model’, composed of multiple functional categories including conserved regions and coding regions. The significance threshold was defined through Bonferroni multi-test correction for 27 traits tested.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to express our gratitude to the families of the subjects used in this study. We would also like to acknowledge the teams of technicians at the Miami Brain Endowment Bank. This work was supported by NIH National Institute of Drug Abuse R01 DA043980.

References

1. Center for Disease Control. Increase in fatal drug overdoses across the United States driven by synthetic opioids before and during the COVID-19 pandemic. Publication Number CDCHAN-00438 December Available at <https://emergencycdc.gov/han/2020/han00438asp> (Accessed February 14, 2021). 2020;

2. Zhou H, Rentsch CT, Cheng Z, Kember RL, Nunez YZ, Sherva RM, et al. Association of OPRM1 Functional Coding Variant With Opioid Use Disorder: A Genome-Wide Association Study. *JAMA Psychiatry* [Internet]. 2020 Jun 3; Available from: 10.1001/jamapsychiatry.2020.1206
3. Hancock DB, Markunas CA, Bierut LJ, Johnson EO. Human Genetics of Addiction: New Insights and Future Directions. *Curr Psychiatry Rep*. 2018 Mar 5;20(2):8. [PubMed: 29504045]
4. Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Børglum AD, Breen G, et al. Psychiatric Genomics: An Update and an Agenda. *AJP*. 2018 Jan 1;175(1):15–27.
5. Koob GF, Volkow ND. Neurocircuitry of addiction. *Neuropsychopharmacology*. 2010 Jan;35(1):217–38. [PubMed: 19710631]
6. Browne CJ, Godino A, Sallery M, Nestler EJ. Epigenetic Mechanisms of Opioid Addiction. *Biol Psychiatry*. 2020 Jan 1;87(1):22–33. [PubMed: 31477236]
7. Goldstein RZ, Volkow ND. Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. *Nat Rev Neurosci*. 2011 Oct 20;12(11):652–69. [PubMed: 22011681]
8. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA [Internet]. Vol. 337, *Science*. 2012. p. 1190–5. Available from: 10.1126/science.1222794 [PubMed: 22955828]
9. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet*. 2013 Feb;45(2):124–30. [PubMed: 23263488]
10. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh P-R, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015 Nov;47(11):1228–35. [PubMed: 26414678]
11. Jung I, Schmitt A, Diao Y, Lee AJ, Liu T, Yang D, et al. A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat Genet*. 2019 Oct;51(10):1442–9. [PubMed: 31501517]
12. Akhtar-Zaidi B, Cowper-Sal-lari R, Corradin O, Saiakhova A, Bartels CF, Balasubramanian D, et al. Epigenomic enhancer profiling defines a signature of colon cancer. *Science*. 2012 May 11;336(6082):736–9. [PubMed: 22499810]
13. Cohen AJ, Saiakhova A, Corradin O, Luppino JM, Lovrenert K, Bartels CF, et al. Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome. *Nat Commun*. 2017;8(1):1–13. [PubMed: 28232747]
14. Morrow JJ, Bayles I, Funnell APW, Miller TE, Saiakhova A, Lizardo MM, et al. Positively selected enhancer elements endow osteosarcoma cells with metastatic competence. *Nat Med*. 2018 Feb;24(2):176–85. [PubMed: 29334376]
15. Blackwood CA, McCoy MT, Ladenheim B, Cadet JL. Oxycodone self-administration activates the mitogen-activated protein kinase/ mitogen- and stress-activated protein kinase (MAPK-MSK) signaling pathway in the rat dorsal striatum. *Sci Rep*. 2021 Jan 28;11(1):2567. [PubMed: 33510349]
16. Qiao X, Zhu Y, Dang W, Wang R, Sun M, Chen Y, et al. Dual-specificity phosphatase 15 (DUSP15) in the nucleus accumbens is a novel negative regulator of morphine-associated contextual memory. *Addict Biol*. 2021;26(1):e12884. [PubMed: 32043707]
17. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016 Jul 8;44(W1):W90–7. [PubMed: 27141961]
18. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013 Apr 15;14:128. [PubMed: 23586463]
19. Al-Hasani R, Bruchas MR. Molecular mechanisms of opioid receptor-dependent signaling and behavior. *Anesthesiology*. 2011 Dec;115(6):1363–81. [PubMed: 22020140]
20. Brynildsen JK, Mace KD, Cornblath EJ, Weidler C, Pasqualetti F, Bassett DS, et al. Gene coexpression patterns predict opiate-induced brain-state transitions. *Proc Natl Acad Sci U S A*. 2020 Aug 11;117(32):19556–65. [PubMed: 32694207]

21. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, NY, USA: Association for Computing Machinery; 2016. p. 785–94. (KDD '16).
22. McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction [Internet]. arXiv [stat.ML]. 2018. Available from: <http://arxiv.org/abs/1802.03426>
23. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017 Nov 28;8(1):1826. [PubMed: 29184056]
24. Sallari RC, Sinnott-Armstrong NA, French JD, Kron KJ, Ho J, Moore JH, et al. Convergence of dispersed regulatory mutations predicts driver genes in prostate cancer. *bioRxiv* [Internet]. 2017 Aug 26; Available from: 10.1101/097451
25. Bailey SD, Desai K, Kron KJ, Mazrooei P, Sinnott-Armstrong NA, Treloar AE, et al. Noncoding somatic and inherited single-nucleotide variants converge to promote ESR1 expression in breast cancer. *Nat Genet*. 2016 Oct;48(10):1260–6. [PubMed: 27571262]
26. Lutz P-E, Almeida D, Filliol D, Jollant F, Kieffer BL, Turecki G. Increased functional coupling of the mu opioid receptor in the anterior insula of depressed individuals. *Neuropsychopharmacology*. 2021 Apr;46(5):920–7. [PubMed: 33531622]
27. Mendez EF, Wei H, Hu R, Stertz L, Fries GR, Wu X, et al. Angiogenic gene networks are dysregulated in opioid use disorder: evidence from multi-omics and imaging of postmortem human brain. *Mol Psychiatry* [Internet]. 2021 Aug 12; Available from: 10.1038/s41380-021-01259-y
28. Martins SS, Fenton MC, Keyes KM, Blanco C, Zhu H, Storr CL. Mood and anxiety disorders and their association with non-medical prescription opioid use and prescription opioid-use disorder: longitudinal evidence from the National Epidemiologic Study on Alcohol and Related Conditions [Internet]. Vol. 42, *Psychological Medicine*. 2012. p. 1261–72. Available from: 10.1017/s0033291711002145 [PubMed: 21999943]
29. Polimanti R, Walters RK, Johnson EC, McClintick JN, Adkins AE, Adkins DE, et al. Leveraging genome-wide data to investigate differences between opioid use vs. opioid dependence in 41,176 individuals from the Psychiatric Genomics Consortium [Internet]. Available from: 10.1101/765065
30. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005 Oct 25;102(43):15545–50. [PubMed: 16199517]
31. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*. 2003 Jun 15;34(3):267–73. [PubMed: 12808457]
32. Bettinger JC, Davies AG. The role of the BK channel in ethanol response behaviors: evidence from model organism and human studies. *Front Physiol*. 2014 Sep 9;5:346. [PubMed: 25249984]
33. Wolen AR, Phillips CA, Langston MA, Putman AH, Vorster PJ, Bruce NA, et al. Genetic dissection of acute ethanol responsive gene networks in prefrontal cortex: functional and mechanistic implications. *PLoS One*. 2012 Apr 12;7(4):e33575. [PubMed: 22511924]
34. Abdul Rahman NZ, Greenwood SM, Brett RR, Tossell K, Ungless MA, Plevin R, et al. Mitogen-Activated Protein Kinase Phosphatase-2 Deletion Impairs Synaptic Plasticity and Hippocampal-Dependent Memory. *J Neurosci*. 2016 Feb 24;36(8):2348–54. [PubMed: 26911683]
35. Cui W-Y, Seneviratne C, Gu J, Li MD. Genetics of GABAergic signaling in nicotine and alcohol dependence. *Hum Genet*. 2012 Jun;131(6):843–55. [PubMed: 22048727]
36. Ribeiro AF, Correia D, Torres AA, Boas GRV, Rueda AVL, Camarini R, et al. A transcriptional study in mice with different ethanol-drinking profiles: possible involvement of the GABA(B) receptor. *Pharmacol Biochem Behav*. 2012 Aug;102(2):224–32. [PubMed: 22579910]
37. Behesti H, Fore TR, Wu P, Horn Z, Leppert M, Hull C, et al. ASTN2 modulates synaptic strength by trafficking and degradation of surface proteins. *Proc Natl Acad Sci U S A*. 2018 Oct 9;115(41):E9717–26. [PubMed: 30242134]
38. Inoue R, Nishizawa D, Hasegawa J, Nakayama K, Fukuda K-I, Ichinohe T, et al. Effects of rs958804 and rs7858836 single-nucleotide polymorphisms of the ASTN2 gene on pain-related

- phenotypes in patients who underwent laparoscopic colectomy and mandibular sagittal split ramus osteotomy. *Neuropsychopharmacol Rep*. 2021 Mar;41(1):82–90. [PubMed: 33476460]
39. Hillis SD, Anda RF, Felitti VJ, Marchbanks PA. Adverse childhood experiences and sexual risk behaviors in women: a retrospective cohort study. *Fam Plann Perspect*. 2001 Sep;33(5):206–11. [PubMed: 11589541]
 40. Montez JK, Hayward MD. Cumulative childhood adversity, educational attainment, and active life expectancy among U.S. adults. *Demography*. 2014 Apr;51(2):413–35. [PubMed: 24281740]
 41. Murthy S, Gould E. How Early Life Adversity Influences Defensive Circuitry. *Trends Neurosci*. 2020 Apr;43(4):200–12. [PubMed: 32209452]
 42. Levis SC, Bentzley BS, Molet J, Bolton JL, Perrone CR, Baram TZ, et al. On the early life origins of vulnerability to opioid addiction. *Mol Psychiatry* [Internet]. 2019 Dec 10; Available from: 10.1038/s41380-019-0628-5
 43. Enoch M-A. The role of early life stress as a predictor for alcohol and drug dependence. *Psychopharmacology*. 2011 Mar;214(1):17–31. [PubMed: 20596857]
 44. Widom CS, Marmorstein NR, White HR. Childhood victimization and illicit drug use in middle adulthood. *Psychol Addict Behav*. 2006 Dec;20(4):394–403. [PubMed: 17176174]
 45. Vaiserman AM, Koliada AK. Early-life adversity and long-term neurobehavioral outcomes: epigenome as a bridge? *Hum Genomics*. 2017 Dec 16;11(1):34. [PubMed: 29246185]
 46. Barnett Burns S, Almeida D, Turecki G. The Epigenetics of Early Life Adversity: Current Limitations and Possible Solutions. In: Grayson DR, editor. *Progress in Molecular Biology and Translational Science*. Academic Press; 2018. p. 343–425.
 47. Leza L, Siria S, López-Goñi JJ, Fernández-Montalvo J. Adverse childhood experiences (ACEs) and substance use disorder (SUD): A scoping review. *Drug Alcohol Depend*. 2021 Jan 29;221:108563. [PubMed: 33561668]
 48. Dasgupta N, Beletsky L, Ciccarone D. Opioid Crisis: No Easy Fix to Its Social and Economic Determinants. *Am J Public Health*. 2018 Feb;108(2):182–6. [PubMed: 29267060]
 49. Altekruze SF, Cosgrove CM, Altekruze WC, Jenkins RA, Blanco C. Socioeconomic risk factors for fatal opioid overdoses in the United States: Findings from the Mortality Disparities in American Communities Study (MDAC). *PLoS One*. 2020 Jan 17;15(1):e0227966. [PubMed: 31951640]
 50. Case A, Deaton A, Stone AA. Decoding the mystery of American pain reveals a warning for the future. *Proc Natl Acad Sci U S A*. 2020 Oct 6;117(40):24785–9. [PubMed: 32958666]
 51. Martin M Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. 2011 May 2;17(1):10–2.
 52. Li H Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM [Internet]. arXiv [q-bio.GN]. 2013. Available from: <http://arxiv.org/abs/1303.3997>
 53. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009 Aug 15;25(16):2078–9. [PubMed: 19505943]
 54. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008 Sep 17;9(9):R137. [PubMed: 18798982]
 55. Ramírez F, Ryan DP, Grünig B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016 Jul 8;44(W1):W160–5. [PubMed: 27079975]
 56. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013 Mar;14(2):178–92. [PubMed: 22517427]
 57. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004 Sep 15;5(10):R80. [PubMed: 15461798]
 58. Carroll TS, Liang Z, Salama R, Stark R, de Santiago I. Impact of artifact removal on ChIP quality metrics in ChIP-seq and ChIP-exo data. *Front Genet*. 2014 Apr 10;5:75. [PubMed: 24782889]
 59. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014 Aug 1;30(15):2114–20. [PubMed: 24695404]

60. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017 Apr;14(4):417–9. [PubMed: 28263959]
61. Srivastava A, Malik L, Sarkar H, Zakeri M, Almodaresi F, Sonesson C, et al. Alignment and mapping methodology influence transcript abundance estimation. *Genome Biol*. 2020 Sep 7;21(1):239. [PubMed: 32894187]
62. Love MI, Hogenesch JB, Irizarry RA. Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. *Nat Biotechnol*. 2016 Dec;34(12):1287–91. [PubMed: 27669167]
63. Sonesson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res* [Internet]. 2015;4. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/pmc4712774/>
64. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. [PubMed: 25516281]
65. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012 Mar 15;28(6):882–3. [PubMed: 22257669]
66. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat Mach Intell*. 2020 Jan;2(1):56–67. [PubMed: 32607472]
67. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015 Mar;47(3):291–5. [PubMed: 25642630]
68. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015 Oct 1;526(7571):68–74. [PubMed: 26432245]
69. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet*. 2019 Jan;51(1):63–75. [PubMed: 30478444]
70. Zhou H, Sealock JM, Sanchez-Roige S, Clarke T-K, Levey DF, Cheng Z, et al. Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. *Nat Neurosci*. 2020 Jul;23(7):809–18. [PubMed: 32451486]
71. Boraska V, Franklin CS, Floyd JAB, Thornton LM, Huckins LM, Southam L, et al. A genome-wide association study of anorexia nervosa. *Mol Psychiatry*. 2014 Oct;19(10):1085–94. [PubMed: 24514567]
72. Stahl EA, Breen G, Forstner AJ, McQuillin A, Ripke S, Trubetskoy V, et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat Genet*. 2019 May;51(5):793–803. [PubMed: 31043756]
73. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*. 2010 Nov;42(11):937–48. [PubMed: 20935630]
74. Dubois PCA, Trynka G, Franke L, Hunt KA, Romanos J, Curtotti A, et al. Multiple common variants for celiac disease influencing immune gene expression. *Nat Genet*. 2010 Apr;42(4):295–302. [PubMed: 20190752]
75. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012 Oct 31;491(7422):119–24. [PubMed: 23128233]
76. Howard DM, Adams MJ, Clarke T-K, Hafferty JD, Gibson J, Shiralil M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat Neurosci*. 2019 Mar;22(3):343–52. [PubMed: 30718901]
77. Manning AK, Hivert M-F, Scott RA, Grimsby JL, Bouatia-Naji N, Chen H, et al. A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat Genet*. 2012 May 13;44(6):659–69. [PubMed: 22581228]

78. Levey DF, Gelernter J, Polimanti R, Zhou H, Cheng Z, Aslan M, et al. Reproducible Genetic Risk Loci for Anxiety: Results From ~200,000 Participants in the Million Veteran Program. *AJP*. 2020 Mar 1;177(3):223–32.
79. Allen HL, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. 2010 Sep 29;467(7317):832–8. [PubMed: 20881960]
80. Okbay A, Baselmans BML, De Neve J-E, Turley P, Nivard MG, Fontana MA, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet*. 2016 Apr 18;48(6):624–33. [PubMed: 27089181]
81. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014 Jul 24;511(7510):421–7. [PubMed: 25056061]
82. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet*. 2019 Feb;51(2):237–44. [PubMed: 30643251]
83. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature*. 2010 Aug;466(7307):707–13. [PubMed: 20686565]
84. Bradfield JP, Qu H-Q, Wang K, Zhang H, Sleiman PM, Kim CE, et al. A Genome-Wide Meta-Analysis of Six Type 1 Diabetes Cohorts Identifies Multiple Associated Loci. *PLoS Genet*. 2011 Sep 29;7(9):e1002293. [PubMed: 21980299]
85. Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, et al. GWAS of 126,559 Individuals Identifies Genetic Variants Associated with Educational Attainment. *Science*. 2013 Jun 21;340(6139):1467–71. [PubMed: 23722424]
86. Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*. 2016 May 11;533(7604):539–42. [PubMed: 27225129]
87. Linnér RK, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat Genet*. 2019 Jan 14;51(2):245–57 [PubMed: 30643258]

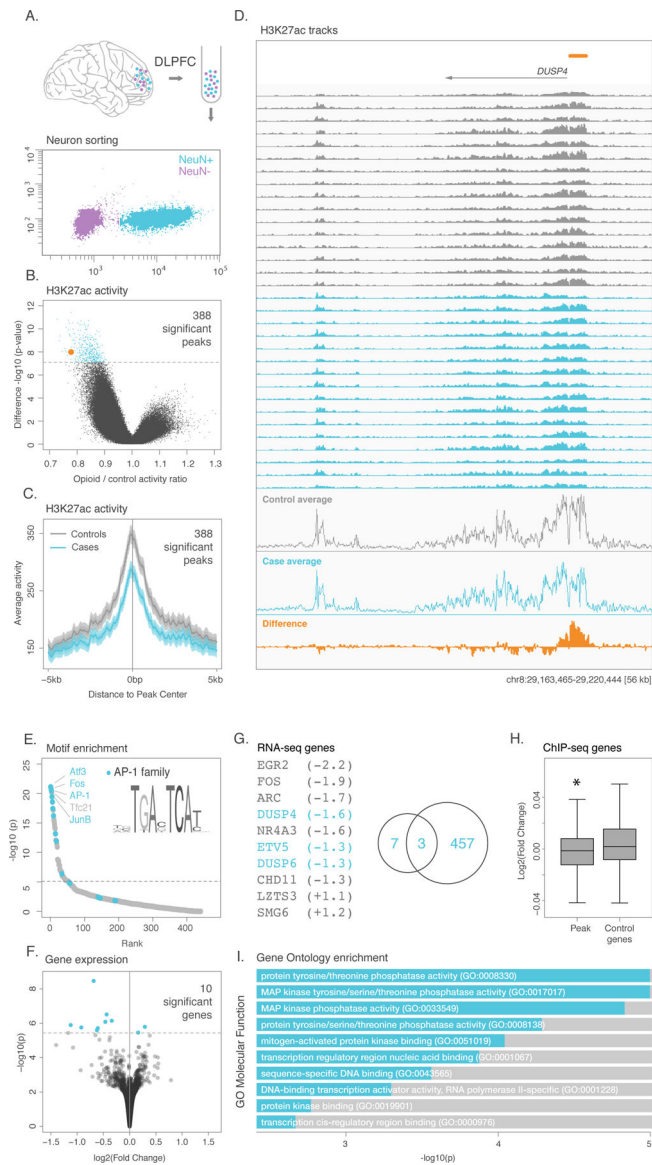


Figure 1. Identification of homogeneous H3K27ac hypoacetylation events

(A) Schematic overview of experimental strategy. NeuN⁺ nuclei were identified using FACS from post mortem dorsolateral prefrontal cortex tissues, followed by H3K27ac ChIP-seq. (B) Volcano plot of hypoacetylation analysis of ChIP-seq samples. 388 differentially acetylated peaks identified at genome-wide significance are shown in blue. Example locus *DUSP4* highlighted in orange corresponding orange bar in D. (C) Aggregate H3K27ac ChIP-seq signal across controls (grey) and opioid cases (blue) for 388 differentially acetylated regions. (D) Browser image of H3K27ac ChIP-seq results at example locus *DUSP4*. Orange bar highlights differentially acetylated peak. (E) Motifs enriched in differentially acetylated peaks ranked by enrichment p-value. (F) Volcano plot of differentially expressed genes identified through RNA-seq of bulk tissue samples (n=24 cases and n=27 controls). Bonferroni multi-test correction threshold is shown in dotted line. (G) Comparison of differential peaks and genes. (left) Differentially expressed genes with

\log_2 fold change (cases/controls) shown in parentheses. Genes in blue overlap gene targets of hypoacetylated peaks. (right) Venn diagram showing 10 differentially expressed genes (left circle) compared to 460 genes linked to hypoacetylated peaks (right circle). **(H)** Tukey boxplot of \log_2 fold-change (cases/controls) in gene expression of gene targets associated with hypoacetylated peaks versus randomly selected control genes not associated with a hypoacetylated peak. Welch's two sample t-test $p=0.00067$. Outliers not shown. **(I)** Gene ontology enrichment of differentially expressed genes.

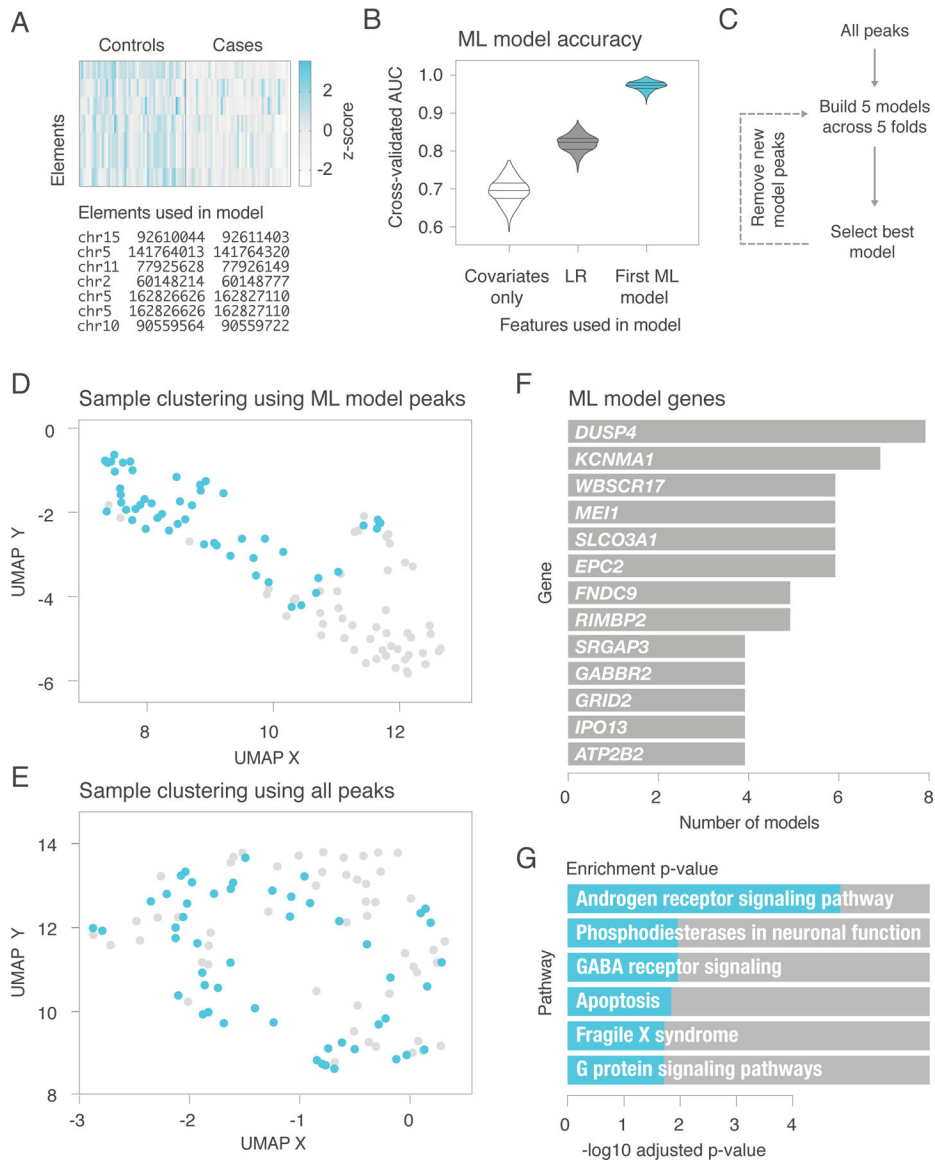


Figure 2. Machine learning reveals models and key features that distinguish opioid cases from controls

(A) Heatmap of the seven peaks that distinguish cases for controls in the first ML model. (B) Distribution of AUC derived from multiple 80/20 train/test splits (5-fold cross validation with 1,000 repetitions) shown for the top machine learning model (blue), models derived from the 388 hypoacetylated peaks identified (grey) and models derived only using covariates (white). (C) Schematic representing iterative strategy to identify additional models with high predictive accuracy. (D) UMAP projection of opioid cases (blue) and control samples (grey) based on peaks utilized by 108 machine learning models and (E) based on all ChIP-seq peaks. (F) Gene targets that are shared across 4 or more ML models. (G) Pathways enriched across all ML model genes.

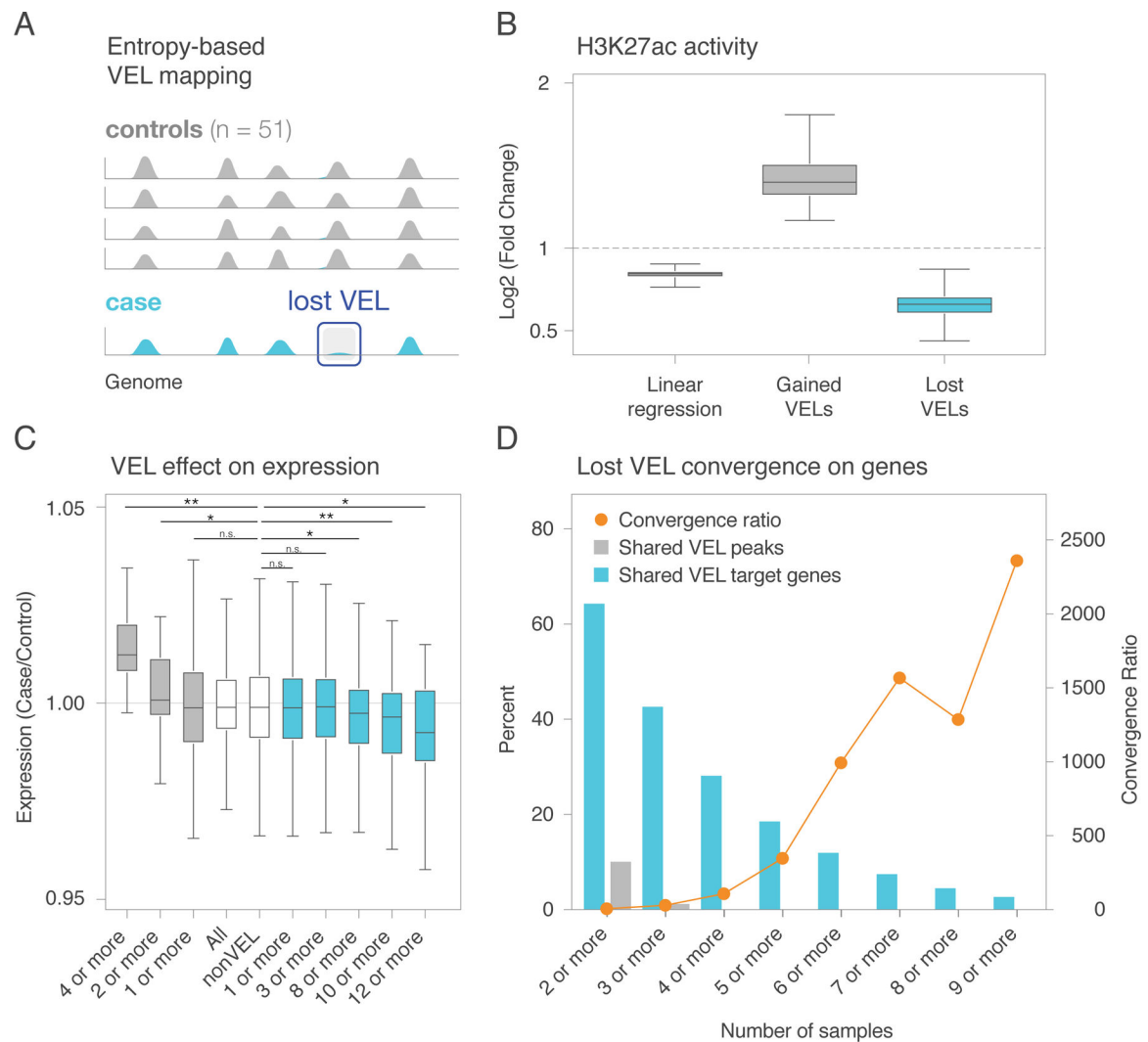


Figure 3. Identification of Variable Enhancer Loci (VELs)

(A) Schematic overview of VEL identification. Each opioid case is compared to the full set of control samples to identify case-specific alterations in H3K27ac. (B) Log₂ fold change (cases/controls) in ChIP-seq signal for lost VELs (blue), gained VELs (grey), and hypoacetylated peaks (black). (C) Tukey boxplot of fold change in average gene expression for genes associated with VELs. Genes associated with gained VELs (grey) and lost VELs (blue) in multiple samples are shown as well as control gene lists shown in white. Outliers not shown. Kruskal-wallis ANOVA $P < 0.0001$. Dunn's post-hoc test is shown * $P < 0.05$ ** $P < 0.009$. (D) Comparison of VELs and VEL gene targets across opioid cases. Percentage of VEL peaks shared across multiple cases (grey bar). Percentage of VEL gene targets shared across multiple samples (blue bar). Convergence ratio, i.e. the ratio of shared genes to shared VEL peaks (orange line).

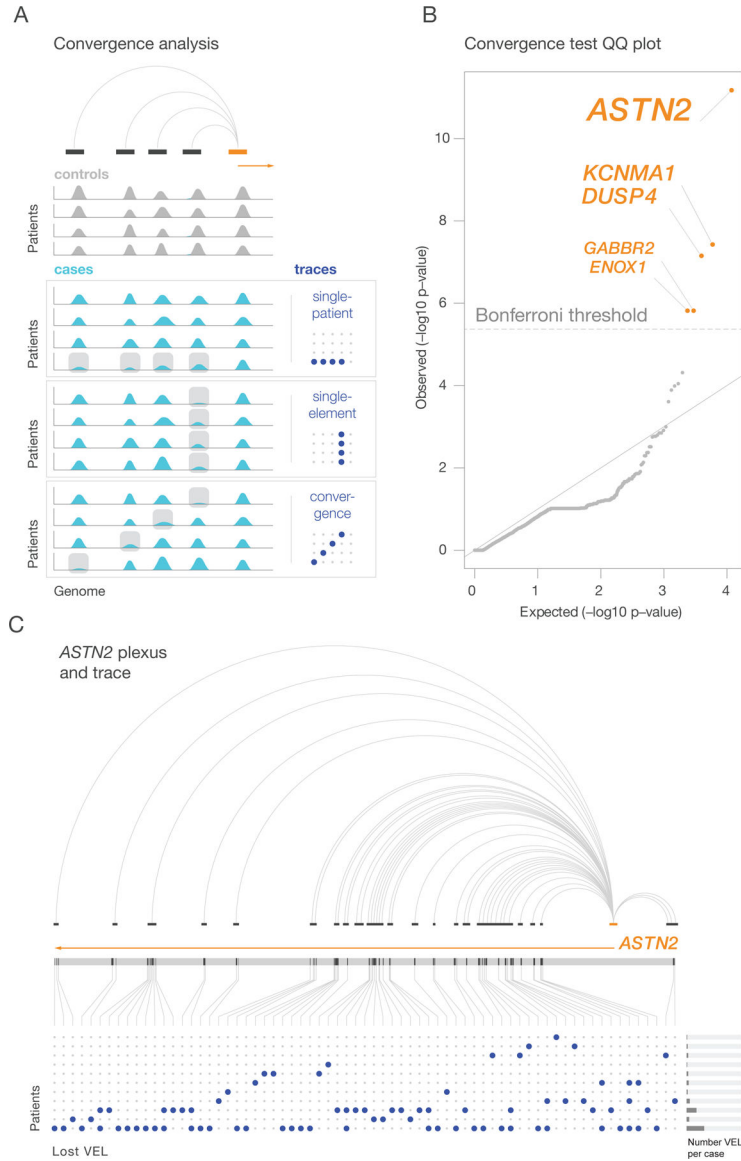


Figure 4. VEL convergence of shared target genes with plexus analysis
(A) Schematic representing plexus convergence strategy. (top) H3K27ac peaks (black) associated with common target promoter (orange) in promoter capture Hi-C (arcs). Example traces are shown for multiple VELs associated with the same target gene in one opioid case (single patient), multiple samples with the same VEL (single element), different VELs associated with the same target gene across cases (convergence). **(B)** Quantile-quantile plot of plexus analysis of VEL convergence. Five genome-wide significant genes are labelled (font proportional to p-value). **(C)** Plexus VEL trace for top gene *ASTN2* is shown with genomic location and Hi-C interaction indicated above. Bar graphs indicating the total number of VELs identified per case is shown (right).

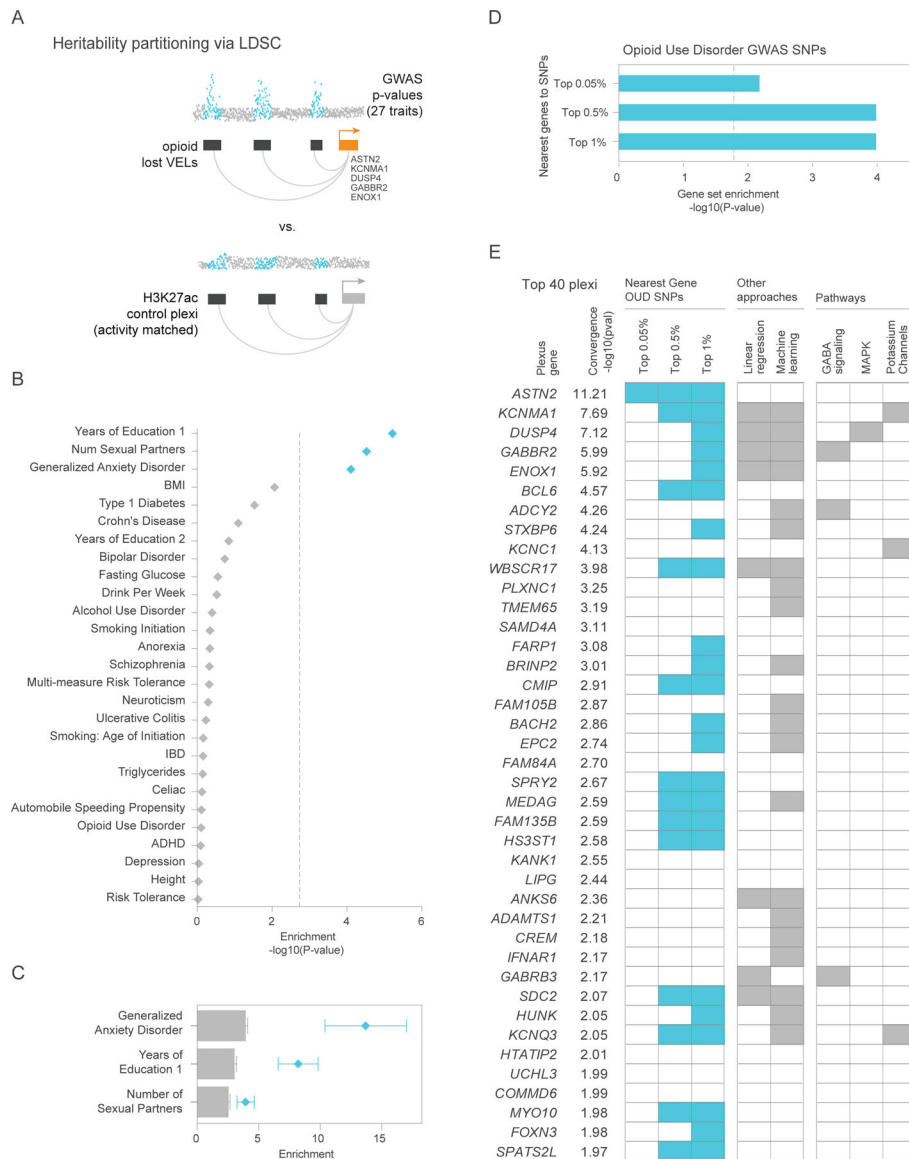


Figure 5. Convergent VEL genes are enriched for disease heritability

(A) Schematic representing analytical approach using LDSC to compare published GWAS results to top 5 plexus regions in (B) and to control regions with matched neuronal H3K27ac activity (C). (B) LDSC heritability enrichment ($-\log_{10}$ p-value) for VEL regions associated with top 5 plexus genes (Fig. 4B). Dotted line indicates Bonferroni multi-test correction threshold. Traits with significant enrichment are shown in blue. (C) Heritability enrichment (proportion of heritability / proportion of SNPs in peak set) with standard error is shown for VEL regions (blue dots) and 10 control sets of randomly selected neuronal peaks with activity profiles matched to VELs (grey bars). (D) GSEA analysis comparing genes ranked by plexus p-value to OUD candidate gene sets defined as genes nearest top 1%, 0.5% and 0.05% of SNPs identified by OUD GWAS (2). (E) Heatmap representing top 40 genes defined by plexus analysis (left). Genes nearest the top 1%, 0.5% and 0.05% of OUD GWAS SNPs are indicated (blue). P-value based on 10,000 permutations. Genes also identified by

previous approaches including linear regression, T tests and machine learning models are highlighted in center. Genes linked to key pathways are indicated (right).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript