OMB Number: 4040-0010
Expiration Date: 12/31/2022

APPLICATION FOR FEDERAL ASSISTANCE
SF 424 (R&R)

3. DATE RECEIVED BY STATE	**State Application Identifier** PA

1. TYPE OF SUBMISSION*

❍ Pre-app cat on ● App cat on ❍ Changed/Corrected App cat on

4.a. Federal Identifier DA053020

b. Agency Routing Number

2. DATE SUBMITTED | **Application Identifier**

c. Previous Grants.gov Tracking Number GRANT13066849

5. APPLICANT INFORMATION **Organizational DUNS*:** 0521841160000

Legal Name*: Carnegie Mellon University
Department:
Division:
Street1*: 5000 Forbes Ave
Street2:
City*: Pittsburgh
County: PA
State*: PA: Pennsylvania
Province:
Country*: USA: UNITED STATES
ZIP / Postal Code*: 152133890

Person to be contacted on matters involving this application
Prefix: Ms. First Name*: Rebecca Middle Name: Last Name*: Harrold Suffix:
Position/Title: Research Administrator
Street1*: 5000 Forbes Ave
Street2:
City*: Pittsburgh
County:
State*: PA: Pennsylvania
Province:
Country*: USA: UNITED STATES
ZIP / Postal Code*: 152133890
Phone Number*: 4122682813 Fax Number: 4122686279 Email: osp-preaward@andrew.cmu.edu

6. EMPLOYER IDENTIFICATION NUMBER *(EIN) or (TIN)** 1250969449A1

7. TYPE OF APPLICANT* O: Private Institution of Higher Education

Other (Specify):
Small Business Organization Type ❍ Women Owned ❍ Socially and Economically Disadvantaged

8. TYPE OF APPLICATION*

❍ New ● Resubmission
❍ Renewal ❍ Continuation ❍ Revision

If Revision, mark appropriate box(es).
❍ A. Increase Award ❍ B. Decrease Award ❍ C. Increase Duration
❍ D. Decrease Duration ❍ E. Other *(specify)* :

Is this application being submitted to other agencies?* ❍Yes ●No What other Agencies?

9. NAME OF FEDERAL AGENCY*
National Institutes of Health

10. CATALOG OF FEDERAL DOMESTIC ASSISTANCE NUMBER
TITLE:

11. DESCRIPTIVE TITLE OF APPLICANT'S PROJECT*
Integrating rodent-primate cell types and epigenomics to identify conservation in substance addiction

12. PROPOSED PROJECT
Start Date* Ending Date*
07/01/2021 06/30/2025

13. CONGRESSIONAL DISTRICTS OF APPLICANT
PA-018

SF 424 (R&R) APPLICATION FOR FEDERAL ASSISTANCE

14. PROJECT DIRECTOR/PRINCIPAL INVESTIGATOR CONTACT INFORMATION

Prefix: First Name*: BaDoi Middle Name: Nguyen Last Name*: Phan Suffix:
Position/Title: Medical Student, MD/PhD Training Program
Organization Name*: Carnegie Mellon University
Department: Computational Biology
Division:
Street1*: 5000 Forbes Ave
Street2:
City*: Pittsburgh
County: PA
State*: PA: Pennsylvania
Province:
Country*: USA: UNITED STATES
ZIP / Postal Code*: 152133890
Phone Number*: 4086135493 Fax Number: Email*: badoi.phan@pitt.edu

15. ESTIMATED PROJECT FUNDING

a. Total Federal Funds Requested*	$364,495.00
b. Total Non-Federal Funds*	$0.00
c. Total Federal & Non-Federal Funds*	$364,495.00
d. Estimated Program Income*	$0.00

16.IS APPLICATION SUBJECT TO REVIEW BY STATE EXECUTIVE ORDER 12372 PROCESS?*

a. YES ○ THIS PREAPPLICATION/APPLICATION WAS MADE AVAILABLE TO THE STATE EXECUTIVE ORDER 12372 PROCESS FOR REVIEW ON:
DATE:

b. NO ● PROGRAM IS NOT COVERED BY E.O. 12372; OR

○ PROGRAM HAS NOT BEEN SELECTED BY STATE FOR REVIEW

17. By signing this application, I certify (1) to the statements contained in the list of certifications* and (2) that the statements herein are true, complete and accurate to the best of my knowledge. I also provide the required assurances * and agree to comply with any resulting terms if I accept an award. I am aware that any false, fictitious, or fraudulent statements or claims may subject me to criminal, civil, or administrative penalties. (U.S. Code, Title 18, Section 1001)

● I agree*

** The list of certifications and assurances or an Internet site where you may obtain this list is contained in the announcement or agency specific instructions*

18. SFLLL or OTHER EXPLANATORY DOCUMENTATION File Name:

19. AUTHORIZED REPRESENTATIVE

Prefix: First Name*: Christine Middle Name: Last Name*: Bediliion Suffix:
Position/Title*: Executive Director
Organization Name*: Carnegie Mellon University
Department: Office of Sponsored Research
Division: Carnegie Mellon University
Street1*: 5000 Forbes Ave
Street2:
City*: Pittsburgh
County: PA
State*: PA: Pennsylvania
Province:
Country*: USA: UNITED STATES
ZIP / Postal Code*: 152133890
Phone Number*: 4122688746 Fax Number: 412-268-6279 Email*: osp-preaward@andrew.cmu.edu

Signature of Authorized Representative*
Post Populate Submitter Signature

Date Signed*
12/07/2020

20. PRE-APPLICATION File Name:

21. COVER LETTER ATTACHMENT File Name:Submission_Letter_resub_final.pdf

424 R&R and PHS-398 Specific
Table Of Contents

OMB Number 4040-0010
Expiration Date 12/31/2022

Project/Performance Site Location(s)

Project/Performance Site Primary Location

❍ I am submitting an application as an individual, and not on behalf of a company, state, local or tribal government, academia, or other type of organization.

Organization Name:	Carnegie Mellon University
Duns Number:	0521841160000
Street1*:	5000 Forbes Ave
Street2:	
City*:	Pittsburgh
County:	PA
State*:	PA: Pennsylvania
Province:	
Country*:	USA: UNITED STATES
Zip / Postal Code*:	152133890
Project/Performance Site Congressional District*:	PA-018

Additional Location(s) File Name:

Track ng Number:

Fund ng Opportun ty Number: . Rece ved Date:

OMB Number: 4040-0010
Expiration Date: 12/31/2022

RESEARCH & RELATED Other Project Information

1. Are Human Subjects Involved?* ❍ **Yes** ● **No**

1.a. If YES to Human Subjects

Is the Project Exempt from Federal regulations? ❍ Yes ❍ No

If YES, check appropriate exemption number: _ 1 _ 2 _ 3 _ 4 _ 5 _ 6 _ 7 _ 8

If NO, is the IRB review Pending? ❍ **Yes** ❍ **No**

IRB Approval Date:

Human Subject Assurance Number

2. Are Vertebrate Animals Used?* ● **Yes** ❍ **No**

2.a. If YES to Vertebrate Animals

Is the IACUC review Pending? ❍ Yes ● No

IACUC Approval Date: 09-18-2020

Animal Welfare Assurance Number A3352-01

3. Is proprietary/privileged information included in the application?* ❍ **Yes** ● **No**

4.a. Does this project have an actual or potential impact - positive or negative - on the environment?* ❍ **Yes** ● **No**

4.b. If yes, please explain:

4.c. If this project has an actual or potential impact on the environment, has an exemption been authorized or an environmental assessment (EA) or environmental impact statement (EIS) been performed? ❍ Yes ❍ No

4.d. If yes, please explain:

5. Is the research performance site designated, or eligible to be designated, as a historic place?* ❍ **Yes** ● **No**

5.a. If yes, please explain:

6. Does this project involve activities outside the United States or partnership with international collaborators?* ❍ **Yes** ● **No**

6.a. If yes, identify countries:

6.b. Optional Explanation:

	Filename
7. Project Summary/Abstract*	Abstract_resub_final.pdf
8. Project Narrative*	Project Narrative_resub.pdf
9. Bibliography & References Cited	References_resub_final.2.pdf
10.Facilities & Other Resources	Facilities-Other-Resources_resub_final.pdf
11.Equipment	Equipment_resub.pdf

Project Summary/Abstract

Substance use disorders (SUD) of many highly addictive drugs affect more than 100 million people worldwide. Genetic variations associated with complex neuro-behavioral traits, such as drug addiction, are likely to impact enhancers which have a high degree of cell type-specificity and can be conserved across species. Furthermore, variation in addiction behavior has been linked to genetic variation in humans. Thus, it follows that molecular mechanisms driving addiction behavior, specifically at cell type-specific cis-regulatory elements (CREs), might also be conserved between primates and rodents. Therefore, identifying which cells utilizing genes and CREs affected by human SUD risk variants can provide insight into the context-dependent molecular basis of SUD genetic risk.

This project proposes to identify the gene markers and CREs of cell types in the nucleus accumbens (NAc) that are conserved or clade-specific to primates and rodents and of these, which are enriched for SUD human genetic risk variants. The proposal comprises of the following aims: **Aim 1:** identify the conserved, active NAc marker genes and relation to human SUD genetic risk. **Aim 2:** identify the conserved, active NAc CREs and relation to human SUD genetic risk. Together, these experiments could reveal primate-rodent gene and CREs atlas of conserved and species-specific cell types of the reward system by integrating single-nuclei genomics data across multiple mammalian species. Understanding which NAc genes and CREs may be affected by human risk variants for SUD, particularly those measurably conserved in primate and rodents, will direct future studies into molecular mechanisms of SUD, revolutionize the playbook for translating human polygenic disease and disorders to model organisms, and ultimately progress towards curbing the substance addiction epidemic. Thus, this work integrates closely with my clinical interests in addiction medicine. This proposal outlines a combination of rigorous mentored research training, longitudinal clinical experiences, coursework, and professional and leadership development activities. The intellectual, technical, and professional skills refined during this training period will be instrumental in my development as an independent physician scientist.

Project Narrative

Substance use disorders affect more than 100 million people worldwide and are influenced by the genetic makeup of the individual. In this proposal, I will investigate how genetic differences in humans could influence sets of brain cell types shared across human, monkey, and rodents. These results will provide a foundation to link the power of animal models to specific features of SUDs in humans.

Bibliography

1. GBD 2016 Alcohol and Drug Use Collaborators. The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Psychiatry* **5,** 987–1012 (2018).
2. Ducci, F. & Goldman, D. The genetic basis of addictive disorders. *Psychiatr Clin North Am* **35,** 495–519 (2012).
3. Dick, D. M. The genetics of addiction: where do we go from here? *J Stud Alcohol Drugs* **77,** 673–675 (2016).
4. Waaktaar, T., Kan, K.-J. & Torgersen, S. The genetic and environmental architecture of substance use development from early adolescence into young adulthood: a longitudinal twin study of comorbidity of alcohol, tobacco and illicit drug use. *Addiction* **113,** 740–748 (2018).
5. Hällfors, J. *et al.* Genome-wide association study in Finnish twins highlights the connection between nicotine addiction and neurotrophin signaling pathway. *Addict Biol* **24,** 549–561 (2019).
6. Dickson, P. E. *et al.* Association of novelty-related behaviors and intravenous cocaine self-administration in Diversity Outbred mice. *Psychopharmacology* **232,** 1011–1024 (2015).
7. Saul, M. C. *et al.* Genetic variation and sex differences are missed opportunities for addiction biology. *BioRxiv* (2020). doi:10.1101/2020.02.14.949784
8. Flagel, S. B. *et al.* Genetic background and epigenetic modifications in the core of the nucleus accumbens predict addiction-like behavior in a rat model. *Proc. Natl. Acad. Sci. USA* **113,** E2861-70 (2016).
9. Dickson, P. E. *et al.* Systems genetics of intravenous cocaine self-administration in the BXD recombinant inbred mouse panel. *Psychopharmacology* **233,** 701–714 (2016).
10. Dickson, P. E. *et al.* Sex and strain influence attribution of incentive salience to reward cues in mice. *Behav. Brain Res.* **292,** 305–315 (2015).
11. Volkow, N. D. & Morales, M. The brain on drugs: from reward to addiction. *Cell* **162,** 712–725 (2015).
12. Scofield, M. D. *et al.* The Nucleus Accumbens: Mechanisms of Addiction across Drug Classes Reflect the Importance of Glutamate Homeostasis. *Pharmacol. Rev.* **68,** 816–871 (2016).
13. Schall, T. A., Wright, W. J. & Dong, Y. Nucleus accumbens fast-spiking interneurons in motivational and addictive behaviors. *Mol. Psychiatry* (2020). doi:10.1038/s41380-020-0683-y
14. Lee, J. H. *et al.* Dopaminergic regulation of nucleus accumbens cholinergic interneurons demarcates susceptibility to cocaine addiction. *Biol. Psychiatry* (2020). doi:10.1016/j.biopsych.2020.05.003
15. Avey, D. *et al.* Single-Cell RNA-Seq Uncovers a Robust Transcriptional Response to Morphine by Glia. *Cell Rep.* **24,** 3619–3629.e4 (2018).
16. Fullard, J. F. *et al.* Open chromatin profiling of human postmortem brain infers functional roles for non-coding schizophrenia loci. *Hum. Mol. Genet.* **26,** 1942–1951 (2017).
17. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19,** 1442–1453 (2016).
18. Dong, X. *et al.* Enhancers active in dopamine neurons are a primary link between genetic variation and neuropsychiatric disease. *Nat. Neurosci.* **21,** 1482–1492 (2018).
19. Tak, Y. G. & Farnham, P. J. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin* **8,** 57 (2015).
20. Brown, C. D., Mangravite, L. M. & Engelhardt, B. E. Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PLoS Genet.* **9,** e1003649 (2013).
21. van Arensbergen, J. *et al.* High-throughput identification of human SNPs affecting regulatory element activity. *Nat. Genet.* **51,** 1160–1169 (2019).
22. Bastle, R. M. & Maze, I. Chromatin regulation in complex brain disorders. *Curr. Opin. Behav. Sci.* **25,** 57–65 (2019).
23. Forrest, M. P. *et al.* Open Chromatin Profiling in hiPSC-Derived Neurons Prioritizes Functional Noncoding Psychiatric Risk Variants and Highlights Neurodevelopmental Loci. *Cell Stem Cell* **21,** 305–318.e8 (2017).
24. Shu, C. *et al.* Epigenome-wide study of brain DNA methylation among opioid users and controls. *BioRxiv* (2020). doi:10.1101/2020.11.10.377069
25. Yoon, K.-J. *et al.* Modeling a genetic risk for schizophrenia in iPSCs and mice reveals neural stem cell

deficits associated with adherens junctions and polarity. *Cell Stem Cell* **15,** 79–91 (2014).
26. Bryois, J. *et al.* Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat. Commun.* **9,** 3121 (2018).
27. Corces, M. R. *et al.* Single-cell epigenomic identification of inherited risk loci in Alzheimer's and Parkinson's disease. *BioRxiv* (2020). doi:10.1101/2020.01.06.896159
28. Roadmap Epigenomics Consortium *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–330 (2015).
29. Heinz, S., Romanoski, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* **16,** 144–154 (2015).
30. Nott, A. *et al.* Brain cell type-specific enhancer-promoter interactome maps and disease-risk association. *Science* **366,** 1134–1139 (2019).
31. Hrvatin, S. *et al.* A scalable platform for the development of cell-type-specific viral drivers. *Elife* **8,** (2019).
32. Blankvoort, S., Witter, M. P., Noonan, J., Cotney, J. & Kentros, C. Marked Diversity of Unique Cortical Enhancers Enables Neuron-Specific Tools by Enhancer-Driven Gene Expression. *Curr. Biol.* **28,** 2103–2114.e5 (2018).
33. Minnoye, L. *et al.* Cross-species analysis of enhancer logic using deep learning. *Genome Res.* (2020). doi:10.1101/gr.260844.120
34. Satpathy, A. T. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37,** 925–936 (2019).
35. McCole, R. B., Erceg, J., Saylor, W. & Wu, C. Ultraconserved Elements Occupy Specific Arenas of Three-Dimensional Mammalian Genome Organization. *Cell Reports* (2018).
36. Sanges, R. *et al.* Highly conserved elements discovered in vertebrates are present in non-syntenic loci of tunicates, act as enhancers and can be transcribed during development. *Nucleic Acids Res.* **41,** 3600–3618 (2013).
37. Klein, J. C., Keith, A., Agarwal, V., Durham, T. & Shendure, J. Functional characterization of enhancer evolution in the primate lineage. *Genome Biol.* **19,** 99 (2018).
38. Flores, M. A. & Ovcharenko, I. Enhancer reprogramming in mammalian genomes. *BMC Bioinformatics* **19,** 316 (2018).
39. Yang, S. *et al.* Functionally conserved enhancers with divergent sequences in distant vertebrates. *BMC Genomics* **16,** 882 (2015).
40. Chen, L., Fish, A. E. & Capra, J. A. Prediction of gene regulatory enhancers across species reveals evolutionarily conserved sequence properties. *PLoS Comput. Biol.* **14,** e1006484 (2018).
41. He, J. *et al.* Transcriptional diversity of medium spiny neurons in the primate striatum. *BioRxiv* (2020). doi:10.1101/2020.10.25.354159
42. Srinivasan, C. *et al.* Addiction-associated genetic variants implicate brain cell type- and region-specific cis-regulatory elements in addiction neurobiology. *BioRxiv* (2020). doi:10.1101/2020.09.29.318329
43. Stanley, G., Gokce, O., Malenka, R. C., Südhof, T. C. & Quake, S. R. Continuous and discrete neuron types of the adult murine striatum. *Neuron* **105,** 688–699.e8 (2020).
44. Gokce, O. *et al.* Cellular Taxonomy of the Mouse Striatum as Revealed by Single-Cell RNA-Seq. *Cell Rep.* **16,** 1126–1137 (2016).
45. Zeisel, A. *et al.* Molecular architecture of the mouse nervous system. *Cell* **174,** 999–1014.e22 (2018).
46. Savell, K. E. *et al.* A dopamine-induced gene expression signature regulates neuronal function and cocaine response. *Sci. Adv.* **6,** eaba4221 (2020).
47. Tran, M. N. *et al.* Single-nucleus transcriptome analysis reveals cell type-specific molecular signatures across reward circuitry in the human brain. *BioRxiv* (2020). doi:10.1101/2020.10.07.329839
48. Karlsson Linnér, R. *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.* **51,** 245–257 (2019).
49. Pasman, J. A. *et al.* GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat. Neurosci.* **21,** 1161–1170 (2018).
50. Johnson, E. C. *et al.* A large-scale genome-wide association study meta-analysis of cannabis use disorder. *Lancet Psychiatry* (2020). doi:10.1016/S2215-0366(20)30339-4
51. Munn-Chernoff, M. A. *et al.* Shared genetic risk between eating disorder- and substance-use-related phenotypes: Evidence from genome-wide association studies. *Addict Biol* e12880 (2020). doi:10.1111/adb.12880
52. Zhou, H. *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.* **23,** 809–818 (2020).
53. Karlsson Linner, R. *et al.* Multivariate genomic analysis of 1.5 million people identifies genes related to addiction, antisocial behavior, and health. *BioRxiv* (2020). doi:10.1101/2020.10.16.342501
54. Mallard, T. T. *et al.* Multivariate GWAS elucidates the genetic architecture of alcohol consumption and misuse, corrects biases, and reveals novel associations with disease. *BioRxiv* (2020). doi:10.1101/2020.09.21.304196
55. Li, Y. *et al.* An atlas of gene regulatory elements in adult mouse cerebrum. *BioRxiv* (2020). doi:10.1101/2020.05.10.087585
56. Bose, J., Hedden, S. L., Lipari, R. N. & Park-Lee, E. Key Substance Use and Mental Health Indicators in the United States:Results from the 2017 National Survey on Drug Use and Health. (2018). at <https://www.samhsa.gov/data/report/2017-nsduh-annual-national-report>
57. Eddie, D., Greene, M. C., White, W. L. & Kelly, J. F. Medical burden of disease among individuals in recovery from alcohol and other drug problems in the united states: findings from the national recovery survey. *J Addict Med* **13,** 385–395 (2019).
58. Wang, Q. Q., Kaelber, D. C., Xu, R. & Volkow, N. D. COVID-19 risk and outcomes in patients with substance use disorders: analyses from electronic health records in the United States. *Mol. Psychiatry* (2020). doi:10.1038/s41380-020-00880-7
59. Welty, L. J. *et al.* Health Disparities in Drug- and Alcohol-Use Disorders: A 12-Year Longitudinal Study of Youths After Detention. *Am. J. Public Health* **106,** 872–880 (2016).
60. Teeters, J. B., Lancaster, C. L., Brown, D. G. & Back, S. E. Substance use disorders in military veterans: prevalence and treatment challenges. *Subst. Abuse Rehabil.* **8,** 69–77 (2017).
61. Pullen, E. & Oser, C. Barriers to substance abuse treatment in rural and urban communities: counselor perspectives. *Subst Use Misuse* **49,** 891–901 (2014).
62. Hoggatt, K. J., Frayne, S. M., Saechao, F. S., Yano, E. M. & Washington, D. L. Substance Use Disorder-Related Disparities in Patient Experiences of Primary Care. *Health Equity* **3,** 193–197 (2019).
63. Koob, G. F. & Volkow, N. D. Neurocircuitry of addiction. *Neuropsychopharmacology* **35,** 217–238 (2010).
64. Ikemoto, S. & Bonci, A. Neurocircuitry of drug reward. *Neuropharmacology* **76 Pt B,** 329–341 (2014).
65. Salgado, S. & Kaplitt, M. G. The nucleus accumbens: A comprehensive review. *Stereotact Funct Neurosurg* **93,** 75–93 (2015).
66. Hong, S.-I., Kang, S., Chen, J.-F. & Choi, D.-S. Indirect Medium Spiny Neurons in the Dorsomedial Striatum Regulate Ethanol-Containing Conditioned Reward Seeking. *J. Neurosci.* **39,** 7206–7217 (2019).
67. Golden, S. A. *et al.* Nucleus Accumbens Drd1-Expressing Neurons Control Aggression Self-Administration and Aggression Seeking in Mice. *J. Neurosci.* **39,** 2482–2496 (2019).
68. Wang, X. *et al.* Parvalbumin Interneurons of the Mouse Nucleus Accumbens are Required For Amphetamine-Induced Locomotor Sensitization and Conditioned Place Preference. *Neuropsychopharmacology* **43,** 953–963 (2018).
69. Ribeiro, E. A. *et al.* Transcriptional and physiological adaptations in nucleus accumbens somatostatin interneurons that regulate behavioral responses to cocaine. *Nat. Commun.* **9,** 3149 (2018).
70. Bracci, E., Centonze, D., Bernardi, G. & Calabresi, P. Dopamine excites fast-spiking interneurons in the striatum. *J. Neurophysiol.* **87,** 2190–2194 (2002).
71. Lansink, C. S., Goltstein, P. M., Lankelma, J. V. & Pennartz, C. M. A. Fast-spiking interneurons of the rat ventral striatum: temporal coordination of activity with principal cells and responsiveness to reward. *Eur. J. Neurosci.* **32,** 494–508 (2010).
72. Wiltschko, A. B., Pettibone, J. R. & Berke, J. D. Opposite effects of stimulant and antipsychotic drugs on striatal fast-spiking interneurons. *Neuropsychopharmacology* **35,** 1261–1270 (2010).
73. Seney, M. *et al.* Transcriptional alterations in opioid use disorder reveal an interplay between neuroinflammation and synaptic remodeling. *BioRxiv* (2020). doi:10.1101/2020.09.14.296707
74. Kendler, K. S. & Prescott, C. A. Cannabis use, abuse, and dependence in a population-based sample of female twins. *Am. J. Psychiatry* **155,** 1016–1022 (1998).
75. Kendler, K. S. & Prescott, C. A. Cocaine use, abuse and dependence in a population-based sample of female twins. *Br. J. Psychiatry* **173,** 345–350 (1998).
76. Erzurumluoglu, A. M. *et al.* Meta-analysis of up to 622,409 individuals identifies 40 novel smoking behaviour associated genetic loci. *Mol. Psychiatry* (2019). doi:10.1038/s41380-018-0313-0
77. Liu, M. *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.* **51,** 237–244 (2019).
78. Cheng, Z. *et al.* Genome-wide Association Study Identifies a Regulatory Variant of RGMA Associated

With Opioid Dependence in European Americans. *Biol. Psychiatry* **84,** 762–770 (2018).
79. Cabana-Domínguez, J., Shivalikanjli, A., Fernàndez-Castillo, N. & Cormand, B. Genome-wide association meta-analysis of cocaine dependence: Shared genetics with comorbid conditions. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **94,** 109667 (2019).
80. Kalsi, G. *et al.* Genome-Wide Association of Heroin Dependence in Han Chinese. *PLoS One* **11,** e0167388 (2016).
81. Kranzler, H. R. *et al.* Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat. Commun.* **10,** 1499 (2019).
82. Hodge, R. D. *et al.* Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573,** 61–68 (2019).
83. Xu, X. *et al.* Species and cell-type properties of classically defined human and rodent neurons and glia. *Elife* **7,** (2018).
84. Bakken, T. E. *et al.* Evolution of cellular diversity in primary motor cortex of human, marmoset monkey, and mouse. *BioRxiv* (2020). doi:10.1101/2020.03.31.016972
85. Krienen, F. M. *et al.* Innovations present in the primate interneuron repertoire. *Nature* **586,** 262–269 (2020).
86. Gallagher, M. D. & Chen-Plotkin, A. S. The Post-GWAS Era: From Association to Function. *Am. J. Hum. Genet.* **102,** 717–730 (2018).
87. Edwards, S. L., Beesley, J., French, J. D. & Dunning, A. M. Beyond GWASs: illuminating the dark road from association to function. *Am. J. Hum. Genet.* **93,** 779–797 (2013).
88. Schaub, M. A., Boyle, A. P., Kundaje, A., Batzoglou, S. & Snyder, M. Linking disease associations with regulatory information in the human genome. *Genome Res.* **22,** 1748–1759 (2012).
89. Nestler, E. J. & Lüscher, C. The molecular basis of drug addiction: linking epigenetic to synaptic and circuit mechanisms. *Neuron* **102,** 48–59 (2019).
90. Hamilton, P. J. & Nestler, E. J. Epigenetics and addiction. *Curr. Opin. Neurobiol.* **59,** 128–136 (2019).
91. Walker, D. M. & Nestler, E. J. Neuroepigenetics and addiction. *Handb Clin Neurol* **148,** 747–765 (2018).
92. Robison, A. J. & Nestler, E. J. Transcriptional and epigenetic mechanisms of addiction. *Nat. Rev. Neurosci.* **12,** 623–637 (2011).
93. Zoonomia Consortium. A comparative genomics multitool for scientific discovery and conservation. *Nature* **587,** 240–245 (2020).
94. Zhou, H. *et al.* In vivo simultaneous transcriptional activation of multiple genes in the brain using CRISPR-dCas9-activator transgenic mice. *Nat. Neurosci.* **21,** 440–446 (2018).
95. Volkow, N. D. *et al.* Decreased striatal dopaminergic responsiveness in detoxified cocaine-dependent subjects. *Nature* **386,** 830–833 (1997).
96. Volkow, N. D. *et al.* Decreases in dopamine receptors but not in dopamine transporters in alcoholics. *Alcohol. Clin. Exp. Res.* **20,** 1594–1598 (1996).
97. Wang, G. J. *et al.* Dopamine D2 receptor availability in opiate-dependent subjects before and after naloxone-precipitated withdrawal. *Neuropsychopharmacology* **16,** 174–182 (1997).
98. Volkow, N. D. *et al.* Low level of brain dopamine d_2 receptors in methamphetamine abusers: association with metabolism in the orbitofrontal cortex. *Focus (Madison)* **1,** 150–157 (2003).
99. Charbogne, P., Kieffer, B. L. & Befort, K. 15 years of genetic approaches in vivo for addiction research: Opioid receptor and peptide gene knockout in mouse models of drug abuse. *Neuropharmacology* **76 Pt B,** 204–217 (2014).
100. Lynch, W. J., Nicholson, K. L., Dance, M. E., Morgan, R. W. & Foley, P. L. Animal models of substance abuse and addiction: implications for science, animal welfare, and society. *Comp Med* **60,** 177–188 (2010).
101. Farrell, M. R., Schoch, H. & Mahler, S. V. Modeling cocaine relapse in rodents: Behavioral considerations and circuit mechanisms. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **87,** 33–47 (2018).
102. Svensson, V. *et al.* Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* **14,** 381–387 (2017).
103. Parsana, P. *et al.* Addressing confounding artifacts in reconstruction of gene co-expression networks. *Genome Biol.* **20,** 94 (2019).
104. McCall, M. N., Illei, P. B. & Halushka, M. K. Complex sources of variation in tissue expression data: analysis of the gtex lung transcriptome. *Am. J. Hum. Genet.* **99,** 624–635 (2016).
105. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14,** 955–958 (2017).

106. Welch, J. *et al.* Integrative inference of brain cell similarities and differences from single-cell genomics. *BioRxiv* (2018). doi:10.1101/459891
107. Kiselev, V. Y. *et al.* SC3: consensus clustering of single-cell RNA-seq data. *Nat. Methods* **14,** 483–486 (2017).
108. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58,** 267–288 (1996).
109. Duane, S., Kennedy, A. D., Pendleton, B. J. & Roweth, D. Hybrid Monte Carlo. *Physics Letters B* **195,** 216–222 (1987).
110. Xiao, J., Cao, H. & Chen, J. False discovery rate control incorporating phylogenetic tree increases detection power in microbiome-wide multiple testing. *Bioinformatics* **33,** 2873–2881 (2017).
111. Saunders, A. *et al.* Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* **174,** 1015–1030.e16 (2018).
112. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* **8,** 281–291.e9 (2019).
113. van Dijk, D. *et al.* Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174,** 716–729.e27 (2018).
114. Lun, A. T. L., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17,** 75 (2016).
115. Gong, W., Kwak, I.-Y., Pota, P., Koyano-Nakagawa, N. & Garry, D. J. DrImpute: imputing dropout events in single cell RNA sequencing data. *BMC Bioinformatics* **19,** 220 (2018).
116. Polański, K. *et al.* BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* **36,** 964–965 (2020).
117. Mohammadi, S., Davila-Velderrain, J. & Kellis, M. A multiresolution framework to characterize single-cell state landscapes. *Nat. Commun.* **11,** 5399 (2020).
118. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550–550 (2014).
119. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43,** e47 (2015).
120. Bulik-Sullivan, B. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).
121. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50,** 621–629 (2018).
122. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11,** e1004219 (2015).
123. Sey, N. Y. A. *et al.* A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nat. Neurosci.* **23,** 583–593 (2020).
124. Timshel, P. N., Thompson, J. J. & Pers, T. H. Genetic mapping of etiologic brain cell types for obesity. *Elife* **9,** (2020).
125. Calderon, D. *et al.* Inferring Relevant Cell Types for Complex Traits by Using Single-Cell Gene Expression. *Am. J. Hum. Genet.* **101,** 686–699 (2017).
126. Corces, M. R. *et al.* Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. *Nat. Genet.* **52,** 1158–1168 (2020).
127. Lawler, A. J. *et al.* Cell type-specific oxidative stress genomic signatures in the globus pallidus of dopamine depleted mice. *J. Neurosci.* (2020). doi:10.1523/JNEUROSCI.1634-20.2020
128. Zhang, X., Kaplow, I. M., Wirthlin, M., Park, T. Y. & Pfenning, A. R. HALPER facilitates the identification of regulatory element orthologs across species. *Bioinformatics* **36,** 4339–4340 (2020).
129. Hickey, G., Paten, B., Earl, D., Zerbino, D. & Haussler, D. HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29,** 1341–1342 (2013).
130. Paten, B. *et al.* Cactus: Algorithms for genome multiple sequence alignment. *Genome Res.* **21,** 1512–1528 (2011).
131. Armstrong, J. *et al.* Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature* **587,** 246–251 (2020).
132. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359 (2012).
133. Granja, J. M. *et al.* ArchR: An integrative and scalable software package for single-cell chromatin accessibility analysis. *BioRxiv* (2020). doi:10.1101/2020.04.28.066498

134. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
135. Li, Z. *et al.* scOpen: chromatin-accessibility estimation of single-cell ATAC data. *BioRxiv* (2019). doi:10.1101/865931
136. Pliner, H. A. *et al.* Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol. Cell* **71,** 858–871.e8 (2018).
137. Yager, L. M., Garcia, A. F., Wunsch, A. M. & Ferguson, S. M. The ins and outs of the striatum: role in drug addiction. *Neuroscience* **301,** 529–541 (2015).

Facilities and other resources

Laboratory Facilities. The Pfenning laboratory is physically located between shared space for neurobiology researchers and the Molecular Bio-Imaging Center (MBIC), with many shared resources to which the Dr. Pfenning has 24/7 access, including fluorescence microscopy. The lab consists of 755 sq. ft. of wet-bench space. The Pfenning wet laboratory and nearby collaborative labs house all necessary equipment and facilities required for the molecular biology aspects of the project.

The Pfenning lab has newly renovated office space to provide the PI, postdocs, students, both undergraduate and graduate, and the lab manager space outside the lab to have small conferences and discussions and provide a productive working environment. The PI has an office, with a small conference area for private discussions and meetings. An additional space of 357 sq. ft. is available for students, postdocs, and technicians, including four desks and a medium sized conference area. Within the wet lab, there is a desk for the lab manager and two desks for anyone to use while performing experiments.

Office and Computer. I dedicated access to two offices: 1) a workspace in the experimental laboratory and 2) a designated desk in the Computational Biology Department office space reserved for CPCB students in my cohort. I have full access to all hardware and software needed for data entry and analysis pertinent to this proposal, including:

- Sole use of a new Apple Mac Mini (3.0Ghz 6 Core Intel i5, 16Gb RAM, 512 Gb SSD)
- Sole use of two new HD 24” Dell U2419HC monitors
- Sole use of a new Apple iPad Air for portable use as a digital lab notebook
- Access to all lab-shared computers (Apple Mac lab computer, Dell laptop, and other lab iPads)
- Access to the CBD Lane Compute Cluster (all shared departmental and private Pfenning lab nodes)
- Access to the Pittsburgh Supercomputing Center
- Software for image processing (ImageJ, Adobe Illustrator, Inkscape)
- Software for data analysis (Matlab 2020, Mathematica 2020)

Animal Facilities. The Mellon Institute Central Vivarium, shared between many laboratories in biological sciences, has space to accommodate at least 100 more cages. This facility is equipped for BSL-3 level virus work.

Heterogeneous Distributed Computing. The Carnegie Mellon University (CMU) School of Computer Science (SCS) research facility has a large number and wide variety of computers available for faculty and graduate student use -- approximately 4000 machines. About one-third are Linux/Unix on Intel and AMD platforms, 60% are Windows systems, and Macintosh computers make up the remainder. Every incoming graduate student is provided with a new, high-powered personal computer; some receive a dual-boot configuration – both Windows and Linux. SCS facilities include a rich variety of computing infrastructure services of very high quality: email, shared file service (AFS), authentication, remote access services (VPN, iPass), backup, printing, software licensing, hardware repair, and so on. The SCS environment also includes a growing number of high-performance compute clusters; support services are available for the entire life-cycle of the cluster, including help for specification and purchasing. For all aspects of computing, there is a dedicated support (Help) staff within the facility, which provides full support for users, applications, machines, and services, via a menu of premium support services. For deep learning, the computational biology cluster at CMU is currently equipped with 3 GPU nodes: “2 x Quantum TXR231-1000R (128GB system memory, 4 x GTX 1080 GPU's)”.

Beyond these college resources, the University maintains computation facilities of various kinds for general use. Pittsburgh Supercomputing Center (PSC) is a joint effort of Carnegie Mellon and the University of Pittsburgh together with Westinghouse Corporation. It is supported by several federal agencies, the Commonwealth of Pennsylvania and private industry. It is a leading partner in the TeraGrid, the National Science Foundation’s cyberinfrastructure program. It operates several supercomputing-class machines, including an SGI UV shared memory machine with 4096 cores and 32TBytes of shared memory. We currently have XSEDE resources through the Pittsburgh Supercomputing Center (Grant MCB190162 “Investigating the Evolution, Spatial Profiles and Gene Therapy Access Points of Neuronal Cell Populations Using Single Cell Sequencing and Analysis

Approaches"; PSC GPU (Bridges GPU): 5,000.0 GPU Hours; PSC Storage (Bridges Pylon): 8,000.0 GB; PSC Large Memory (Bridges Large): 14,100.0 Memory Hours; PSC Regular Memory (Bridges): 14,500.0 Sus)

Networking. Carnegie Mellon operates a fully interconnected, multimedia, multiprotocol campus network. The system incorporates state-of-the-art commercial technology and spans all campus buildings in a redundant 10Gbps backbone infrastructure that enables access to all campus systems, including the PSC supercomputers. The University also provides Wi-Fi connectivity in all campus buildings; it has recently upgraded to equipment that supports the 802.11n standard, which provides wireless speeds in excess of 100Mbps.

SCS has redundant 10Gbps links to the Carnegie Mellon campus network. The University has redundant 1Gbps links to a combination of providers for internet connectivity. These include Sprint and Level3 for commodity internet traffic, and the Three Rivers Optical Exchange (3ROX) for connections to a number of high-speed research and education networks, including Internet2, National Lambda Rail, ESnet, and Teragrid. The University can also arrange advanced point-to-point research connectivity through services such as Internet2's Dynamic Circuit Network.

General Facilities Information. Carnegie Mellon's School of Computer Science is the largest academic organization devoted to the study of computers. Its seven degree-granting departments—the Computer Science Department, the Human-Computer Interaction Institute, the Institute for Software Research, the Computational Biology Department, the Language Technologies Institute, the Machine Learning Department, and the Robotics Institute—include over 250 faculty, 700 graduate students, and a 250-member professional technical staff. SCS also collaborates with other University Research Centers, including the Software Engineering Institute (SEI), the Pittsburgh Supercomputing Center (PSC), the Information Networking Institute (INI), the Institute for Complex Engineered Systems (ICES), the Center for the Neural Basis of Cognition (CNBC), and the Entertainment Technology Center (ETC).

Environment. The research environment in computational genomics is especially strong. The Department of Computational Biology within the School of Computer Science provides a close-knit community of several faculty developing new computational approaches to analyze genomic data. This includes Dr. Ziv Bar-Joseph, Dr. Eric Xing, Dr. Carl Kingsford, Dr. Jian Ma, Dr. Russell Schwartz, and Dr. Seyoung Kim. Dr. Joel McManus who is affiliated with department is also developing high-throughput reporter assay methods and integrating computational and experimental approaches.

The research environment for neuroscience at Carnegie Mellon is excellent. Three neuroscientist faculty (Dr. Sandy Kuhlman, Dr. Aryn Gittis, Dr. Eric Yttri, and Dr. Alison Barth) in the department of Biological Sciences share contiguous laboratory and graduate student/post-doc office space. The Pfenning laboratory is directly next-door. Across the five groups are 12 graduate students, 2 post-docs, 6 technicians, and one research scientist. This organization enables daily, informal contact between investigators and trainees, facilitating resource sharing and collaboration. In addition, there are monthly research presentations given by neuroscience graduate students and post-docs. Due to close collaborations with these labs, the Pfenning lab has access to any equipment needed to preserve and slice tissue for histology or dissection of specific structures within the brain for down-stream molecular analysis. The local atmosphere is for neuroscience research is dynamic and interactive. CMU and Pitt located less than half a mile apart, together provide a rich intellectual environment for neuroscience research in Pittsburgh. This is formalized in the dual-university Center for the Neural Basis of Cognition, of which Dr. Pfenning is an active member.

The University of Pittsburgh (*Pitt*) is a major academic medical center with a well-established record of high quality, federally funded research. Pitt consistently ranks among the top 10 recipients of NIH funding and has recently risen to the fifth-highest recipient; the University and its affiliates received over $315 million in NIH support in 2017. Pitt's many research centers devoted to neuroscience create an exceptional training environment that is particularly strong in the areas of sensory biology, movement disorders, computational research, and neuroimaging. Pitt's programs for graduate medical education are highly regarded, with the Psychiatry and Neurology departments consistently ranking in the top 10 programs nationally. Pitt is truly a collaborative place, evidenced by its close ties to CMU. I have found the faculty to be extremely accessible, and I will continue to make invaluable connections with leading neuroscientists and physician-scientists throughout my training.

The University of Pittsburgh Medical Center (UPMC) is a non-profit health system comprising over 30 hospitals and 2700 physicians. In 2016, UPMC was ranked 12th nationally in *U.S. News and World Report's* Honor Roll of America's Best Hospitals. UPMC provides a large, diverse patient population for translational studies. Further, UPMC provides a dynamic hospital system for medical students to train. No other medical schools have access to UPMC facilities, guaranteeing access to the world-renowned faculty and resources of UPMC for Pitt medical students. Since there are few non-UPMC healthcare facilities in the region for specialized care, Pitt students encounter a diverse patient population from rural and urban areas within a ~150-mile radius of Pittsburgh, encompassing all of western Pennsylvania as well parts of West Virginia, New York, and Ohio. UPMC's generous financial support of biomedical research at Pitt further stabilizes research funding. Overall, UPMC contributes substantially to the financial and translational success of research at the University of Pittsburgh, in addition to providing excellent clinical training to MSTP students (see "Additional Education Information" for MSTP details).

The Pitt-CMU Center for the Neural Basis of Cognition (CNBC) is one of the nation's largest and most well-regarded cross-institutional programs in neuroscience. The program includes graduate students and faculty from Pitt's Center for Neuroscience (CNUP), CMU Biological Sciences, Computational Biology, Computer Science, Biomedical Engineering, Psychology, Statistics, Philosophy, and Machine Learning. The program includes over 200 faculty across numerous departments at Pitt, CMU, the CNBC, and UPMC, and numerous opportunities to interact with these faculty at multiple seminar series on both general and specialized neuroscience topics. The CNUP is supported in part by a large NIH pre-doctoral training grant and was one of eight centers nationwide to participate in the Carnegie Foundation's Initiative on the Doctorate to assess graduate education in neuroscience.

Equipment

General Equipment: The Pfenning lab is well-equipped for molecular biology with heat blocks, shakers, microcentrifuges, refrigerator, -20°C freezer, and -80°C freezer. Shared facilities in close proximity to the lab provide a Tapestation 4200 electrophoresis system, dry ice chest, wet ice chest, downdraft sink (for animal perfusions), and O_2/CO_2 tanks for tissue oxygenation. Space and fixtures include two fume hoods, a dark room, and dedicated rooms in the animal facility with temperature and lighting control.

Project-specific Equipment: The Pfenning lab specifically owns and operates the following equipment to perform nuclei isolation, single nucleus genomic assays, and in-house sequencing:

Nuclei Prep:

- Dedicated fresh tissue, semi-automated vibratome (Leica, VT1200)
- Countess II Fluorescence Automated Cell Counter (ThermoFisher, AMQAX1000)
- EVOS Upright Fluorescence Imaging System (Life Technologies, AMF4300)
- Fixed angle, refrigerated centrifuge (Eppendorf, 5430R/5428)
- Swinging-bucket, high-capacity refrigerated centrifuge (Eppendorf, 22331 Hamburg)

Single-nucleus Genomics:

- 10X Chromium Controller (10X Genomics)
- MiSeq Desktop Sequencer System (Illumina, SY-410-1003)
- SimpliAmp Thermocylcer (Life Technologies, A24811)
- QuantStudio 3 Real-Time PCR System (ThermoFisher, A28132)
- Qubit 3.0 Fluorometer (Life Technologies, Q33216)
- Dedicated DNA Bio-safety hood (Justright, 891200)
- Dedicated RNA Bio-safety hood (Justright, 891200)

Office space and software: All Pfenning lab students and personnel have dedicated office space within the immediate experimental lab space and in the Computational Biology Department (CBD) office space. I will work at my personal office desk and perform experiments within the experiment lab space. The Pfenning laboratory has individual PCs associated with each molecular biology system (RT-PCR, MiSeq), plus additional portable iPads for documenting experiments (8 total). For computational analysis with access to the two dedicated high-performance clusters, the CBD Lane Cluster and the Pittsburgh Supercomputing Center. The Pfenning Lab Lane cluster private partitions has more than 100Tb of dedicated cluster storage, backed up nightly, in addition to 1TB of cloud storage via CMU Box, a free service for CMU Students.

Media and Library Services: At CMU, I have full access to extensive services including 3D printing and poster printing via the CMU Media Services office and CMU School of Computer Science, as well as numerous licenses for biostatistics, cloud-based data storage, automatic file backup systems. I also have access to the Pitt main campus and Health Sciences Library Systems, which provide a suite of office software as well as literature search, data analysis, and scientific writing tutorials and consultation services.

OMB Number 4040-0010
Expiration Date 12/31/2022

RESEARCH & RELATED Senior/Key Person Profile (Expanded)

PROFILE - Project Director/Principal Investigator

Prefix: First Name*: BaDoi Middle Name Nguyen Last Name*: Phan Suffix:

Position/Title*: Medical Student, MD/PhD Training Program
Organization Name*: Carnegie Mellon University
Department: Computational Biology
Division:
Street1*: 5000 Forbes Ave
Street2:
City*: Pittsburgh
County: PA
State*: PA: Pennsylvania
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 152133890

Phone Number*: 4086135493 Fax Number:

E-Mail*: badoi.phan@pitt.edu

Credential, e.g., agency login: badoiphan

Project Role*: PD/PI Other Project Role Category:

Degree Type: Degree Year:

Attach Biographical Sketch*: File Name: Biosketch_resub_final_Phan.pdf

Attach Current & Pending Support: File Name:

Track ng Number:

Fund ng Opportun ty Number: . Rece ved Date:

PROFILE - Senior/Key Person

Prefix: Prof. First Name*: Andreas Middle Name Last Name*: Pfenning Suffix: Ph.D

Position/Title*: Assistant Professor
Organization Name*: Carnegie Mellon University
Department: Computational Biology
Division:
Street1*: 5000 Forbes Ave
Street2:
City*: Pittsburgh
County: PA
State*: PA: Pennsylvania
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 15213-3890

Phone Number*: 412-268-5518 Fax Number:

E-Mail*: apfenning@cmu.edu

Credential, e.g., agency login: APFENNING

Project Role*: Other (Specify) Other Project Role Category: SPONSOR

Degree Type: PhD Degree Year: 2012

Attach Biographical Sketch*: File Name: biosketch_Pfenning.pdf

Attach Current & Pending Support: File Name:

PROFILE - Senior/Key Person

Prefix: Prof. First Name*: Ryan Middle Name Last Name*: Logan Suffix: Ph.D

Position/Title*: Assistant Professor of Psychiatry
Organization Name*: University of Pittsburgh
Department: Psychiatry
Division:
Street1*: 450 Technology Drive
Street2:
City*: Pittsburgh
County:
State*: PA: Pennsylvania
Province:
Country*: USA: UNITED STATES
Zip / Postal Code*: 152190000

Phone Number*: 4123833084 Fax Number:

E-Mail*: loganw@upmc.edu

Credential, e.g., agency login: RWLOGAN

Project Role*: Other (Specify) Other Project Role Category: CO-SPONSOR

Degree Type: PhD Degree Year: 2009

Attach Biographical Sketch*: File Name: Biosketch_Logan.pdf

Attach Current & Pending Support: File Name:

OMB No. 0925-0001 and 0925-0002 (Rev. 03/2020 Approved Through 02/28/2023)

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Phan, BaDoi

eRA COMMONS USER NAME (credential, e.g., agency login): **badoiphan**

POSITION TITLE: MSTP Trainee

EDUCATION/TRAINING

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
Johns Hopkins University, Baltimore, MD	BS	05/2016	Biomedical Engineering
University of Pittsburgh, Pittsburgh, PA	M.D.	05/2025	Medicine
Carnegie Mellon University, Pittsburgh, PA	Ph.D.	05/2023	Computational Biology

A. Personal Statement

My overall career goal is to apply computational and biologic research to investigate neuropsychiatric disorders. I have always been interested in applying computational approaches to investigating biological questions. Before joining the medical scientist training program (MSTP) at Carnegie Mellon University (CMU) and the University of Pittsburgh (Pitt), I applied bioinformatics and genomics approaches to investigate convergent gene pathways of neurodevelopmental disorders at the Lieber Institute for Brain Development. I collaborated experimental neuroscientists to apply systematic computation towards unbiased image quantification. My research efforts produced a first author publication at *Nature Neuroscience* and co-author of ten other publications and preprints.

Currently, I am mentored by Dr. Andreas Pfenning, who leads the Neurogenomics lab at CMU and is an expert in computational biology and neuroscience. His lab environment is a perfect synergy of experimental and computational neuroscience research. During my first year in the lab, I contributed to a research talk presented at the 2020 NIDA Genetics Consortium Meeting that saw that addiction-associated risk variants enrich within cell type-specific gene-regulatory regions. This co-first author work is currently in review at the *Journal of Neuroscience*, and these findings drive the questions I propose to investigate in this proposal. I will test the hypothesis that human SUD risk variants enrich in nucleus accumbens cell type marker genes and cis-regulatory elements that are conserved in primates and rodents. In completing this training fellowship, I will gain skills to generate high-quality single cell genomic data, build algorithms to model species conservation at single-cell resolution, and investigate genetic variation in substance use addiction.

My training plan capitalizes on the extensive experimental and computational resources of Pitt, CMU, and the joint CMU-Pitt Computational PhD training program. These centers are at the forefront of technology development in genomics and neuroscience research making them the ideal institutions for interdisciplinary training. I have formed my thesis committee including my advisor and prominent experts in single-cell genomic algorithms, addiction neurobiology, statistics and human genetics of neuropsychiatric disorders. I will also take advantage of the region's superb clinical instruction. In years 2-3 of my PhD, I will do two longitudinal clerkships including one with expert addiction psychiatrist, Dr. Antoine Douaihy, to gain clinical skills in addiction medicine and learn where in the field can my research make measurable impacts.

After completing my MD/PhD training, I plan to pursue a residency in medicine or psychiatry to prepare myself to take on an independent investigator position at an academic medical center, where I will balance a career in research and patient care. My research program will apply my intersectional training in computational biology and neuroscience to explore the mechanisms of complex brain disorders with strong genetic influences. The patients that I will see in clinic will drive the focus and goals of my research. I enjoy bridging the distinct perspectives across the different domains of my medical scientist training: medicine, neuroscience, computational biology. I am excited to continue my training in the invigorating and inclusive environment of CMU and Pitt. The field of neurogenomics is rapidly growing, and I look forward to contributing to it as a young investigator.

Selected Publications:

a. Srinivasan C*, **Phan BN***, Lawler AJ, Ramamurthy E, Kleyman M, Brown AR, Kaplow IM, Wirthlin ME, Pfenning AR. Addiction-associated genetic variants implicate brain cell type- and region-specific cis-regulatory elements in addiction neurobiology. bioRxiv 2020.09.29.318329; doi: https://doi.org/10.1101/2020.09.29.318329. *In review*
b. Seney ML, Moon-Kim S, Glausier JR, Hildebrand MA, XiangningXue, Zong W, Wang J, Shelton MA, **Phan BN**, Srinivasan C, Pfenning AR, Tseng GC, Lewis DA, Freyberg Z, Logan RW. Transcriptional alterations in opioid use disorder reveal an interplay between neuroinflammation and synaptic remodeling. bioRxiv 2020.09.14.296707; doi: https://doi.org/10.1101/2020.09.14.296707. *In review*
c. **Phan BN**, Bohlen JF, Davis BA, Ye Z, Chen HY, Mayfield B, Sripathy SR, Cerceo Page S, Campbell MN, Smith HL, Gallop D, Kim H, Thaxton CL, Simon JM, Burke EE, Shin JH, Kennedy AJ, Sweatt JD, Philpot BD, Jaffe AE, Maher BJ. A myelin-related transcriptomic profile is shared by Pitt-Hopkins syndrome models and human autism spectrum disorder. **Nat Neurosci.** 2020 Feb 3; PubMed PMID: 32015540; PubMed Central PMCID: PMC7065955.

B. Positions and Honors

Positions and Employment

2013 - 2013 Summer Research Assistant, National University of Singapore, Dept. of Biomedical Engineering, Singapore
2014 - 2017 Research Assistant, Lieber Institute for Brain Development, Baltimore, MD
2015 - 2015 Summer Research Assistant, Oxford University, Dept. of Psychiatry, Oxford
2015 - 2017 Teaching Assistant, Johns Hopkins University, Dept. of Biomedical Engineering, Baltimore, MD
2017 - 2017 Summer Rotation Student, Carnegie Mellon University, Dept. of Computational Biology, Pittsburgh, PA
2018 - 2018 Summer Rotation Student, University of Pittsburgh, Dept. of Computational and Systems Biology, Pittsburgh, PA
2019 – present Graduate Student, Carnegie Mellon University, Dept. of Computational Biology, Pittsburgh, PA

Other Experience and Professional Memberships

2015 - Member, Alpha Eta Mu Beta Biomedical Engineering Honors Society
2016 - Member, Tau Beta Pi Engineering Honors Society
2018 - Member, American Psychiatry Association
2018 - 2019 Founder/coordinator, Knitt Med Student Group, University of Pittsburgh
2020 - Treasurer, American Physician Scientist Association, University of Pittsburgh Chapter

Honors

2012 - 2015 Dean's List, Johns Hopkins University
2013 - 2013 Study Abroad Scholarship, Benjamin A. Gilman International Scholarship
2013 - 2015 Technology Fellowship, Johns Hopkins University, Center for Education Resources
2015 - 2015 Vredenburg Study Abroad Scholarship, Johns Hopkins University, School of Engineering
2016 - 2016 David T. Yue Undergraduate Teaching Award, Johns Hopkins University, Dept. of Biomedical Engineering
2016 - 2017 Provost Undergraduate Research Award, Johns Hopkins University
2017 - 2018 Recognition of Excellence in Collaborative Learning, University of Pittsburgh School of Medicine
2019 BGSA Conference Travel Award, University of Pittsburgh School of Medicine

C. Contributions to Science

1. Autism Spectrum Disorder Transcriptomics: I continued my research on TCF4 with Dr. Brady Maher's group and became co-mentored by Dr. Andrew Jaffe (Lieber Institute) on RNA sequencing (RNA-seq) data analysis and bioinformatics. During this time, I applied bioinformatic and computational analyses to identify transcriptional changes in the brains of five independent mouse models of PTHS and identified considerable overlap in differentially expressed genes (DEGs). I identified oligodendrocyte-specific genes were dysregulated and confirmed that this transcriptional signature is also present in two additional mouse models of syndromic autism spectrum disorder (ASD). Moreover, I found significant overlap of syndromic ASD mouse DEGs with identified ASD risk genes and human idiopathic ASD postmortem brain RNA-seq

associated with myelination. These results from seven independent mouse models of ASD are validated in human brain, implicating disruptions in myelination as shared mechanisms in ASD pathophysiology.

a. **Phan BN***, Bohlen JF*, Davis BA, Ye Z, Chen HY, Mayfield B, Sripathy SR, Cerceo Page S, Campbell MN, Smith HL, Gallop D, Kim H, Thaxton CL, Simon JM, Burke EE, Shin JH, Kennedy AJ, Sweatt JD, Philpot BD, Jaffe AE, Maher BJ. A myelin-related transcriptomic profile is shared by Pitt-Hopkins syndrome models and human autism spectrum disorder. **Nat Neurosci**. 2020 Feb 3;PubMed PMID: 32015540; PubMed Central PMCID: PMC7065955.

2. Pitt-Hopkins Syndrome Research: I was a member of Dr. Brady Maher's lab at the Lieber Institute for Brain Development investigating the gene TCF4. Heterozygous TCF4 mutation confers Pitt-Hopkins Syndrome (PTHS), a rare congenital disorder marked by severe intellectual disability, which my lab modeled using in utero electroporation shRNA knockdown or CRISPR mutation to study the effects of decreased TCF4 expression on layer ⅔ pyramidal neurons in rodent medial prefrontal cortex. In this space, I analyzed genome-wide binding targets of TCF4 from ChIP-seq of mouse neuroblastoma cell lines and primary cultures of neural progenitor cells. I contributed to creating CRISPR tools to mutate TCF4. Lastly, I developed algorithms to process and analyze calcium imaging of *ex vivo* whole-brain slice preparations for systematic, higher-throughput electrophysiological investigation of TCF4 knockdown and knockout. Piecing these projects together, our lab identified that TCF4 binds within the genes SCN10a and KCNQ1 to regulate expression of these genes in primary neural progenitor cell cultures. These investigations defined the effects of decreased TCF4 mutations and identified potential drug-able targets and pharmacological studies for Pitt-Hopkins Syndrome. My contributions to these studies were included in two publications in Neuron and Molecular Psychiatry.
    a. Page SC, Hamersky GR, Gallo RA, Rannals MD, Calcaterra NE, Campbell MN, Mayfield B, Briley A, **Phan BN,** Jaffe AE, Maher BJ. The schizophrenia- and autism-associated gene, transcription factor 4 regulates the columnar distribution of layer 2/3 prefrontal pyramidal neurons in an activity-dependent manner. **Mol Psychiatry**. 2018 Feb;23(2):304-315. doi: 10.1038/mp.2017.37. Epub 2017 Mar 14. PubMed PMID: 28289282; PubMed Central PMCID: PMC5599320.
    b. Rannals MD, Hamersky GR, Page SC, Campbell MN, Briley A, Gallo RA, **Phan BN,** Hyde TM, Kleinman JE, Shin JH, Jaffe AE, Weinberger DR, Maher BJ. Psychiatric Risk Gene Transcription Factor 4 Regulates Intrinsic Excitability of Prefrontal Neurons via Repression of SCN10a and KCNQ1. **Neuron**. 2016 Apr 6;90(1):43-55. doi: 10.1016/j.neuron.2016.02.021. Epub 2016 Mar 10. PubMed PMID: 26971948; PubMed Central PMCID: PMC4824652.
3. smFISH image quantification: In the era of single-cell transcriptomics and spatial transcriptomics, multiplexed single-molecule fluorescent in situ hybridization (smFISH) both validates and complements single-cell genomics. In collaboration with the Martinowich lab at LIBD, I helped to develop algorithms for quantifying multiplexed smFISH imaging datasets from the mouse hypothalamic nucleus. The tool allows for quantifying spatial expression of single RNA transcripts and aggregating transcript expression at single-nucleus resolution for differential spatial expression analysis. Later versions of these algorithms further deal with challenges of applying smFISH to postmortem human brain. Dotdotdot, the package developed from these efforts is a resource for processing smFISH data and a manuscript describing its application is in review.

    a. Maynard KR, Tippani M, Takahashi Y, **Phan BN**, Hyde TM, Jaffe AE, Martinowich K. dotdotdot: an automated approach to quantify multiplex single molecule fluorescent in situ hybridization (smFISH) images in complex tissues. **Nucleic Acids Research**, Volume 48, Issue 11, 19 June 2020, Page e66, https://doi.org/10.1093/nar/gkaa312, PMID: 32383753, PMCID: PMC7293004
    b. Maynard KR, Hobbs JW, **Phan BN**, Gupta A, Rajpurohit S, Williams C, Rajpurohit A, Shin JH, Jaffe AE, Martinowich K. BDNF-TrkB signaling in oxytocin neurons contributes to maternal behavior. **Elife**. 2018 Sep 7;7PubMed PMID: 30192229; PubMed Central PMCID: PMC6135608.
4. Addiction epigenetics: As a PhD student in Dr. Andreas Pfenning's lab at Carnegie Mellon University, I investigated the genetic contributions from cell type- and region-specific open chromatin integrating cross-species epigenomic data and machine learning approaches summarized in co-first author manuscript in review at the Journal of Neuroscience. Our study on cell types and brain regions contributing to heritability of addiction-associated traits suggests that the conserved non-coding regions within cortical excitatory and striatal medium spiny neurons contribute to genetic predisposition for nicotine, alcohol, and cannabis use behaviors. This computational framework can flexibly integrate epigenomic data across species to screen for putative causal variants in a cell type- and tissue-specific manner across numerous complex traits, including a collaboration with the Logan lab on human transcriptomics of opioid use disorder.

a. Srinivasan C*, **Phan BN***, Lawler AJ, Ramamurthy E, Kleyman M, Brown AR, Kaplow IM, Wirthlin ME, Pfenning AR. Addiction-associated genetic variants implicate brain cell type- and region-specific cis-regulatory elements in addiction neurobiology. bioRxiv 2020.09.29.318329; doi: https://doi.org/10.1101/2020.09.29.318329. *In review*
b. Seney ML, Moon-Kim S, Glausier JR, Hildebrand MA, XiangningXue, Zong W, Wang J, Shelton MA, **Phan BN**, Srinivasan C, Pfenning AR, Tseng GC, Lewis DA, Freyberg Z, Logan RW. Transcriptional alterations in opioid use disorder reveal an interplay between neuroinflammation and synaptic remodeling. bioRxiv 2020.09.14.296707; doi: https://doi.org/10.1101/2020.09.14.296707. *In review*

Complete List of Published Work in My Bibliography:
https://www.ncbi.nlm.nih.gov/myncbi/badoi.phan.1/bibliography/public/

D. Additional Information: Research Support and/or Scholastic Performance

Scholastic Performance

Johns Hopkins University **GPA: 3.76**, graduated with General Honors & Departmental Honors
University of Pittsburgh, School of Medicine courses are graded Satisfactory/Unsatisfactory (S/U) during first 2 years. The courses Pediatric Inpatient Medicine and Obstetrics/Gynecology Medicine are graded Satisfactory/High Satisfactory/Honors (S/HS/H)

YEAR	COURSE TITLE	GRADE	YEAR	COURSE TITLE	GRADE
JOHNS HOPKINS UNIVERSITY					
2012	Intro to Programming, Java	S	2014	Proj Design: Pharmacokinetics	A
2012	BME Modeling & Design	S	2014	Prob/Stat for Sci & Eng	A-
2012	Organic Chemistry I	S	2014	Systems & Controls	A
2012	Calculus III	S	2014	Models & Simulations	A
2012	General Physics I	S	2015	Expository Writing	A
2013	Introduction to MATLAB	S	2015	Systems Bioengineering II	B+
2013	Organic Chemistry II	A	2015	Comp Bio/Bioinformatics II	A
2013	Linear Algebra	A+	2015	Proj Design: Pharmacodynamics	A
2013	General Physics II	B+	2015	Stat Mechanics/Thermodynamics	A
2013	BME Design Group	B+	2015	BME Design Group	A
2013	Diff Equations w/ Applications	A	2015	Systems Bioengineering III	B+
2013	Cell & Tissue Eng Lab	A	2015	Computer Vision	B
2013	Biomaterals I	B+	2016	BME Design Group	A-
2013	Molecules & Cells	A-	2016	Data Structures	A-
2014	Systems Bioengineering I	B+	2016	Intro to Genomic Research	A-
UNIVERSITY OF PITTSBURGH, SCHOOL OF MEDICINE					
2017	Patient, Physician & Society 1	S	2018	Introduction to Patient Care 2	S
2017	Basic Science Fundamentals 1	S	2018	Neuroscience/Psychiatry	S
2017	Introduction to Patient Care 1	S	2018	OS2:Body Fluid Homeostasis	S
2017	Evidenced Based Medicine	S	2018	PPS3 Population Health	S
2017	Basic Science Fundamentals 2	S	2018	Introduction to Patient Care 3	S
2017	Basic Science Fundamentals 3	S	2018	GI/Endocrine/Hematology/Skin-Musculoskeletal/Repro & Dvlp	S
2017	Prof Dev II: Methods and Analysis	S	2019	Introduction to Patient Care 4	S
2017	Laboratory Research Rotation	S	2019	Integrated Case Studies	S
2018	PPS2 Behavioral Medicine	S	2019	Pediatric Inpatient Medicine	S
			2019	Obstetrics/Gynecology Medicine	H
CARNEGIE MELLON UNIVERSITY, SCHOOL OF COMPUTER SCIENCE					
2019	Intro to Comp Structural Bio	B	2020	Computational Genomics	A-
2019	Comp Bio Seminars	S	2020	Comp Bio Writing Workshop	S
2019	Machine Learning	B	2020	Cell Systems Modeling	A-
2020	Comp Bio Seminars	S	2020	Ethics for Medical Scientists	S

BIOGRAPHICAL SKETCH			
NAME: Andreas Robert Pfenning			
eRA COMMONS USER NAME (credential, e.g., agency login): APFENNING			
POSITION TITLE: Assistant Professor of Computational Biology			
EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable.)*			
INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date	FIELD OF STUDY
Carnegie Mellon University	B.S.	05/2006	Computer Science
Duke University	Ph.D.	05/2012	Computational Biology
Massachusetts Institute of Technology	Postdoc	12/2015	Computational Biology

A. Personal Statement

My goal is to understand the principles that govern neurological disorders and complex vertebrate behaviors from a genetic and evolutionary perspective. I have a broad base of knowledge in computational biology, computer science, machine learning, neurobiology, genomics, genetics, and epigenetics. My previous research has identified convergent evolution in gene expression patterns associated with vocal learning across birds and mammals and has identified mouse/human conserved immune epigenetic immune signatures associated with Alzheimer's disease.

The expertise necessary to mentor Mr. Phan on this project is broad – involving both a variety of biological disciplines and techniques. At the biological level, my training in studying response to neural activity as well as cortico-basal ganglia circuits has created a foundation for my laboratory's focus on studying the mechanisms of addiction through a NIDA Avenir Award and striatal neuron subtypes through a Brain Initiative UG3. In addition, a consistent theme throughout my career has been the careful application of comparative genomic methods to identify conserved gene regulatory features associated with complex phenotypes (Gjoneska, Pfenning, et. al, 2015). Together, these experiences provide substantial preparation for mentoring Mr. Phan in connecting features of addiction in human studies with animal models. Already, Mr. Phan has been a bridge between my group, whose expertise lies in neurogenomics, with Dr. Logan's group, who have deeper expertise in addiction (Seney et. al. 2020).

A variety of computational and experimental techniques to are required to dissect the cell type-specific basis of addiction behavior. Experimentally, we have published a new technique for isolating neuron subtypes in the striatum (Lawler et. al. 2020) and have used it interpret the genetic basis of human addiction phenotypes (Srinivasan, Phan et. al., 2020). Computationally, we are developing methods as a part of the Zoonomia consortium to study gene regulation across species (Zhang et. al., 2020). By integrating the high-throughput experimental biology with advanced computational techniques, we are able to provide insight into fundamental open in how genetics relates to phenotype.

1. Gjoneska E*, **Pfenning AR***, Mathys M, Quon G, Kundaje A, Tsai L-H, Kellis M. Conserved epigenomic signals in mice and human reveal immune basis of Alzheimer's disease. (2015) Nature. PMID: 25693568
2. Seney M, Kim S-M, Wang J, Hildebrand M, Xue X, Glausier J, Zong W, Shelton M, Phan B, Srinivasan C, **Pfenning AR**, Tseng G, Lewis D, Freyberg Z, Logan R. Transcriptional alterations in opioid use disorder reveal an interplay between neuroinflammation and synaptic remodeling. (2020) bioRxiv.
3. Lawler AJ, Brown AR, Bouchard RS, Toong N, Kim J, Velraj N, Fox G, Kleyman M, Kang B, Gittis AH, **Pfenning AR**. Cell type-specific oxidative stress genomic signatures in the globus pallidus of dopamine depleted mice. (2020). J. Neurosci. PMID: 33188066
4. Srinivasan C*, Phan B*, Lawler AJ, Ramamurthy E, Kleyman M, Brown AR, Kaplow IM, Wirthlin ME, **Pfenning AR**. Addiction-associated genetic variants implicate brain cell type-and region-specific cis-regulatory elements in addiction neurobiology. (2020) bioRxiv.
5. Zhang X*, Kaplow IM*, Wirthlin M, Park TY, **Pfenning AR**. HALPER facilitates the identification of regulatory element orthologs across species. Bioinformatics. (2020) PMID: 32407523

B. Positions and Honors

Positions Employment

2016- Assistant Professor, Department of Computational Biology, School of Computer Science,

Carnegie Mellon University, Pittsburgh, PA

Other Experience and Professional Memberships

2017- Member, Vertebrate Genome Project Consortium
2017- Member, Society for Molecular Biology and Evolution
2012- Member, AAAS
2006- Member, Society for Neuroscience

Honors

2018 Avenir Award for Genetics or Epigenetics of Substance Abuse (NIDA, NIH)
2018 AFAR Junior Faculty Fellow in Aging Research
2018 Sloan Research Fellowship for Computational & Evolutionary Molecular Biology
2016 Okawa Foundation Research Grant in Bioinformatics
2014 Kavli Foundation Blog – Biggest Science Stories of 2014
2013 Best Postdoc Poster Presentation, Harvard Med. Genetics Retreat
2013 ISMB/ECCB Conference Program Committee
2010 Best Poster Presentation, Duke Computational Biology
2006 Richard Schoenwald Phi Beta Kappa Research Award

C. Contributions to Science

Full list of papers at:
http://www.ncbi.nlm.nih.gov/sites/myncbi/1zmzIHeEPeiQl/bibliography/45315448/public/?sort=date&direction=ascending.

Contribution 1: Genetic and epigenetic basis of neurodegeneration and aging

The increasing sizes of genome-wide association studies are leading to the identification of an increasing number of loci affecting predisposition to neurological and psychiatric disorders. However, their interpretation remains a challenge, as greater than 80% of causal mutations are predicted to affect non-coding regulatory regions. As a part of the epigenome roadmap project, I showed that Alzheimer's disease loci are enriched CD14+ immune cell enhancers (defined by ChIP-Seq for histone modifications) compared to other cell types and neural tissues. Parallel epigenomic experiments in mouse showed thousands of enhancers, mostly near immune genes, were increasing in the histone mark H3K27ac. Strikingly, the human regions orthologous to those increasing enhancers were strongly enriched for known threshold ($p < 10^{-6}$) and sub-threshold ($p < 10^{-3}$) Alzheimer's disease loci, suggesting the conserved gene expression patterns are mediated by conserved (and causal) gene regulatory mechanisms at immune enhancers. Still, a major challenge facing the field is interpreting cell type-specific mechanisms underlying neurological disorders. Even a small piece of brain tissue can contain dozens of molecularly distinct cell types, each of which could play different roles in disease predisposition and progression. We have developed a new genomic technology, SNAIL, to isolate neuron subtypes without the use of double transgenic mice. We applied that technology to identify cell type-specific HIF signaling underlying neurodegeneration in a mouse model of Parkinson's disease.

1. Lawler AJ, Brown AR, Bouchard RS, Toong N, Kim J, Velraj N, Fox G, Kleyman M, Kang B, Gittis AH, **Pfenning AR**. Cell type-specific oxidative stress genomic signatures in the globus pallidus of dopamine depleted mice. (2020). J. Neurosci
2. Glorioso CA*, **Pfenning AR***, Lee SS, Bennett DA, Sibille EL, Kellis M, Guarente LP. Rate of brain aging and APOE ε4 are synergistic risk factors for Alzheimer's disease. (2019) Life Science Alliance. PMID: 31133613.
3. Klein HU, McCabe C, Gjoneska E, Sullivan SE, Kaskow BJ, Tang A, Smith RV, Xu J, **Pfenning AR**, Bernstein BE, Meissner A, Schneider JA, Mostafavi S, Tsai LH, Young-Pearse TL, Bennett DA, De Jager PL. Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and Alzheimer's human brains. (2019) Nat Neurosci. PMID: 30559478.
4. Gjoneska E*, **Pfenning AR***, Mathys M, Quon G, Kundaje A, Tsai L-H, Kellis M. Conserved epigenomic signals in mice and human reveal immune basis of Alzheimer's disease. (2015) Nature. PMID: 25693568

Contribution 2: Convergent evolution of vocal learning behavior and complex traits

A central finding of my PhD thesis was that gene expression patterns associated with vocal learning behavior, in the brain regions that control the behavior, have undergone convergent evolution in birds and mammals.

Vocal learning, or the ability to mimic sounds, is a necessary component of human speech production and has undergone convergent evolution, having evolved independently in different avian lineages as well as mammals (including humans in the form of speech). I developed a dynamic programming algorithm that was able to find molecular similarities between songbirds and humans in the brain regions that control vocal learning. For one of these brain region pairs (songbird RA and human primary motor cortex) I was able to show that these molecular signatures were stronger in vocal species (songbirds, parrots, hummingbirds and human) compared to vocal non-learning species (dove, quail, and macaque), thus showing convergent evolution in gene expression patterns. This model implies that genetic differences across species at enhancer regions are likely to be underlying complex traits like vocal learning. A central goal of my laboratory is to develop resources and computational tools to predict these convergent regulatory events. We have done so by helping to develop the Zoonomia project, a collection of 252 aligned mammalian genomes. We also assisted in applying those resources to studying the evolutionary basis of COVID-19 susceptibility.

1. Zhang X*, Kaplow IM*, Wirthlin M, Park TY, **Pfenning AR**. HALPER facilitates the identification of regulatory element orthologs across species. Bioinformatics. (2020) PMID: 32407523
2. Genereux DP, Serres A, Armstrong J, Johnson J, Marinescu VD, Murén E, Juan D, Bejerano G, Casewell NR, Chemnick LG, Damas J, Di Palma F, Diekhans M, Fiddes IT, Garber M, Gladyshev VN, Goodman L, Haerty W, Houck ML, Hubley R, Kivioja T, Koepfli K-P, Kuderna LFK, Lander ES, Meadows JRS, Murphy WJ, Nash W, Noh HJ, Nweeia M, **Pfenning AR**, Pollard KS, Ray D, Shapiro B, Smit A, Springer M, Steiner CC, Swofford R, Taipale J, Teeling EC, Turner-Maier J, Alfoldi J, Birren B, Ryder OA, Lewin H, Paten B, Marques-Bonet T, Lindblad-Toh K, Karlsson EK. A comparative genomics multitool for scientific discovery and conservation. (2020). Nature. PMID: 33177664
3. Damas J, Hughes GM, Keough KC, Painter CA, Persky NS, Corbo M, Hiller M, Koepfli K-P, **Pfenning AR**, Zhao H, Genereux DP, Swofford R, Pollard KS, Ryder OA, Nweeia MT, Lindblad-Toh K, Teeling EC, Karlsson EK, Lewin HA. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. (2020) PNAS. PMID: 33060197
4. Wirthlin M, Chang EF, Knörnschild M, Krubitzer LA, Mello CV, Miller CT, **Pfenning AR**, Vernes SC, Tchernichovski, Yartsev MM. A Modular Approach to Vocal Learning: Disentangling the Diversity of a Complex Behavioral Trait. (2019) Neuron. PMID: 31600518
5. **Pfenning, AR**, Hara E, Whitney O, Rivas M, Roulhac P, Ganapathy G, Hartemink AJ, Jarvis ED. Convergent transcriptional specializations in the brains of humans and song learning birds. (2014) Science. PMID: 25504733

Contribution 3: Molecular mechanisms of the transcriptional response to neural activity

A central goal in neurobiology, and a focus of my research for the past decade, is to understand how genes are regulated in response to neural activity. A stimulus of neural activity causes large-scale changes in gene expression in the brain, as is required for neural plasticity and as a key component of learning and memory. This work has taught me that the tools of epigenomics can provide the key insights to understand this process. I begin this research using in silico approaches, ChIP-Seq for the neural transcription factor CaRF, and, as a part of my dissertation, a microarray analysis looking at genes that respond to neural activity during song production in the zebra finch. Surprisingly, we found that very different patterns of gene expression were activated across different brain regions. Furthermore, the genes that were induced these in these brain regions were also likely to have higher expression before the bird began to sing. Based on this observation, I hypothesized that the differences in gene expression across brain regions could be due to the basal epigenetic state of the brain region. As a postdoc, I had the opportunity to finally test this hypothesis, and demonstrated that the differences in activated gene expression across brain regions were, in fact, associated with differences in epigenetic state (measured by H3K27ac) of those brain regions before the bird began to produce song. To further study the epigenomics of activity regulated genes I collaborated with the lab of Prof. Li-Huei Tsai at MIT, and participated in the analysis of ChIP-Seq data to show that double-stranded DNA breaks occur in response to neural activity at enhancers with high H3K27ac, high H3K4me1, bind CBP and contain the CTCF insulator element. In parallel I begun a computational and experimental collaboration with Prof. Jesse Gray at Harvard Medical School to develop a high-throughput way to measure how genetic variation impacts how enhancers response to activity. My design and analysis of a reporter assay helped to determine sequence differences between sequences acting as promoters compared to enhancers during the response to neural activity.

1. **Pfenning AR**, Kim TK, Spotts JM, Hemberg M, Su D, West AE. Genome-wide identification of calcium-response factor (CaRF) binding sites predicts a role in regulation of neuronal signaling pathways. (2010) PLoS One. PMID: 20523734

2. Whitney O*, **Pfenning AR***, Howard JT, Blatti CA, Lie F, Ward JM, Wang R, Kellis M, Mukherjee S, Sinha S, Hartemink AJ, West AE, Jarvis ED. Core and region enriched networks of behaviorally regulated genes and the singing genome. (2014) Science. PMID:
3. Madabhushi R, Gao F, **Pfenning AR**, Pan L, Yamakawa S, Seo J, Rueda R, Phan T, Pao P-C, Stott RT, Gjoneska E, Nott A, Cho S, Kellis M, Tsai L-H. Programmed DNA double strand breaks govern the expression of early response genes in neurons. (2015) Cell. PMID: 26052046
4. Nguyen TA, Jones RD, Snavely A, **Pfenning AR**, Kirchner R, Hemberg M, Gray JM. High-throughput functional comparison of promoter and enhancer activities. (2016) Genome Research. PMID: 27311442

Contribution 4: Genetic and epigenetic basis of addiction

Recent large genome-wide association studies (GWAS) have identified multiple confident risk loci linked to addiction-associated behavioral traits. Genetic variants linked to addiction-associated traits lie largely in non-coding regions of the genome, likely disrupting the function of cis-regulatory elements (CREs) to regulate the expression of genes. These CREs tend to be highly cell type-specific and can even be context-specific, which creates barriers to their characterization, but also provides an opportunity to link genetic variation to specific cell types, neural circuits, and biological processes. Thus, to understand the genetics of addiction, the Pfenning laboratory is studying CREs and the genes they regulate in the striatum and cortex. To build profiles of relevant CRE, we used a technology we developed in the laboratory, cSNAIL, to isolate relevant neuron subtypes from the cortex, dorsal striatum, and ventral striatum of mouse (Lawler et. al., 2020). After mapping these regulatory elements to human, we found that different addiction-associated traits were linked to different medium spiny neuron (D1 and D2) and interneuron (PVALB and SST) subtypes (Srinivasan, Phan, et. al., 2020). Moreover, we have begun applying convolutional neural network models to make predictions about how these genetic variants will impact cell type-specific gene expression (Srinivasan, Phan et. al., 2020). To move beyond these broader populations of cells and shift towards systems closer to human, we are collaborating with the Stauffer laboratory at the University of Pittsburgh to build the most comprehensive characterization of striatum medium spiny neuron subtypes described to date, validating populations associated with the olfactory tubercle, islands of Calleja, D1/D2 hybrid cells, core vs. shell, and patch vs. matrix (He e. al., 2020). Finally, in a collaboration with Ryan Logan's laboratory at the University of Pittsburgh, identified an enrichment of addiction-associated loci near genes in response to opioid use disorder, suggesting that the genes that respond to the disorder may themselves be mediated the predisposition to addiction behavior (Seney et al, 2020). These efforts build upon earlier efforts of mine to link pathways studied in the context of addiction to genes induced by salt craving (. Overall, the techniques that we have developed to experimentally isolate neurons subtypes, computationally predict the impact of genetic variation, and to map relevant results across species provide the foundation to systematically connect candidate addiction-associated genetic variants to neural circuits and behavior.

6. Lawler AJ, Brown AR, Bouchard RS, Toong N, Kim J, Velraj N, Fox G, Kleyman M, Kang B, Gittis AH, **Pfenning AR**. Cell type-specific oxidative stress genomic signatures in the globus pallidus of dopamine depleted mice. (2020). J. Neurosci. PMID: 33188066
7. Srinivasan C*, Phan B*, Lawler AJ, Ramamurthy E, Kleyman M, Brown AR, Kaplow IM, Wirthlin ME, **Pfenning AR**. Addiction-associated genetic variants implicate brain cell type-and region-specific cis-regulatory elements in addiction neurobiology. (2020) bioRxiv.
8. He J, Kleyman M, Chen J, Alikaya A, Rothenhoefer KM, Ozturk BE, Wirthlin ME, Fish K, Byrne LCT, **Pfenning AR**, Stauffer WR. Transcriptional Diversity of Medium Spiny Neurons in the Primate Striatum. (2020) bioRxiv.
9. Seney M, Kim S-M, Wang J, Hildebrand M, Xue X, Glausier J, Zong W, Shelton M, Phan B, Srinivasan C, **Pfenning AR**, Tseng G, Lewis D, Freyberg Z, Logan R. Transcriptional alterations in opioid use disorder reveal an interplay between neuroinflammation and synaptic remodeling. (2020) bioRxiv.
10. Liedtke W, McKinley MJ, Walker LL*, Zhang H*, **Pfenning AR***, Drago J, Hochendoner SJ, Hilton DL, Lawrence AJ, Denton DA. Relation of addiction genes to hypothalamic gene changes subserving genesis and gratification of a classic instinct, sodium appetite. (2011) Proc Natl Acad Sci. PMID: 21746918

D. Research Support

Ongoing Research Support

1DP1DA046585 (Pfenning) 08/15/2018-06/30/2023 3.00 summer
NIH $300,000 DC
Interpreting the gene regulatory mechanisms underlying the predisposition to addictions disorders

The major goal of this project are to build a framework to study the function of both human and mouse brain enhancer regions in vivo to work towards deciphering biological mechanism underlying substance use disorders.

UG3-MH-120094 (Stauffer) 09/01/2019-06/30/2022 2.00 calendar
University of Pittsburgh/NIH $107,589 DC (Pfenning lab sub amount)
A massive library of AAVs to target transcriptionally-defined primate cell types
The major goal of this project is to develop tools to target specific cell types in the primate brain and retina.

A022258 (Pfenning) 07/01/2018-06/30/2021 0.25 calendar
American Federation for Aging Research $22,454 DC
Cell type-specific epigenetic decay underlying brain aging
The major goal of this project is to test the hypothesis that a process of “epigenetic decay”, where cell type-specific features are lost, is underlying aging in the brain

A019625 (Pfenning) 12/01/2016-12/31/2020 1.50 calendar
Alzheimer’s Disease Research Foundation $199,013 DC
Interpreting Alzheimer’s disease-associated genetic variation at enhancer regions
The major goal of this project is to make progress towards a cure for Alzheimer’s disease using genomic approaches.

Completed Research Support

2016 (Pfenning) 12/01/2016 – 12/01/2017 1 calendar
The Okawa Foundation $10,000 (no salary)
Interpreting Alzheimer’s-associated genetic variation using genomic analysis
The goal of the project is to use computational methods on genomic data to infer the pathways underlying Alzheimer's disease predisposition.
Status: We have identified that Alzheimer’s associated mutations are significantly enriched in enhancers that are specific to microglia.

2018 (Pfenning) 8/15/2018-8/15/2019 1 calendar
Whitehall Foundation $75,000 DC/year
Gene regulatory mechanisms underlying motor learning
The goal of the project is to identify the molecular genetic basis of motor learning behavior. To accomplish that goal, we will track changes in gene expression and open chromatin as mice learn a joystick control task.

2018 (Pfenning) 9/15/2018-9/15/2020 2 calendar
Alfred P. Sloan Foundation $32,500 DC/year
Cell type-specific epigenetic decay underlying brain aging
We are using computational techniques to trace the evolution and tissue-specificity of regulatory elements across vertebrates.

OMB No. 0925-0001 and 0925-0002 (Rev. 03/2020 Approved Through 02/28/2023)

BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. DO NOT EXCEED FIVE PAGES.

NAME: Logan, Ryan, W.

eRA COMMONS USER NAME (credential, e.g., agency login): rwlogan

POSITION TITLE: Graduate Faculty Mentor

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE (if applicable)	END DATE MM/YYYY	FIELD OF STUDY
University of Maine	BA	05/2004	Psychology
University of Maine	PHD	12/2009	Neuroscience
Rutgers University	Postdoctoral Fellow	10/2011	Neuroscience & Immunology
The Jackson Laboratory	Postdoctoral Fellow	06/2012	Computational Biology
University of Pittsburgh	Postdoctoral Fellow	06/2015	Molecular Neuroscience

A. Personal Statement

The overall goal of the Logan laboratory is to understand the mechanisms underlying the vulnerability to addiction and substance use disorders. Our laboratory is focused particularly on the role of sleep and circadian rhythms in reward, motivation, and dependence and relapse. Almost every cell in the body and brain expresses the molecular machinery necessary for cellular and molecular rhythms. The machinery is comprised of genes and proteins that form a series of transcriptional—translational feedback loops that modulate a wide array of cellular and molecular functions depending on the tissue and cell-type. We found that circadian genes and proteins that form the molecular clock in striatal and cortical neural circuits are critical in the modulation of reward and motivated behaviors, especially in their control in the behavioral response to cocaine and opioids. Currently, our research is using human postmortem brain tissues from subjects with opioid use disorder to further understand cellular and molecular mechanisms of human opioid dependence, then use these discoveries in humans to dissect translationally relevant mechanisms in animal models. Our current projects related to this proposal investigate: 1) Circadian regulation of transcriptional programs within dopamine receptor 1 and 2 neurons of the ventral striatum and their potential role in the behavioral responses to psychostimulants and opiates; 2) Genetic mechanisms underlying the relationship between circadian rhythm and addiction-related behaviors using the genetically heterogenous Diversity Outbred and Collaborative Cross mouse populations. Our goal is to discover genetic variants contributing to addiction vulnerability and the genes (or variants) possible role in specific cell-types of neural substrates of drug addiction.

1. Seney, M.L., Moon-Kim,S., Glausier, J.R., Hildebrand, M.A., Xue, X., Zong, W., Wang, J., Shelton, M.A., **Phan, B.N.**, Srinivasan, C., **Pfenning, A.R.**, Tseng, G.C., Lewis, D.A., Freyberg, Z., **Logan, R.W.** (2020). Transcriptional alterations in opioid use disorder reveal an interplay between neuroinflammation and synaptic remodeling. bioRxiv 2020.09.14.296707; doi: https://doi.org/10.1101/2020.09.14.296707

2. **Logan RW**, Parekh PK, Kaplan GN, Becker-Krail DD, Williams WP 3rd, Yamaguchi S, Yoshino J, Shelton MA, Zhu X, Zhang H, Waplinger S, Fitzgerald E, Oliver-Smith J, Sundarvelu P, Enwright JF 3rd, Huang YH, McClung CA. NAD+ cellular redox and SIRT1 regulate the diurnal rhythms of tyrosine hydroxylase and conditioned cocaine reward. Mol Psychiatry. 2019 Nov;24(11):1668-1684. PubMed PMID: 29728703; PubMed Central PMCID: PMC6215755.

3. Cahill KM, Huo Z, Tseng GC, **Logan RW***, Seney ML*. Improved identification of concordant and discordant gene expression signatures using an updated rank-rank hypergeometric overlap approach. Sci Rep. 2018 Jun 25;8(1):9588. PubMed PMID: 29942049; PubMed Central PMCID: PMC6018631.

4. **Logan RW**, Robledo RF, Recla JM, Philip VM, Bubier JA, Jay JJ, Harwood C, Wilcox T, Gatti DM, Bult CJ, Churchill GA, Chesler EJ. High-precision genetic mapping of behavioral traits in the diversity outbred mouse population. Genes Brain Behav. 2013 Jun;12(4):424-37. PubMed PMID: 23433259; PubMed Central PMCID: PMC3709837.

B. Positions and Honors

Positions and Employment

2004 - 2006 Graduate Research Assistant, University of Maine, Dept. of Psychology
2007 - 2009 Graduate Research Assistant, University of Maine, Graduate School of Biomedical Sciences
2009 - 2011 Postdoctoral Associate, Rutgers University, Dept. of Animal Sciences
2011 - 2012 Postdoctoral Associate, The Jackson Laboratory
2012 - 2015 Postdoctoral Scholar, University of Pittsburgh, Dept. of Psychiatry, Translational Neuroscience Program
2015 - Assistant Professor, University of Pittsburgh, Dept. of Psychiatry, Translational Neuroscience Program
2016 - Graduate Student Training Faculty, Center for Neuroscience University of Pittsburgh
2018 - Graduate Faculty Mentor, University of Pittsburgh School of Medicine

Other Experience and Professional Memberships

2005 - Member, Research Society on Alcoholism
2005 - Member, Society for Neuroscience
2011 - Member, International Behavioral and Neural Genetics Society
2011 - Member, Society for Research on Biological Rhythms
2015 - Member, Society of Biological Psychiatry
2018 - ZDA1 SXM-M 16 S, NIDA Cutting-Edge Basic Research Awards (CEBRA) Special Emphasis Panel Study Section
2018 - ZDA1 SXM-M 09, NIDA CEBRA Special Emphasis Panel Study Section
2019 - ZDA1 IXR-Q 07 S, NIDA U01 for Identification of Genetic and Genomic Variants by Next-Gen Sequencing in Non-human Animal Models Study Section
2019 - ZDA1 SXM-M 22 S, NIDA CEBRA Special Emphasis Panel Study Section

Honors

2006 Travel Award, ISBRA/RSA Conference, International Society for Biomedical Research on Alcoholism
2010 Presidential Research Travel Award, Society for Neuroscience
2011 Junior Investigator Travel Award, Research Society on Alcoholism
2011 Loan Repayment Program, National Institute on Alcohol Abuse and Alcoholism
2011 HHMI Course Scholarship, Genetics of Addiction, The Jackson Laboratory
2012 Junior Investigator Travel Award, Research Society on Alcoholism
2012 Training Position, NIDA T32, University of Pittsburgh
2012 Presenter Travel Award, International Behavioural and Genetics Society
2012 HHMI Course Scholarship, Genetics of Addiction, The Jackson Laboratory
2014 Training Position, NINDS T32, University of Pittsburgh
2014 NARSAD Young Investigator Award, Brain and Behavior Research Foundation
2014 Career Development Institute for Psychiatry Fellow, University of Pittsburgh / Stanford University
2015 Chairman's Choice Travel Award, Society of Biological Psychiatry
2016 Outstanding Junior Faculty Travel Award, International Behavioral and Neural Genetics Society
2016 Young Investigator Travel Award, Molecular Psychiatry Association
2016 Loan Repayment Program, National Institute on Drug Abuse
2018 ACNP Conference Travel Award - Senior Level Selection, American College of

	Neuropsychopharmacology
2018	Hamilton Family Prize for Basic Neuroscience Research in Psychiatry, University of Pittsburgh School of Medicine
2018	Loan Repayment Program Renewal, National Institute on Drug Abuse

C. Contribution to Science

1. Emerging evidence from human studies suggests disruptions to the circadian system are associated with mood and addiction disorders. However, the mechanisms underlying these relationships are less clear. As such, we were the first to demonstrate chronic drug use and withdrawal has consequences core features of the circadian system (i.e., period, phase, and amplitude). We also discovered certain circadian rhythm phenotypes (e.g., period) are co-inherited, presumably via common genetic factors, with addiction-related phenotypes, such as alcohol drinking and cocaine reward. These initial findings provided the foundation for studies probing the genetic and molecular mechanisms by which circadian rhythms modulate drug reward, seeking, and motivation. Our recent contribution demonstrates molecular rhythms and bioenergetic pathways directly control dopamine neurotransmission and the behavioral response to cocaine. We continue to investigate the cellular, molecular, and genetic mechanisms linking rhythms to the vulnerability and progression of substance abuse and dependence, primarily focusing psychostimulants (e.g., cocaine) and more recently, opioids (i.e., fentanyl).

    a. Parekh PK, **Logan RW**, Ketchesin KD, Becker-Krail D, Shelton MA, Hildebrand MA, Barko K, Huang YH, McClung CA. Cell-Type-Specific Regulation of Nucleus Accumbens Synaptic Plasticity and Cocaine Reward Sensitivity by the Circadian Protein, NPAS2. J Neurosci. 2019 Jun 12;39(24):4657-4667. PubMed PMID: 30962277; PubMed Central PMCID: PMC6561687.
    b. Freyberg Z, **Logan RW**. The Intertwined Roles of Circadian Rhythms and Neuronal Metabolism Fueling Drug Reward and Addiction. Curr Opin Physiol. 2018 Oct;5:80-89. PubMed PMID: 30631826; PubMed Central PMCID: PMC6322667.
    c. **Logan RW**, Parekh PK, Kaplan GN, Becker-Krail DD, Williams WP 3rd, Yamaguchi S, Yoshino J, Shelton MA, Zhu X, Zhang H, Waplinger S, Fitzgerald E, Oliver-Smith J, Sundarvelu P, Enwright JF 3rd, Huang YH, McClung CA. NAD+ cellular redox and SIRT1 regulate the diurnal rhythms of tyrosine hydroxylase and conditioned cocaine reward. Mol Psychiatry. 2019 Nov;24(11):1668-1684. PubMed PMID: 29728703; PubMed Central PMCID: PMC6215755.
    d. **Logan RW**, McCulley WD 3rd, Seggio JA, Rosenwasser AM. Effects of withdrawal from chronic intermittent ethanol vapor on the level and circadian periodicity of running-wheel activity in C57BL/6J and C3H/HeJ mice. Alcohol Clin Exp Res. 2012 Mar;36(3):467-76. PubMed PMID: 22013893; PubMed Central PMCID: PMC3266959.

2. Findings from human and animals directly links circadian rhythms to immune function and various diseases, including several types of cancer. Over several decades, human epidemiological studies have reported chronic shift-workers (e.g., rotating work schedules) are at a much higher risk for breast, ovarian, and lung cancers. The innate immune system, our first line of defense against pathogens, is critical for attacking and killing pre-cancerous and cancerous cells. A primary effector of the innate immune system involves the activation of natural killer cells. Through a series of animal studies, we demonstrated the circadian timing system temporally regulates the cytotoxic ability of natural killer cells in response to tumor cells. We also found disruption of the circadian system, both at the systemic and molecular levels, impairs natural killer cell function, reducing cytotoxicity of invading tumor cells, and leads to significant adenocarcinomas (i.e., lung cancer). We are actively extending this work by using ontogenetic tools to directly modulate the central circadian pacemaker of the suprachiasmatic nucleus to modulate systemic circadian timing in an effort to further investigate the ability of the circadian timing system to control innate immune function. These studies will be important for the development of "chronotherapeutics" for cancer and other diseases.

    a. **Logan RW**, Wynne O, Maglakelidze G, Zhang C, O'Connell S, Boyadjieva NI, Sarkar DK. β-Endorphin neuronal transplantation into the hypothalamus alters anxiety-like behaviors in prenatal alcohol-exposed rats and alcohol-non-preferring and alcohol-preferring rats. Alcohol Clin Exp Res. 2015 Jan;39(1):146-57. PubMed PMID: 25623413; PubMed Central PMCID: PMC4521638.

b. **Logan RW**, Zhang C, Murugan S, O'Connell S, Levitt D, Rosenwasser AM, Sarkar DK. Chronic shift-lag alters the circadian clock of NK cells and promotes lung cancer growth in rats. J Immunol. 2012 Mar 15;188(6):2583-91. PubMed PMID: 22308312; PubMed Central PMCID: PMC3294088.
c. **Logan RW**, Sarkar DK. Circadian nature of immune function. Mol Cell Endocrinol. 2012 Feb 5;349(1):82-90. PubMed PMID: 21784128.
d. **Logan RW**, Arjona A, Sarkar DK. Role of sympathetic nervous system in the entrainment of circadian natural-killer cell function. Brain Behav Immun. 2011 Jan;25(1):101-9. PubMed PMID: 20816749; PubMed Central PMCID: PMC2991610.

3. Another component of the laboratory, in collaboration with Drs. Marianne Seney (Psychiatry) George Tseng (Computational Biology), is focused on developing bioinformatics tools for analyzing large-scale gene expression in relation to mood and addiction disorders across species (e.g., mouse and human). We are developing new 'unbiased' approaches to discover and prioritize molecular pathways related to various disease states in mouse and human brains. We recently applied these tools to human post-mortem brain tissues in collaboration with Dr. Seney and mouse brain tissues in collaboration with Dr. Eric Nestler. As we are also integrating transcriptomics and proteomics approaches in our own studies, we will continue to adapt and develop new computational tools for probing brain and behavior relationships related to psychiatric disorders.

a. Seney ML, Cahill K, Enwright JF 3rd, **Logan RW**, Huo Z, Zong W, Tseng G, McClung CA. Diurnal rhythms in gene expression in the prefrontal cortex in schizophrenia. Nat Commun. 2019 Aug 9;10(1):3355. PubMed PMID: 31399567; PubMed Central PMCID: PMC6689017.
b. Walker DM, Cates HM, Loh YE, Purushothaman I, Ramakrishnan A, Cahill KM, Lardner CK, Godino A, Kronman HG, Rabkin J, Lorsch ZS, Mews P, Doyle MA, Feng J, Labonté B, Koo JW, Bagot RC, **Logan RW**, Seney ML, Calipari ES, Shen L, Nestler EJ. Cocaine Self-administration Alters Transcriptome-wide Responses in the Brain's Reward Circuitry. Biol Psychiatry. 2018 Dec 15;84(12):867-880. PubMed PMID: 29861096; PubMed Central PMCID: PMC6202276.
c. Seney ML, Huo Z, Cahill K, French L, Puralewski R, Zhang J, **Logan RW**, Tseng G, Lewis DA, Sibille E. Opposite Molecular Signatures of Depression in Men and Women. Biol Psychiatry. 2018 Jul 1;84(1):18-27. PubMed PMID: 29548746; PubMed Central PMCID: PMC6014892.
d. Cahill KM, Huo Z, Tseng GC, **Logan RW***, Seney ML*. Improved identification of concordant and discordant gene expression signatures using an updated rank-rank hypergeometric overlap approach. Sci Rep. 2018 Jun 25;8(1):9588. PubMed PMID: 29942049; PubMed Central PMCID: PMC6018631.

Complete List of Published Work in My Bibliography:
https://www.ncbi.nlm.nih.gov/myncbi/ryan.logan.1/bibliography/public/

D. Additional Information: Research Support and/or Scholastic Performance

Ongoing Research Support

5 P50 DA039841, National Institute on Drug Abuse
McClung (PI)
10/01/16-09/01/21
Center for Systems Neurogenetics of Addiction
To investigate the genetics underlying circadian rhythms, behavioral impulsivity, and cocaine using novel genetically diverse mouse populations.
Role: Co-Investigator

5 R01 AA025626, National Institute on Alcohol Abuse and Alcoholism
Hasler (PI)
09/10/17-05/31/22
Proximal Prospective Associations between Circadian Alignment, Reward Function and Alcohol Use in Adolescents
To establish the extent of proximal prospective associations between sleep/circadian factors, reward function, and alcohol use.
Role: Co-Investigator

4 R33 DA041872, National Institute on Drug Abuse
Logan (PI)
04/01/18-03/31/21
Generating novel mouse tools to investigate brain region and cell-type specific circadian molecular mechanisms of reward and motivation
To further clarify the role of the molecular clock in the transition to addiction and provide the broader scientific community with novel transgenic mice to further investigate the molecular mechanisms of direct and indirect pathway regulation.
Role: PI

1 R01 HL150432, National Heart, Lung, and Blood Institute
Logan (PI)
09/23/19-08/31/22
Cell-type specific role of circadian-dependent transcription in fentanyl-induced synaptic and behavioral plasticity
To investigate the role of cell-type specific molecular rhythms in the striatum in opioid self-administration and opioid-induced sleep disruptions
Role: PI

1 R01 DA051390, National Institute of Drug Abuse
Logan (PI)
01/01/20-01/01/24
Molecular rhythm alterations in human postmortem brain associated with opioid use disorder
To investigate transcriptional rhythm changes in corticostriatal circuits of human postmortem brains from subjects with opioid use disorder and use pathway-specific viral targeting to disrupt rhythms in these circuits in mice to determine their role in opioid dependence
Role: PI

Completed Research Support

R21 DA041872, National Institute on Drug Abuse
Logan (PI)
04/01/16-03/31/18
Generating novel mouse tools to investigate brain region and cell-type specific circadian molecular mechanisms of reward and motivation
To optimize CRISPR for large genomic knockin strategies to generate novel mouse models for the study of addiction-related behaviors.
Role: PI

Young Investigator Award [NCE 01/2017], NARSAD, Brain and Behavior Research Foundation
Logan (PI)
01/15/14-01/14/16
Identifying neurotherapeutic epigenetic targets for the treatment of bipolar disorder
Investigate the epigenetic mechanisms of valproic acid and their circadian transcriptional targets in an animal model of bipolar mania.
Role: PI

K01 DA038654, National Institute on Drug Abuse
Logan (PI)
02/01/15-05/31/19
The role of the circadian transcription factor NPAS2 in the nucleus accumbens to regulate cocaine reward
To investigate the role of NPAS2 to modulate cocaine reward behavior via the interaction with immune-related molecular pathways in the striatum.
Role: PI

PHS Fellowship Supplemental Form

OMB Number: 0925-0001
Expiration Date: 02/28/2023

Introduction

1. Introduction to Application (or Resubmission applications)	Introduction_resub_final.2.pdf

Fellowship Applicant Section

2. Applicant's Background and Goals for Fellowship Training*	Background-Goals-of-Fellowship_resub_final.pdf

Research Training Plan Section

3. Specific Aims*	Phan_Specific Aims_resub_final.3.pdf
4. Research Strategy*	Phan_Research Strategy_resub_final.3.pdf
5. Respective Contributions*	Respective-Contributions_resub_final.pdf
6. Selection of Sponsor and Institution*	Selection-of-Sponsor-and-Institution_resub_final.pdf
7. Progress Report Publication List (or Renewal applications)	
8. Training in the Responsible Conduct of Research*	Responsible-Conduct-of-Research_resub_final.pdf

Sponsor(s), Collaborator(s) and Consultant(s) Section

9. Sponsor and Co-Sponsor Statements	Sponsor_and_Co-sponsor_Statement_resub_final.pdf
10. Letters of Support from Collaborators, Contributors and Consultants	Letters_of_Support.pdf

Institutional Environment and Commitment to Training Section

11. Description of Institutional Environment and Commitment to Training	Additional-Educational-Info_resub_final.pdf
12. Description of Candidate's Contribution to Program Goals	

Other Research Training Plan Section

Vertebrate Animals

The following item is taken from the Research & Related Other Project Information form and repeated here for your reference. Any change to this item must be made on the Research & Related Other Project Information form.

Are Vertebrate Animals Used? ☑ Yes ☐ No

13. Are vertebrate animals euthanized?	☑ Yes	☐ No
If "Yes" to euthanasia Is method consistent with American Veterinary Medical Association (AVMA) guidelines?	☑ Yes	☐ No
If "No" to AVMA guidelines, describe method and provide scientific justification		
14. Vertebrate Animals	Vertebrate-Animals_resub_final.pdf	

PHS Fellowship Supplemental Form

Other Research Training Plan Information

15. Select Agent Research

16. Resource Sharing Plan — Resource-Sharing-Plan_resub_final.2.pdf

17. Authentication of Key Biological and/or Chemical Resources

Additional Information Section

18. Human Embryonic Stem Cells

Does the proposed project involve human embryonic stem cells?* ☐ Yes ☑ No

If the proposed project involves human embryonic stem cells, list below the registration number of the specific cell line(s), using the registry information provided within the agency instructions. Or, if a specific stem cell line cannot be referenced at this time, please check the box indicating that one from the registry will be used:

☐ Specific stem cell line cannot be referenced at this time. One from the registry will be used.

Cell Line(s):

19. Alternate Phone Number:

20. Degree Sought During Proposed Award:

Degree:	If "other", indicate degree type:	Expected Completion Date (MM/YYYY):
OTH: Other	MD/PhD	2020-12

21. Field of Training for Current Proposal*: 104 Computational Biology

22. Current or Prior Kirschstein-NRSA Support?* ☑ Yes ☐ No

If yes, identify current and prior Kirschstein-NRSA support below:

Level*	Type*	Start Date (if known)	End Date (if known)	Grant Number (if known)
Predoctoral	Institutional	06/01/2017	06/30/2018	T32 GM008208

23. Applications for Concurrent Support?* ☐ Yes ☑ No

If yes, describe in an attached file:

24. Citizenship*

U.S. Citizen — U.S. Citizen or Non-Citizen National? ☑ Yes ☐ No

Non-U.S. Citizen
☐ With a Permanent U.S. Resident Visa
☐ With a Temporary U.S. Visa

If you are a non-U.S. citizen with a temporary visa applying for an award that requires permanent residency status, and expect to be granted a permanent resident visa by the start date of the award, check here: ☐

25. Change of Sponsoring Institution — Name of Former Institution:*

PHS Fellowship Supplemental Form

Budget Section

All Fellowship Applicants:

26. Tuition and Fees*:

☐ None Requested ☑ Funds Requested

Year 1	$46,606.00
Year 2	$48,004.00
Year 3	$74,051.00
Year 4	$77,754.00
Year 5	
Year 6 (when applicable)	
Total Funds Requested:	$246,415.00

Senior Fellowship Applicants Only:

	Amount	Academic Period	Number of Months
27. Present Institutional Base Salary:			

28. Stipends/Salary During First Year of Proposed Fellowship:

a. Federal Stipend Requested:	Amount	Number of Months
b. Supplementation from Other Sources:	Amount	Number of Months
	Type (e.g.,sabbatical leave, salary)	
	Source	

Appendix

29. Appendix

INTRODUCTION

Thank you for taking the time to read and consider this revised proposal. I also thank the reviewers of my April 2020 application for their comprehensive feedback and comments, which have improved this proposal tremendously, and for their encouraging impact score of 32 (27 ^{h} percentile). This speaks kindly to the confidence reviewers have in dual-degree trainees who often become experts in *both* clinical medicine and basic science research. In response to their general comments, I have amended the text to make it less dense and more readable. Specific changes to the document are discussed below.

Sponsors, Co-sponsors, Consultants: Since the initial proposal, I have submitted a co-first author preprint for review at the *Journal of Neuroscience* and contributed as co-author to manuscripts from both Pfenning and Logan labs. Both Dr. Pfenning and Dr. Logan have additional funding sources and senior author manuscripts, either published or in review. My mentorship team now includes Dr. Kathryn Roeder, Professor of Statistics, serving as the statistical geneticist and senior training faculty member of my thesis committee. Her feedback will be crucial for thorough analysis of substance use disorder (SUD) risk variants as outlined in the **Research Strategy Aim 1c** and **2c**.

Research Training Plan:

Feasibility of Aims and Wet Lab Experiments.

The proposal has scaled down from studying three brain regions to focusing on the nucleus accumbens (NAc). The NAc is highly relevant to addiction, and NAc cell type-specific rodent studies continue to reveal novel roles in addiction. The Pfenning Lab, in collaboration with the Stauffer Lab, has collected snRNA-seq and snATAC-seq of the macaque striatum and report in a preprint two non-canonical D1 MSN populations restricted to the NAc. Our findings in macaque open questions about primate evolution in the NAc and the relevance to human addiction. Together, inhouse and publicly available datasets provide all data required for **Aim 1** and most of the data for **Aim 2**. The experiments proposed in **Aim 2** (snATAC-seq of rat and human NAc) will have N>=3 per sex and species. These changes reduce the experimental load and make both aims achievable within the stated training timeline.

SUD GWAS risk loci and cis-regulatory element (CRE) conservation.

My co-first author preprint suggests that heterogeneity in SU GWAS could reflect different biological contexts where of risk variants might affect gene expression. I demonstrated that orthologous CREs from mouse NAc cell types, along with complementary data and machine learning approaches, can fine-map human SU risk loci to a select a handful of putative causal variants and predict the target cell types, **Research Strategy Figure 9**. These results steer **Aim 2** to focus on human SUD risk variants at CREs that may be shared between rodent and primate orthologous NAc cell types. Recent multiple genome alignments created from >240 deeply sequenced mammalian species by the Zoonomia Consortium, in which the Pfenning Lab participates, enable our published algorithms to identify confident 1-1 orthologous CREs between mammalian species. My preliminary results applying these tools show ~90% of human and macaque CREs have 1-1 primate orthologs, and that accessibility profiles using these CREs can align unlabeled human and macaque cells together correctly, **Research Strategy Figure 7**. Furthermore, I show that a good fraction of conserved CREs identified across human, macaque, and mouse caudate snATAC-seq retain open chromatin activity within orthologous cell types, **Research Strategy Figure 8**.

Elaboration of single cell analyses.

I plan to use common methods and tailored approaches to analyze the snRNA-seq and snATAC-seq data, as detailed in **Research Strategy Aim 1a** and **2a**. These analyses will benchmark the proposed algorithms for multi-species molecular evolution analyses, **Research Strategy Aim 1b** and **2b**. In **Research Strategy Aims 1c** and **2c**, I outline how I will use the popular methods, MAGMA and partitioned LD score regression, to account for LD when linking sets of genes and CREs to human SUD risk variants. I show preliminary results showing that both gene-based and CRE-based partitioned LD score regression approaches can relate human SU risk variants to orthologous macaque or mouse cell types, **Research Strategy Figure 4** and **9**, respectively.

Training Potential: Lastly, I demonstrate my capacity to gain proficiency in snATAC-seq in Aim 2a by successfully performing the analogous snRNA-seq experiments on mouse brain, **Research Strategy, Figure 6**. Expanding my molecular biology skillset to include snATAC-seq will complement the computational focus of this proposal to achieve my goal to become an independent physician scientist.

APPLICANT'S BACKGROUND AND GOALS FOR FELLOWSHIP TRAINING

A. Doctoral Dissertation and Research Experience

I am interested in how genetic variation converges across genes, cells, tissues, and species to produce neuropsychiatric disease and disorders. I envision that beyond basic scientific inquiry, dissecting the genes to their regulatory pathways of the reward system can provide unique opportunities for novel therapies and perspectives to treat substance use disorders and addiction. My experiences have driven me to pursue a career as a physician scientist, being uniquely trained at the intersection of medicine, neuroscience, genomics, and machine learning. Here, I describe a few of my most impactful experiences, each of which was integral in fostering my intellectual curiosity and broadening my understanding of important issues in neuroscience and medicine.

Undergraduate and Post-bac research: Lieber Institute for Brain Development, MD (2014-2017)
Characterization of schizophrenia-associated protein and mRNA variants of ZNF804a.

- **Advisors:** Brady Maher, PhD; Thomas Michael Hyde, MD/PhD (Letters of Reference);
- **Summary:** I was a member of Dr. Brady Maher's lab at the Lieber Institute for Brain Development investigating a novel, truncated transcript of the transcription factor ZNF804a. I cloned the novel transcript and assayed the predicted protein for intracellular localization in HEK293T cells. As a visiting scholar in Dr. Paul Harrison's lab at Oxford University, I assessed the transcriptional and spatial patterns of this novel transcript in postmortem human fetal brain.
- **Skills acquired:** cell culture, molecular biology, radio-labeled DNA probe *in situ* hybridization
- **Achievements and awards:** Johns Hopkins Vredenburg Summer Research Abroad Scholarship

Discovery of neuronal intrinsic excitability and migration regulation by transcription factor 4.

- **Advisors:** Brady Maher, PhD; Andrew Jaffe, PhD (Letters of Reference)
- **Summary:** My next project transitioned to investigating the gene *TCF4* in neurodevelopment. In this space, I analyzed genome-wide binding targets of TCF4 from ChIP-seq of mouse neuroblastoma cell lines and primary cultures of neural progenitor cells. I contributed to creating CRISPR tools to mutate TCF4. I developed algorithms to process and analyze calcium imaging of *ex vivo* whole-brain slice preparations for systematic, higher-throughput electrophysiological investigation of TCF4 knockdown and knockout. Together, our lab identified that TCF4 binds within the genes SCN10a and KCNQ1 to regulate expression of these genes in primary neural progenitor cell cultures. CRISPR mutation of these genes rescues the intrinsic excitability deficits of cell-autonomous rodent models of TCF4 knockdown or knockout. These investigations defined the effects of decreased TCF4 mutations and identified potential drug-able targets and pharmacological studies for Pitt-Hopkins Syndrome (PTHS).
- **Skills acquired**: ChIP-seq analyses, molecular biology, immunohistochemistry, and image analyses.
- **Achievements and awards:** Co-author on two publications *(Neuron, Molecular Psychiatry),* national poster presentations (BMES, SfN, ACNP, presented by Drs. Stephanie Page and Brady Maher).

Identifying a myelin-related transcriptomic profile is shared by Pitt-Hopkins syndrome models and human autism spectrum disorder.

- **Advisors:** Brady Maher, PhD; Andrew Jaffe, PhD (Letters of Reference)
- **Summary:** The next focus of my in the Maher and Jaffe groups was to investigate the gene pathways disrupted in heterozygous mutations of the mouse gene *transcription factor 4* to model the rare neurodevelopmental disorder Pitt-Hopkins Syndrome and more broadly idiopathic autism spectrum disorder. I performed RNA-seq analyses on multiple mouse models of heterozygous mouse *Tcf4* mutation to identify a myelination-specific transcriptional signature concordant across mouse models of PTHS. Confirmatory electrophysiological and immunohistochemical data revealed that oligodendrocyte maturation is inhibited in these mouse models in a cell-autonomous fashion. Furthermore, this myelin transcriptional signature was disrupted in other mouse models of syndromic autism spectrum disorder and can distinguish in postmortem brain RNA-seq the diagnoses of idiopathic autism spectrum disorder patients.
- **Skills acquired**: RNA-seq data analyses, bioinformatics.
- **Achievements and awards:** A co-first author publication at *Nature Neuroscience,* national poster presentations (SfN, ACNP, presented by Dr. Brady Maher). Received a Johns Hopkins Provost Undergraduate Research Award to perform part of the work.

MD/PhD Rotation Student, Carnegie Mellon University, Pittsburgh, PA (6/2017-8/2017)
Region-specific open chromatin dynamics in the aging mouse brain.

- **Advisor:** Andreas Pfenning, PhD, Department of Computational Biology, Carnegie Mellon University
- **Summary:** I had the opportunity to understand the changing brain epigenetics with age. I co-piloted the assay for transposase-accessible chromatin (ATAC) within the Pfenning lab and sequenced libraries to access the chromatin states of the cortex, striatum, and external globus pallidum across the mouse lifespan. This work contributed to improving the ATAC-seq protocol in fresh brain tissue and experimental design of brain aging epigenetics in the Pfenning lab.
- **Skills acquired:** ATAC-seq library preparation, next-generation sequencing, and ATAC-seq analyses.
- **Achievements and awards:** Presented poster at the annual Pitt-CMU MSTP retreat.

MD/PhD Rotation Student, University of Pittsburgh, Pittsburgh, PA (6/2018-9/2018)
Transcripts in GWAS gene-desert loci introduce new players into human health.

- **Advisor:** Anne-Ruxandra Carvunis, PhD, Department of Computational and Systems Biology, University of Pittsburgh
- **Summary:** I had the opportunity to investigate whether un-characterized transcripts mapped to gene-deserts from the GTEX RNA-seq dataset of postmortem human tissues could possibly produce *de novo* proteins or even be *de novo* human genes. I investigated the mapped transcripts with available human Ribo-seq data to find evidence of translation as well as compared genome conservation of the underlying sequence across 100 vertebrates.
- **Skills acquired:** comparative genomics, bioinformatics, Ribo-Seq data analyses
- **Achievements and awards:** Presented poster at the annual Pitt-CMU MSTP retreat and selected talk at the 2018 Tree Rivers Evolutionary Event regional conference.

PhD Student, Carnegie Mellon University, Pittsburgh, PA (7/2019-present)
Addiction-associated genetic variants implicate cell type- and region-specific cis-regulatory elements.

- **Advisor:** Andreas Pfenning, PhD, Department of Computational Biology, Carnegie Mellon University
- **Summary:** I contributed to work showing preliminary findings that substance use GWAS enriched in cell type-specific and brain region-specific open chromatin areas. This finding is replicated in macaque-orthologous and mouse-orthologous regions of the human genome, and targeted mouse epigenetics with machine learning models can predict the functional effects of GWAS addiction variants.
- **Skills acquired:** comparative genomics, machine learning, single nucleus RNA-seq and analysis
- **Achievements and awards:** A co-first author manuscript in review at the *Journal of Neuroscience*, University of Pittsburgh BGSA Travel grant, selected talk at the 2020 NIDA Cross-cutting Research in Genetics and Epigenetics Meeting (presented by Dr. Pfenning)

Doctoral Dissertation in the Lab of Dr. Andreas Pfenning

The work outlined in this proposal will contribute to a substantial amount of my proposed dissertation. My goal is to understand how genetic variation affects the cell types underlying neural circuits involved in reward and addiction. I will utilize single nucleus technologies and develop new machine learning techniques for comparative genomics to determine conserved neuronal cell types and their associated enhancers to predict the functional impacts of human addiction genetic variants. These powerful techniques will help identify a cellular and genetic basis for addiction genetic predisposition between primates and rodents. With the guidance of my advisors, I hope to develop the skills and expertise in genomics, single-cell genomics, and machine learning that are necessary for me to complete the proposed project and eventually become an independent computational neurobiologist. Thus, the projects outlined in this fellowship application will align closely with my dissertation and career goals.

I have independently developed the ideas behind this project and have mastered several of the assays and analyses necessary to successfully complete the proposed experiments, as demonstrated by my previous research projects. Thus, I believe that my proposed dissertation timeline is realistic and reasonable. Funding for this proposal will support my general research interest of understanding the cell types and gene regulation of the reward circuit. Furthermore, it will provide me with the opportunity to continue my efforts toward understanding the genetic risk of patients with substance use disorder to better inform addiction care.

B. Training Goals and Objectives:

Following my training as an MD/PhD candidate, my career goal is to pursue a tenure-track faculty position at an academic medical center. I plan to combine my interests in neuropsychiatric diseases and genomics in a career as a medical scientist, spending 70% of my time as a principal investigator of a laboratory, 20% as a clinician, and 10% as an educator involved in trainee professional development and advocacy. Specifically, I would like to understand the gene-behavior mechanisms of the brain by studying how genetic variations alter gene regulatory mechanisms of cell type specific neural circuits underlying reward and addiction behavior. As genomic datasets become larger and more biologically complex, researchers in this field will increasingly need to have first-hand clinical experience in neuropsychiatric diseases to ask the right scientific questions as well as the skills to use computational approaches to answer those questions. My research interests in genes, behavior, and machine learning reinforce each other to inform both my research and clinical work to truly transform each other in a way that will allow me to better care for my patients and guide my investigations with clinically relevant research questions.

I first became interested in genetics of neuropsychiatric disorders at the Lieber Institute for Brain Development (LIBD) where I had the opportunity to consider how regulation transcription factor expression alter brain development and produce genetic liability for schizophrenia. I found inspiration in the ways by which bioinformatics could discover relevant biological patterns in complex genomic data. The pursuit of this passion led me to extend computational approaches to analyze rich imaging data addressing in different collaborations at LIBD. My interests in computational biology and neuroscience ultimately led me to the MSTP at the University of Pittsburgh and Carnegie Mellon University and the laboratory of Dr. Andreas Pfenning for my PhD work. Support from the Ruth L. Kirschstein NRSA fellowship will facilitate my training as a physician-scientist.

The opportunities for computational biology research to close gaps in the understanding of genetics in complex brain disorders formed the basis for my proposal. Prior to starting my PhD training, I completed clinical clerkships in Pediatric and Obstetrics/Gynecology. On both rotations, I encountered numerous patients with personal or family histories of SUD yet only seek care for medical conditions that arise as consequence of the SUD. Mitigating SUD by supporting those susceptible to SUD along with treating those affected by SUD would thereby improve both mental and physical health of patients. Therefore, completing of the proposed studies may help identify the genetic and cellular basis of SUD predisposition to translate to preventative interventions and better therapies for patients suffering from SUD as those that I encountered on my clerkships.

Under this fellowship, I also seek to expand my skillset as a computational biologist to study the gene regulatory patterns of cell types in the reward circuit. Before joining the Pfenning Lab, I gained experience in genomic data analyses and molecular biology during my undergraduate and postbaccalaureate training. While these early experiences provided me with an excellent foundation of technical skills to study gene regulatory mechanisms or reward cell types, I seek to expand my expertise during my graduate training with Dr. Pfenning. I seek to apply this proposal toward my development in 3 areas: **1) expansion of molecular biology technical skills**, **2) expand machine learning and algorithm development skills**, and **3) development of scientific communication and collaboration.**

1) Expansion of molecular biology technical skills: One priority of my graduate training is to become a skilled molecular biologist in genomic assays so that I can use this expertise as an independent investigator. The limited, feasible single nucleus genomic assays proposed here were designed with this goal in mind. Specifically, I have generated single nucleus RNA-seq from fresh mouse brain and will extend this approach to fresh rat and frozen postmortem human brains to investigate cell types of reward areas are conserved across species. I will receive support from Dr. Jing He and Dr. Esin Ozturk **(Letters of Support).** I will further my expertise in ATAC-seq by extending the application to the single nuclei setting (snATAC-seq). Developing proficiency in these techniques will enable me to apply novel approaches in single cell genomics as an independent investigator.

2) Expansion of machine learning and algorithm development skills: My second goal is to expand my computational skills into machine learning and developing new algorithms for genomic analyses. I will accomplish this goal through my training in the CPCB PhD program, which draws from over 100 training faculty members who specialize in computer science, machine learning, systems biology, and genomics. The graduate program provides core and elective coursework to build my foundation in algorithms for modeling biology: *Machine Learning*, *Computational Genomics*, and *Graphical Models*. I will also take *Advanced Cellular Neuroscience* to learn more about the cell types of the brain and how they have been modeled in the past. Specifically, I will refine and extend a nested tree probabilistic graphical algorithm to model hierarchical evolutionary conservation of cell type-specific enhancers across rodents and primates, particularly from Dr. Irene Kaplow **(Letter of Support).** I

will further my expertise in genomic data analysis by applying methods required to interpret complex single cell genomic datasets available in the public domain as well as those that I will generate. Moreover, as a CPCB student, member of the Neuroscience Institute, and CNBC, I have ready access to computational and neuroscience events and seminars hosted across both CMU School of Computer Science and Pitt School of Medicine. In May 2020, my addiction research presentation was well-received by neurobiology researchers within the Neuroscience Institute. These cross-departmental and inter-institutional experiences will allow me to stay up to date with methods being developed to address complex questions in genomics and neuroscience.

3) Development of scientific communication skills: A critical skill for success as an independent investigator is the ability to communicate effectively in written and oral formats to new and experienced members in the field. Dr. Pfenning is an excellent orator and writer, and I look forward to his mentorship in not just the conceptual but also the communicative aspects of doing science. To cultivate my presentation skills, I will present data in weekly lab meetings, annually within the Neuroscience Institute and CNBC, and regularly in formal journal clubs and research in progress talks for the CPCB and MSTP. I will also present my findings in poster sessions and/or symposia at the annual Society for Neuroscience and NIDA Genetics and Epigenetics meetings. To prepare for each talk, Pfenning laboratory members will critique my presentations slide by slide. To cultivate my writing, I will write manuscripts for submission to peer-reviewed journals. Each manuscript will go through iterations of feedback and revision at each stage with Dr. Pfenning. All of these experiences will provide a solid foundation for a career of publications and presentations of scientific work. To cultivate my teaching and mentorship skills, I will continue to mentor new scientists to the Pfenning lab, receive feedback from both trainees and Dr. Pfenning. One notable example of my mentees is Chai Srinivasan, the undergraduate co-first author on my preprint manuscript. To cultivate formal teaching experience, I will be guest lecture for Dr. Pfenning's graduate-level course, *Genomics and Epigenetics in the Brain*. As the TA in the Fall 2020 edition of the class, I have given six lectures and held office hours to class participants ranging from undergraduates to postdoctoral fellows.

Through these diverse activities, I will build a strong scientific skill set, continue my clinical development, and strengthen important leadership skills. By training in both the laboratory and clinical environments over the next five years, I hope to develop a robust foundation for a career as a successful physician scientist in academic medicine. During the fellowship, I will engage in intensive basic science research in conserved gene regulatory mechanisms of cell types of the reward pathway as well as continued clinical work during my PhD years and the final two years of medical school training. I will have clinical involvement during my PhD years in the form of two personalized longitudinal clinical clerkships with experts in Addiction Psychiatry and Addiction Medicine. These experiences will allow me to work with physician scientist role models to learn how successful physician scientists approach their dual roles.

Following completion of my MD and PhD, my goal is to pursue a residency in Medicine or Psychiatry at a major academic institution whose department has an institutional T32 training grant and protected research time for residents. After completion of the residency, these T32 grants provide new graduates with two years of salary and research support, allowing for the balance of 70% protected research, 20% clinical, and 10% teaching.

Afterwards, I plan to pursue a postdoctoral fellowship with the goal of eventually becoming a principal investigator at an academic institution. I am committed to pursuing a career as a physician scientist, and the availability of excellent mentors, institutional support, and protected time are integral to my continued development and growth as a trainee. My thesis mentor, institution, graduate school research and coursework, clinical coursework, and professional development were all carefully chosen to prepare me for additional training at the intersection of computational biology, neuroscience, and patient care **(Selection of Sponsor and Institution)**.

I believe that the strategies I have described will allow me to achieve my dual goals in science and medicine. Completion of my PhD and the proposed research strategy will provide me with experimental, computational, and communication skills as I work to become an academic physician scientist and contribute to the field's understanding of neuropsychiatric disorders. Ultimately, creative and meaningful investigation of disease at the genetic and molecular levels will push the boundaries of scientific knowledge and ultimately inform the ways in which I can improve the lives of my patients.

C. Activities Planned Under this Award:

	7/21-6/22	7/22-6/23	7/23-6/24	7/24-6/25
Research				
Aim 1A	20	-	-	-
Aim 1B & C	20	10	-	-
Aim 2A	15	25	-	-
Aim 2B & C	10	30	-	-
Scientific Writing	8	8	2	2
Computational Biology Coursework				
CPCB Journal Club & Prof. Devel.	2	2	-	-
Medical Training				
Longitudinal Clinical Clerkship	5	5	-	-
3rd & 4th year Clerkships	-	-	90	90
Seminar				
CPCB Seminar Series	4	4	-	-
MSTP Workshop Series	2	2	2	2
Meetings				
Individual Meetings w/ Dr. Pfenning	2	2	2	2
Group Meetings, Pfenning Lab	8	8	-	-
Conferences				
Neuroscience (SfN)	2	2	2	2
NIDA Genetics & Epigenetics Meeting	2	2	2	2

Research: I seek to complete **Aims 1-2** by end of Year 3. I anticipate most of my research time will be committed to completing **Aim 2** as I collect the single-nucleus ATAC-seq samples and develop algorithms to analyze the data across species. I have assembled a team of consultants and experts from within as well as outside of the lab to mentor me on potential challenges surrounding **Aims 1A** and **2A** to generate single cell genomic datasets in mouse, rat, and human postmortem brain **(Accomplishing Proposal)**. Throughout this time, I plan to summarize and disseminate findings as posters, talks, and peer-reviewed publications. In Years 2-3, I will complete any manuscripts in preparation to defend my dissertation.

Computational Biology Coursework: I will complete the required core coursework by the end of Year 2 and take two electives focused on neural circuits of the cortex and basal ganglia (relevant to **Aim 1A, 2A**) and probabilistic graphical models (relevant to **Aim 1B, 2B**). I will also complete my thesis proposal by the end of year 1 to advance to candidacy by the beginning of Year 2.

Medical Training: I will participate in two Longitudinal Clinical Clerkships in Year 1 and 2, which will provide me with clinical continuity (one half day per week for 20 weeks each) in addiction psychiatry and addiction medicine subspecialties. In Year 3, I will complete Clinical Clerkship Bootcamp, a 2-week ungraded rotation on a Medicine inpatient service to polish clinical skills and preempt my return to medical school. From Year 4 onwards, I will complete required medical clerkships and complete my MD/PhD training.

Seminars: The CPCB has regularly student research seminars, journal clubs, and faculty presentations that I will attend during my 3 years in the Comp Bio PhD program. These events provide opportunities to critically analyze relevant papers, present my research to peers, and network with visiting experts. I will also participate in the monthly MSTP workshops, which continue to provide opportunities for professional development.

Meetings: Throughout all 4 years, in addition to impromptu interactions with Dr. Pfenning, I will attend our general weekly lab meetings. In addition, I will participate in experimental and computational group meetings that enable lab members and Dr. Pfenning to facilitate our cross-cutting teamwork and collaboration.

Conferences: During each year of the PhD and in my clerkships, I plan to present at and/or attend SfN and the NIDA Genetics and Epigenetics meetings.

Specific Aims:

The prevalence for substance use disorders (SUD) has substantially increased worldwide in the last three decades[1]. Human genetic studies have identified genetic factors underlying of SUD, reporting up to 80% heritability[2–5]. Similarly, rodent studies also found that genetic variation explains a significant portion of addiction behaviors[6–10]. The molecular basis for these SUD predispositions in humans and model organisms, however, remains unknown. Thus, the overall goal of this proposal is to define the molecular basis of genetic predisposition to SUD.

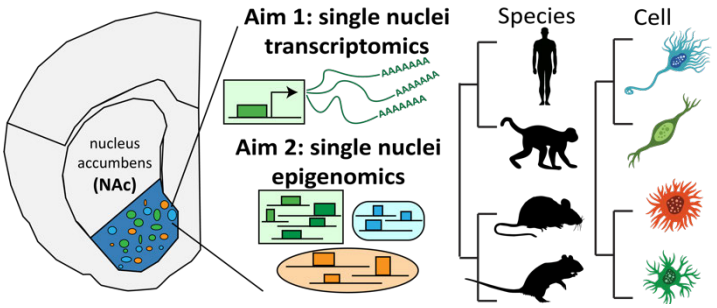


Figure 1. Hypothesis: human SUD risk variants enrich in NAc cell type marker genes and CREs conserved in primates and rodents. I propose to define the molecular basis of genetic predisposition to SUD with single-cell genomics and novel computational algorithms.

SUDs hijack many components of the reward neural circuit, especially cells in the nucleus accumbens (NAc)[11–15]. Genetic risk variants associated with complex neuro-behavioral traits, such as addiction, demonstrably impact gene *cis* regulatory elements (CREs)[16–27]. CREs are largely cell type-specific[28–34] and are conserved across species at varying degrees[35–40]. Therefore, identifying which cells utilizing genes and CREs affected by SUD risk variants can provide insight into the context-dependent molecular basis of SUD genetic risk (**Figure 1**). Our lab in collaboration with Dr. William R. Stauffer **(Letter of Support)** used single nuclei transcriptomics to show that macaques have unique NAc cell types and marker genes that may be relevant to human addiction[41]. We further showed that human and orthologous mouse CREs of several NAc cell types enrich for SUD risk variants[42]. Whether orthologs of these genes and CREs detected in model organisms stay cell type-specific or maintain biological relevance to human SUD remains to be delineated.

Through analyzing our inhouse and publicly available single cell datasets, I am poised to test the overarching hypothesis that human SUD risk variants enrich in NAc cell type marker genes and CREs that are conserved in primates and rodents (**Figure 1**). Understanding which NAc genes and CREs may be affected by human risk variants for SUD, particularly those measurably conserved in primates and rodents, will revolutionize the playbook for translating human polygenic diseases and disorders to model organisms, direct future studies into molecular mechanisms of SUD, and ultimately enable progress to curb the substance addiction epidemic.

Aim 1: Identify conserved, active NAc marker genes and relation to human SUD genetic risk. Due to the evolutionary conservation of mammalian genes, I hypothesize that SUD risk variants enrich near primate-rodent conserved marker genes of NAc cell types. Results of **Aim 1** will systematically associate the heterogeneous collections of SUD loci to candidate cell types and marker genes that classify conserved NAc cell type-specificity.

1a. Evaluate the molecular composition of NAc cell types through identification of marker gene profiles using single nucleus RNA sequencing (snRNA-seq) independently in mice[43–45], rats[46], macaques[41], and humans[47].

1b. Develop a graphical model to jointly compare single cell gene expression across the species and cell types employing the principle of maximum parsimony over known evolution and neurodevelopmental hierarchies.

1c. Determine whether SUD risk variants[48–54] are enriched in conserved or divergent NAc marker genes.

Aim 2: Identify conserved, active NAc cis-regulatory elements and relation to human SUD genetic risk. Due to the lower evolutionary selection for CREs than genes, I will investigate whether human SUD risk variants enrich in primate-conserved or rodent-primate shared cell type-specific CREs. **Aim 2** results will provide an important epigenomic view to complement gene-based findings in **Aim 1** and provide genome-wide CRE candidates in other species that can be targeted to model the spectrum of human polygenic risk *in vivo*.

2a. Identify the putative CREs with single nucleus assays for transposase accessible chromatin (snATAC-seq) in NAc cell types separately in mice[55], rats, macaques, and humans.

2b. Extend the model from **Aim 1b** to model single cell putative CRE activity applying maximum parsimony to the same hierarchies in species and neural cell types.

2c. Determine whether SUD risk variants enrich for conserved, active or divergent NAc cell type-specific CREs.

The identification of genetic risk loci for SUD in humans gives us the opportunity to identify molecular mechanisms contributing to addiction behavior. This proposal will use innovative genomic technologies, integrate the molecular evolution of genes and CREs, and illuminate how human SUD risk variants might target gene regulatory mechanisms in cells of the NAc. Completion of this proposal will provide excellent training in developing machine learning algorithms and applying novel approaches to study human disease mechanisms. Importantly, this proposal will lay the foundation of my academic career as an independent physician-scientist.

RESEARCH STRATEGY

1. Significance. The prevalence for substance use disorders (SUD) has substantially increased worldwide in the last three decades[1]. In the United States, 19.7 million people had a reported SUD in 2017[56]. Individuals with SUD suffer increased mental and medical health burdens[57], which recently includes significantly higher risk for COVID-19[58]. Treatments for many classes of addictive substances do not exist. For those that do, there are physiological and logistic barriers that limit efficacy[59–62]. On my clinical rotations, I encountered too many patients with personal or family histories of SUD who receive care for only medical conditions that arise downstream of the SUD. Proactively identifying and supporting patients susceptible to SUD in addition to improving therapies for SUD would improve the overall health of patients. Therefore, this proposal addresses the critical need to better understand how large segments of the population are predisposed to SUD, identify the molecular equivalents in model organisms for translational research, and move towards more effective treatments for SUD.

The premise of this proposal is that genetic variation in the human population contributes to substance use behaviors. I aim to identify the context-dependent molecular targets for human SUD genetic risk variants, identify the orthologous elements that may be conserved in other primates and rodents, and predict how they might affect different cell types within a key brain region in the reward neural circuit, the nucleus accumbens (NAc).

1.1 What are the cellular contexts of the NAc in addiction? Addiction is a disorder of dopamine reward signaling from the ventral tegmental area (VTA) activating downstream areas, most prominently the NAc[12,63,64]. In the last four decades, many groups have shown that both human and rodent NAc share organization, anatomy, morphology, histology, immunohistochemistry, neural circuits, and behavioral function[65]. Until recently, molecular characterization of NAc cell types has been with a handful of genes. Classically, the main NAc neuronal cells, medium spiny neurons (MSN), are equipped with dopamine receptors D1 or D2 to respond to VTA input. They further integrate signals from other areas (prefrontal cortex, amygdala, etc.) to mediate addiction behaviors[63–68]. More recent works leveraging rodent genetic tools and high-throughput technologies spotlight how other cell types in the NAc like interneurons[13,14,68–72] or glia[15,73] play roles in substance use and addiction pathogenesis.

1.2 Contexts and complexities of human SUD genetic risk. Twin-studies have shown that risk for SUD is highly heritable[3,74,75]. Genome-wide association studies (GWAS) for substance *use* (SU) behaviors have uncovered many loci related to this heritability[48,76–78]. Each GWAS reports one aspect of addiction behavior (initiation, dose, etc.) from an addictive substance (nicotine, alcohol, etc.). Unlike other GWAS traits with one clear measure like height, the spectrum of SU traits makes the results of addiction-related GWAS difficult to interpret. Yet the apparent lack of consensus risk loci might reveal how complex domains of addiction behaviors are linked to the human genome. Results from my co-first author preprint support this idea, showing that patterns emerging from SU risk loci correlate with epigenomic contexts of various cell types and brain regions, including D1 and D2 MSNs in the NAc[42]. These results consistently held up whether the epigenomes came from human or orthologs of mouse. More recently, significant efforts have detected numerous GWAS loci for SUD and teased out nuances between genetic risk for SU versus SUD[5,50,52–54,79–81]. Together, the SU and SUD risk variants provide a sound basis for studying the mechanisms and context-dependent genetic predisposition for addiction.

1.3 Molecular evolution of cell types and their gene usage. Recent advances in single cell or nuclei RNA sequencing (snRNA-seq) reveal molecularly distinct cell populations in the brain, allowing unbiased identification of orthologous cell types across mammalian evolution using highly conserved genes. The few cross-species comparisons find that human neural cell types defined according to gene expression are largely conserved in primates and rodents[82–85]. However, *Krienen et al.* identified a type of primate-specific striatal interneurons not found in rodents[85], and *Hodge et al.* described shifts in gene expression, spatial distribution, and morphology between mouse and humans in otherwise orthologous cell types[82]. These notable examples raise the question: in *which* cell types or species do highly conserved genes change their expression levels or relevance to human disease? Using the data from these papers to answer this question will require me to develop new computational tools that account for the 90-million-year evolutionary gap between primates and rodents.

1.4 How could genetic variation modulate substance addiction? The genetic variants measured in SU and SUD GWAS, similar to those from neuropsychiatric GWAS, enrich within intergenic and intronic regions of genes above background. Myself and others find specific enrichment of this signal to be within likely *cis*-regulatory elements (**CREs**)—which have a high degree of cell type-specificity and can be conserved across species[19,42,86–88]. CREs allow one genome to regulate the genes and gene products for all the cell types in the organism, so genetic variants within CREs may alter cell type-specific function. One hypothesis supposes that genetic variants alter transcription factor binding motifs and inhibit the cell's normal response to a stimulus such as highly addictive drugs[89–92]. One example tested sporadic Alzheimer's Disease (AD) genetic risk variants in microglia

CREs to show that deletion of the CRE that harbor risk variants blunted gene expression in microglia but not in other cells[30]. SUD genetic variation could therefore affect an individual's resiliency to substance use depending on the impact and context of CREs in homeostatic gene regulation, cellular function, or neural circuitry.

1.5 Conservation of cell type-specific CREs and disease. Animal models of genetic diseases have manipulated the homologs of highly penetrant genes. No such paradigm exists to model lowly penetrant, polygenic human conditions. A challenge has been inability to map CREs between species due to non-coding DNA being less conserved than the 1-2% of the genome that codes for genes. However, the Zoonomia Consortium, in which the Pfenning Lab participates, used alignments of >240 deeply sequenced mammalian genomes to find that 3.1% of bases in human genome is under purifying selection[93]. This doubles previous records and powers our ability to detect conserved, functional non-coding DNA sequences. Furthermore, this advancement pairs well with single nucleus assay for transposase accessible chromatin sequencing (snATAC-seq) to map cell type activity of CREs across species. Systematically tracing the conserved, active CRE elements of distinct cell types could revolutionize future experimentation to model human polygenic disease; one direction might find human CREs in risk loci and deliver massively parallel guide RNAs targeting the orthologous CREs in transgenic mice with CRISPR interference or activation[94] to model high and low polygenic disease risk *in vivo*.

To summarize, three major challenges must be considered to understand the molecular basis of human genetic predisposition for SUD. **1)** The heterogeneity in SUD GWAS genetic risk loci may represent differences in cellular contexts within the NAc in which the variants affect gene expression, **2)** The gene expression of human NAc cell types have yet to be compared with model organisms to determine if there is evolutionary cellular specialization, and **3)** the CREs of human NAc cell types, which are likely targets of SUD genetic risk variants, have yet to be mapped or compared with model organisms for evolutionary cellular specialization. These considerations suggest the **hypothesis—to be tested here—that human SUD risk variants enrich in NAc cell type marker genes and CREs that are conserved in primates and rodents.** Although many other brain regions are involved in addiction, practical concerns of acquiring multi-species, multi-region snRNA-seq and snATAC-seq data limit the scope of this study. Extensive functional and behavioral studies of the NAc in human[11,95–98] and rodent[8,12,65,67,99–101] make the NAc an likely region that could mediate components of SUD genetic predisposition.

1.6 Innovation & Training. My proposal integrates a number of technical and conceptual innovations.

a. *Technological innovations*: There are no algorithms that can model the molecular evolution at single cell-resolution jointly across multiple species. I will develop novel machine learning methods to do so using single cell gene expression and open chromatin data, incorporating evolutionary relations of species and cells.
b. *Conceptual innovations*: There are limited information about the cellular context and molecular mechanisms in the NAc on SUD genetic risk variants, and my study will provide critical insight in the role of conserved or recently evolved cell types of the NAc in human genetic risk for SUD. Furthermore, systematic detection of conserved, active CREs in humans and model organisms with single cell epigenomics remains unexplored.

Finally, the aims of this proposal will provide me with an outstanding training opportunity as a physician scientist. Particularly, I will develop expertise in a number of arenas, ranging from machine learning, to single cell genomics, to novel approaches to study complex human disease mechanisms. This proposal, together with the training provided by the CMU-Pitt Computational Biology program and mentorship from my sponsors, Drs. Andreas Pfenning and Ryan Logan, will ideally position me to become an independent physician scientist.

2. GENERAL APPROACH: I initially outline shared strategies that support the hypotheses stated in **both Aims**. Single cell datasets**.** The proposed experiments will use snRNA-seq and snATAC-seq datasets from two rodent and two primate species in the NAc. These datasets, when possible, will come from collected or publicly available sources. I will collect a minority of new snATAC-seq samples from rat and human with guidance from Dr. Jing He, the Stauffer Lab member who generated macaque snRNA-seq and snATAC-seq data[41] **(Letter of Support)**.

2.1 Contribution and training opportunity: These results and new data could establish, for the first time, a molecular basis human SUD genetic risk in the NAc. Not only will the data generated by **Aim 2a** provide me with invaluable training in single cell genomic assays, it will be a rich dataset for comparative genomics and addiction neurobiology, providing single cell-resolution atlases of the NAc to translate human polygenic disease to animal models. The machine learning methods developed in **Aim 1b, 2b** will provide excellent algorithmic training and provide the novel computational tools to investigate cross-species cellular evolution. These resources will align with FAIR principles of data access as they are disseminated to the wider community **(Resource Sharing Plan).**

2.2 Scientific rigor and statistical analysis: An N=3 per sex per species powers discovery of cell type-specific genes/enhancers while adjusting for covariates[102–104]. For reproducibility, I will compare cell type profiles within

species and across experimental conditions when multiple, independent datasets are available. I have also collected independent samples towards this goal. I evaluated QC metrics (map rate, % rRNA, periodicity, complexity) and found that our samples are high-quality. The analyses proposed do not allow for sample blinding.

2.3 Aim 1b, 2b: Nested Tree Probabilistic Graphical Model

Rationale: There is no ideal tool for modeling molecular evolution of gene activity in cell types across species, as the state-of-the-art tool is limited to pairwise comparisons of clusters identified within each species[105,106]. To overcome this limitation, I will extend and develop a novel nested tree probabilistic graphical model (ntPGM), in which gene expression or putative CRE open chromatin feature within a cell type is modeled as a combination of a cell type hierarchy and a species hierarchy.

Nested Tree Model: In a hierarchical ntPGM, each connection has a nested overlaying tree model, where a parent species tree overlays a child cell type tree, and every corresponding hierarchical cell type has a link from parent to child (**Figure 2**). The species tree is pre-defined with known biology. The structure of the cell type tree is preset with a user dendrogram or built in a data-driven manner using hierarchical clustering that enables tuning the resolution of a "cell type"[107].

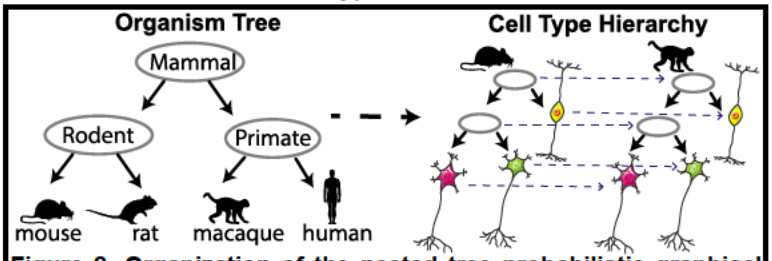


Figure 2. Organization of the nested tree probabilistic graphical model (ntPGM). The method takes in a hierarchical tree of organisms (left) and brain region. Then it is initialized with a hierarchical tree of neural cell types (right). The model jointly learns cell type-specific and species-specific features of gene expression or chromatin accessibility.

Parameters & Training: Modeling gene/CRE feature matrix can be viewed as shifts of distributions at each node with parameters capturing changes in Normal/Binomial distributions. The model is parameterized with vectors of delta means/probabilities and delta variances. To optimize, I will use a likelihood cost function weighing both the reconstruction error of the true data in the feature matrix and a Lasso penalty penalizing non-zero parameters[108]. This penalty encourages a maximum parsimony structure. For example, D1 and D2 MSNs share the marker gene *PPP1R1B* that is upregulated in both cell types; the ntPGM captures this upregulation as a positive delta at the MSN level, rather than at both D1 *and* D2 levels. Therefore, gene expression differences that occur hierarchically in species and cell types are represented throughout the structure of the tree. The model optimizes by an efficient Hamiltonian Monte Carlo[109] sampling to explore and maximize the likelihood of the parameters with respect to the data and Lasso penalty. Training is parallelized feature-wise across compute resources.

Marker Inference: The feature level of any cell is the sum of the mean/probability vector from each node in the ntPGM along the path to the root. Marker features are selected by choosing the features with largest, significant effect size of a cell type at that cell type's resolution (neuron or D1). The position in the species tree thereby measures the degree of marker evolution. I will permute cell labels to compute empirical null distributions of model parameters. I will use phylogenetic-aware false-discovery rate[110] to identify active or evolving markers.

3. AIM 1 APPROACH: Identify conserved, active NAc marker genes and relation to human SUD risk loci.

Species	N_{Sample} (M/F)	N_{Cells} x1,000	Sample Prep, Technology	snRNA-seq Publication
human	5 (5/0)	~13k	Nuclei, 10X v3	Tran et al. 2020[47]
macaque	2 (1/1)	~15k	Nuclei, 10X v3	He, Kleyman, 2020[41]
rat	8 (4/4)	~7k	Nuclei, 10X v2	Savell et al. 2020
mouse	6 (2/4)	~13k	Cell, 10X v1	Zeisel et al. 2018[45]
mouse	26 (26/0)	~1.2k	Cell, SMARTer or SmartSeq2	Gokce et al. 2016[44]
mouse	9 (9/0)	~1.2k	Cell, SmartSeq2	Stanley et al. 2020[43]
mouse	1 (0/1)	~7k	Nuclei, 10X v3	This study, collected

3.1 Rationale: The Pfenning lab, in a joint effort with Dr. William Stauffer **(Collaborator Letter)**, has applied snRNA-seq to deeply profile cellular diversity in the macaque striatum. We found molecularly distinct MSN subtypes from the classical D1 and D2 MSNs that are restricted to the NAc[41]. Independent studies by another group using snRNA-seq of postmortem human NAc has also identified distinct MSN subtypes that share marker genes with the subtypes we identified[47]. Whether these findings describe orthologous NAc subtypes and if these cells are novel features evolved in the primate NAc remain unexplored. Nevertheless, these rich primate snRNA-seq datasets and extensively published datasets profiling the rodent NAc[43–46,111] set the stage for **Aim 1** of this proposal to develop novel computational algorithms and investigate the hypothesis that human SUD risk variants enrich in NAc cell type marker genes that are conserved and remain cell type-specific in primates and rodents.

3.2 Aim 1a, snRNA-seq data analysis: I will gather the processed, annotated NAc cell by gene count matrices from snRNA-seq datasets (above). For each species, I will perform the following steps: remove doublets with *Scrublet*[112], impute dropout with *MAGIC*[113], and depth-normalize with *scran*[114] to overcome the issues of low

sampling in snRNA-seq data[113,115]. For datasets with strong batch effects, I will batch-correct count matrices with *bbknn*[116]. I will use the novel multi-resolution clustering tool *ACTIONet*[117] to transfer labels between inconsistently labeled datasets and loosely align clusters between species. I will use the *Ensembl* gene catalog to subset to genes with 1-1 human orthologs. These processing steps will tune the intra-species data to have hierarchical cell labels present in all species for differential marker gene testing by *ad hoc* cross-species comparisons using *limma-trend*[119] or systematically with ntPGM **(Aim 1b)** at a false discovery rate < 0.05. The *ad hoc* approach will compare the following cell types within and between species: one vs. all, consensus of one vs. one, between terminal cell types, between broader cell types (neuron vs. glia), and between clades (primate-specific).

3.3 Aim 1b, phylogenetic cell type gene evolution: I will fit a ntPGM to the normalized, imputed, batch-corrected, ortholog-matched count matrices of all four species to identify shared or clade/species-specific events at the gene-level of orthologous cell types, as outlined above. I will compare ntPGM marker gene conservation to the *ad hoc* differential methods results from **Aim 1a**.

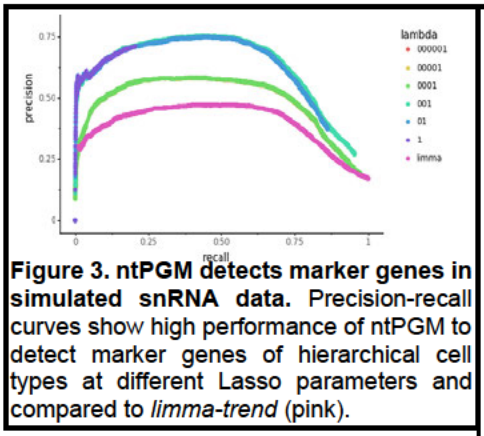


Figure 3. ntPGM detects marker genes in simulated snRNA data. Precision-recall curves show high performance of ntPGM to detect marker genes of hierarchical cell types at different Lasso parameters and compared to *limma-trend* (pink).

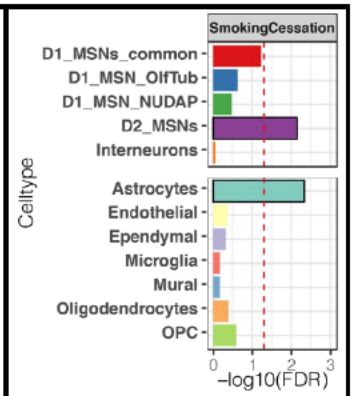


Figure 4. Gene-based LDSC find smoking cessation risk loci enrich for macaque D2 MSNs. Macaque cell types & markers detected in **Figure 5**.

3.4 Aim 1c, relating marker genes to human SUD: A number of methods[120–125] have been published to relate GWAS summary statistics with sets of marker genes while accounting for linkage disequilibrium (LD) structure of a population. I will apply popular tools *MAGMA*[122] and partitioned-LD Score Regression (*LDSC*)[120,121], to determine the cell type(s) enriched by sets of active or diverging genes from the ntPGM or *ad hoc* approach from **Aim 1a**. For *LDSC*, I will take the marker genes' intronic and proximal intergenic regions to be the foreground to intersect with SUD GWAS loci for enrichment analyses as in **Figure 4**.

3.5 Preliminary Results: Figure 3 shows the ntPGM with simulated demonstrating clear gains in marker gene testing compared to *limma-trend*. Preliminary results of **Aim 1b** on inhouse macaque snRNA-seq shows ntPGM can capture a nested multivariate experimental design (**Figure 5**). QC metrics, nuclei capture, library complexity, and sequencing depth were comparable or superior to other published NAc snRNA-seq datasets. In **Figure 5a**, the ntPGM takes in the hierarchical structure of striatal brain regions and find strong similarities across sub-regions, but also noticeable biological differences **(Figure 5b)**. The cell type hierarchy, inferred from the data, recapitulates known biology of striatal cell types **(Figure 5c)**. In **Figure 4,** gene-based partitioned LD score regression analyses show non-coding regions of human genes that are orthologous to macaque D2 MSN and astrocyte marker genes are enriched in GWAS variants of Smoking Cessation.

3.5 Anticipated results: I expect to identify the major cell types and marker genes in each species (Aim 1a) and find conserved marker gene expression levels are stably active across primates and rodents (Aim 1b). E.g., finding that D1 and D2 MSN populations identified in individual species and remain marker genes across primate and rodent would indicate that these genes are essential in maintaining the conserved reward circuit. However, they might reveal human- or primate-specific marker genes that would provide insight to understand apparent differences in reward responses between human and model organisms. Additionally, I expect to find human SUD risk variants enriched within and around non-coding regions of marker genes of orthologous cell types (Aim 1c). A finding that SUD risk variants enrich within noncoding regions of conserved, cell type-specific marker genes for would suggest perturbation of local gene regulators, specifically in these cell types, contribute to human genetic risk for substance use. However, **Aim 1's** approach is open to the alternate hypothesis that there can be cell type-specific

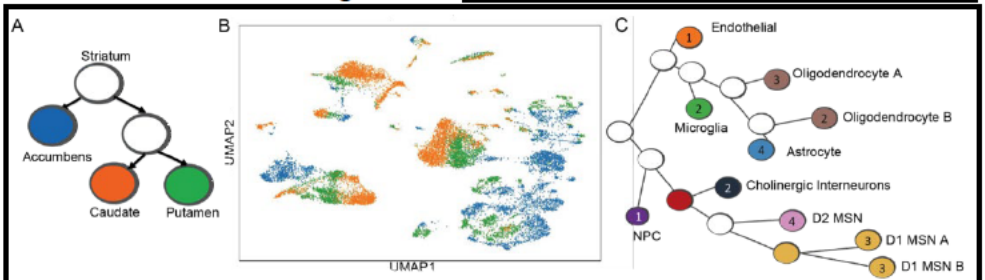


Figure 5. Macaque snRNA-seq analyzed with nested probabilistic graphical models. **(A)** A hierarchy of macaque striatal regions assayed with snRNA-seq based on gene expression, functional, and structural similarity. Here, a tissue tree is used in place of species tree for proof-of-concept for analogous cross-species analysis. **(B)** Using the pipeline detailed in **Aim 1a**, single nucleus RNA-Seq clustering are visualized in two dimensions using UMAP. Colors correspond to striatal regions. **(C)** A hierarchical clustering of cell types recapitulates known biology of cell type similarity **(in review)** [41].

genes that are enriched for human SUD risk variants but are not active across species in that cell type. Such a finding would still prove interesting and study of this cell type could lead to insights into the limits of rodent addiction models to capture the nature of human SUD genetics.

3.6 Potential pitfalls and alternative approaches: A major limitation of this study is stitching together multiple datasets which confounds species with experimental factors (study, batch, sample preparation, sex, and assay). Including macaque and rat species in cross-species conservation analyses increases the computational power to identify clade-specific evolution. Using only human and mouse could instead find convergent evolution or spurious findings from confounds with species. Conveniently, shared effects across all cells in a species propagate to the cell tree root and species tree leaves, so species-specific events further down cell type tree see fewer effects from confounding. This would not control nonlinear or interactions between confounding effects and cell type. To test accuracy of detecting species and clade-specific marker genes, I will fit a ntPGM in leave-one-out analyses for each of the 4 mouse datasets, validating marker gene predictions in the held-out group. Including my own data will help mitigate concerns about biological reproducibility (**Figure 6**).

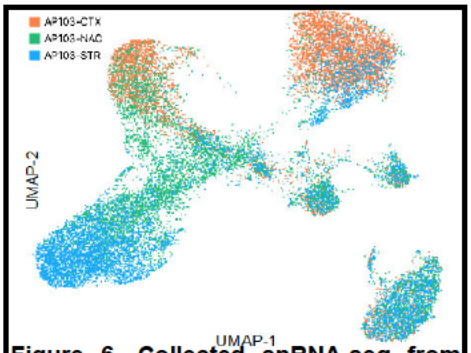


Figure 6. Collected snRNA-seq from fresh mouse brain. Ctx, NAc, and Str, and sequenced with a MiSeq to demonstrate high-quality libraries and adequate nuclei capture. These data test reproducibility for **Aim 1b** and show feasibility in **Aim 2a**.

4. AIM 2 APPROACH: Identify conserved, active NAc CREs and relation to human SUD genetic risk.

Species	N_{Sample} (M/F)	N_{Cells} x1,000	Relevant Tissues	snATAC-seq Publication
macaque	1 (1/0)	~11k	NAc, Caud,	Stauffer Lab, collected
mouse	4 (2/4)	~23k	NAc, CPu, …	Li et al. 2020[55]
human	6 (3/3)	-	NAc	This study, planned
rat	6 (3/3)	-	NAc (or CPu)	This study, planned
human	3 (3/0)	~27k	Caud, …	Corces et al. 2020[126]

4.1 Rationale: The Pfenning lab has made two efforts that build the data and algorithmic foundation for **Aim 2** of the proposal. First, lab members[41,127] and others[27,43,44,46,47,55] generated cell type-specific ATAC-seq or snATAC-seq data in mouse, macaque, and human to study SUD genetic risk. Second, we have developed tools to map 1-1 CRE orthologs across mammals[128] using highly accurate multiple genome alignments[129–131] from 240+ deeply sequenced genomes[93]. Recently, I have found that tens of thousands of putative CREs from mouse bulk cell type-specific ATAC-seq[42] and **preliminary** macaque snATAC-seq are conserved, remain active, and have consistent cell type-specificity in humans **(Figure 8)**. More importantly, these human orthologs of CREs, even from rodents, are significantly enriched for SUD risk variants **(Figure 9)**. While these findings lack the breadth and depth of the proposed **Aim 2**, they enable me to test the hypothesis that human SUD risk variants enrich in primate-conserved or rodent-primate shared cell type-specific CREs.

4.2 Aim 2a, snATAC-seq sample collection: Newly collected rat and human subjects will be sex-matched with specimens in existing datasets (above). N=3 biological replicates per sex in will be collected. Rats will be anesthetized with isoflurane and rapidly decapitated according to approved IACUC protocol TR201900003. Brain sections will be cut in aCSF using a vibratome and the NAc will be rapidly micro-dissected into ice-cold nuclei lysis buffer. I will collaborate with Dr. David Lewis to select postmortem human subjects for snATAC-seq **(Letter of Collaboration)**. Pre-dissected NAc from frozen postmortem human will be thawed on ice in nuclei lysis buffer. Briefly, tissues will be homogenized using a dounce homogenizer in the lysis buffer. Nuclei will be washed, filtered, and isolated with an iodixanol gradient centrifugation. Clean nuclei are resuspended, filtered, and washed in Nuclei Resuspension Buffer (NSB), counted, and incubated with 10X Genomics Tn5 transposases. The 10X Chromium controller will be used to barcode and capture individual tagmented nuclei. Sequencing libraries will be prepared and sent for deep sequencing on an Illumina Novaseq (Genewiz) as in the Stauffer lab.

snATAC-seq data analysis: I will gather and process NAc snATAC-seq data similar to **Aim 1b** to generate normalized CRE activity matrices for each species with cell labels aligned across species with changes appropriate for snATAC-seq data. I will align snATAC reads to the respective genomes using *Bowtie2*[132] and remove PCR duplicates. I will input mapped reads into the snATAC analysis multitool *ArchR*[133] to remove doublets, TF-IDF normalize, cluster iteratively, annotate cells with gene-activity scores, call peaks with MACS2[134] on pseudo-bulk profiles of each cluster, identify reproducible peaks across biological replicates, generate peak accessibility count matrices, batch-correct with *bbknn*[116], and filter out peaks in exons and promoters to restrict to putative CREs. I will impute the CRE peak matrix with *scOpen*[135], which is designed for snATAC sparsity and generates a bounded probability matrix of peak accessibility. I will use HALPER[128] and 240+ genome alignment[93]

to pairwise map CREs between species, select 1-1 CRE orthologs in hg38 coordinates, then perform *ad hoc* cross-species comparisons as in **Aim 1a** using *Cicero*[136] which has differential accessibility testing for snATAC.

4.3 Aim 2b & 2c Phylogenetic CRE activity & relation to human SUD: I will fit a ntPGM to ortholog-matched CRE accessibility matrices to identify shared or clade/species-specific evolutionary events at the CRE-level in orthologous cell types, as in **Aim 1b**. I will compare ntPGM CRE markers to *ad hoc* differential methods results from **Aim 2a**. Here, partitioned-*LDSC*[120,121] can directly take sets of conserved, active or diverging CREs from the ntPGM or *ad hoc* differential tests from **Aim 1a** against a reference of conserved CREs from NAc cell types to identify enrichment in human SUD risk loci while also controlling for biases due to LD in the epigenome.

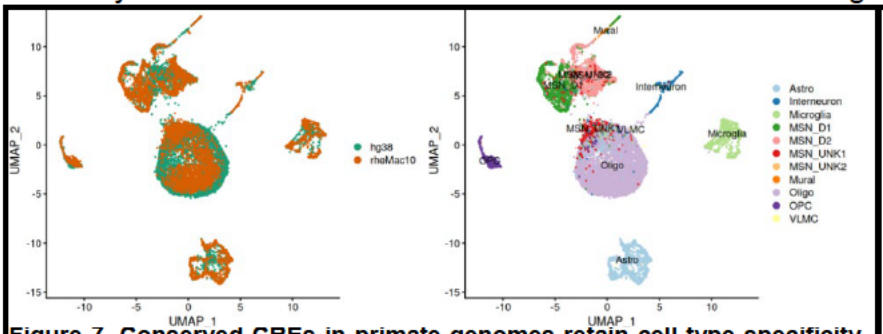


Figure 7. Conserved CREs in primate genomes retain cell type specificity. Cross-species integration of snATAC-seq peaks using 1-1 CRE orthologs between human and macaque. Left plot: macaque snATAC reads in conserved CREs can align unlabeled cells to human cells. Right plot: agreement of cell annotations identified within each species using gene scores of known markers.

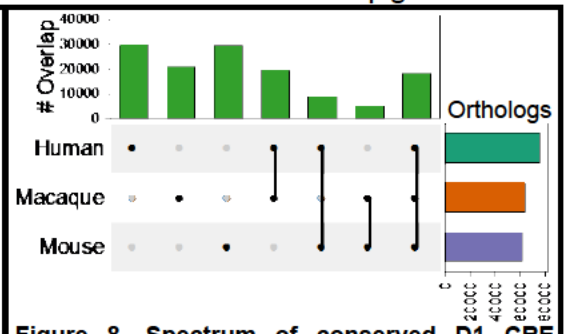


Figure 8. Spectrum of conserved D1 CRE between mouse, macaque, and human. Upset plot of D1 CRE peaks from caudate snATAC-seq in one species that is mappable and is a peak in MSN D1 cells of another species (i.e., conserved, active).

4.4 Preliminary Results: Preliminary results on snATAC-seq from mouse, macaque, and human caudate snATAC-seq demonstrate the feasibility of **Aim 2a** and reveal the complex nature of CRE conservation and retaining cell type-specificity. Only 60-70% of CREs share orthologous DNA between rodents while humans and macaque share >90% of CREs detected in either species. Yet >270,000 CREs can be mapped in all species. **Figure 7** shows that non-coding DNA of human and macaque retains sufficient biological signal for unsupervised algorithms to align orthologous cell types using just 1-1 CRE orthologs. **Figure 8** shows CREs detected in D1 MSN independently in each species range from being active in all species to active only in one. **Figure 9** shows that mouse CREs of D1 and D2 MSNs enriched for human SU risk loci and aid fine-mapping of SNPs in risk loci.

4.5 Anticipated results: I expect to find sets of CREs conserved between primates as well as sets shared between primate-rodent that remain cell type-specific and be able to associate them to human SUD variants. While **Figure 7-9** suggests that **Aim 2** will yield rodent-primate shared or primate-specific cell type-specific CREs, I am open to finding that human-specific CREs are more enriched for SUD risk. This would still prove interesting and suggest instead that human-accelerated functional CREs alters molecular resilience to SUD. Results of **Aim 2** could establish the first cross-species mapping between cell type-specific CREs that can translate human polygenic disease to model organisms.

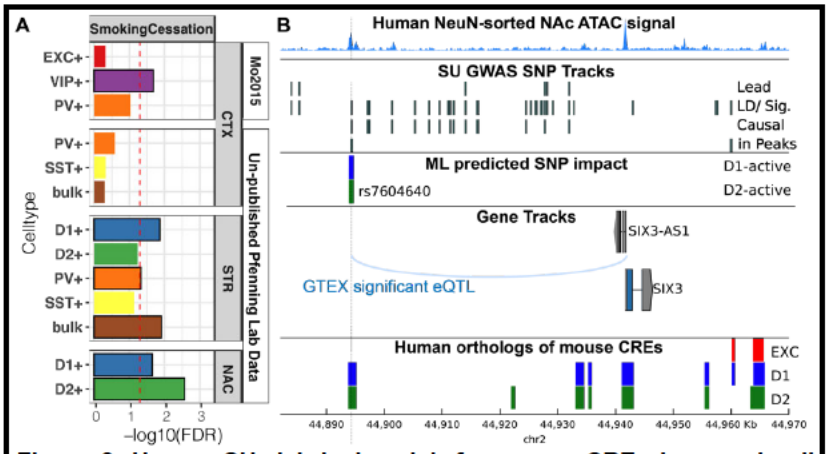


Figure 9. Human SU risk loci enrich for mouse CREs in neural cell types. (A) Partitioned LDSC analyses on human CRE orthologs of multiple mouse cell type-targeted ATAC-seq. **(B)** Fine-mapping SU GWAS variants by intersecting human NAc NeuN+ ATAC, mouse orthologous CREs, deep learning models of SNP effect, and GTEX striatum eQTL data predicts that one of many variants impact D1/2 MSNs in the *SIX3* locus **(in review)**[42].

4.6 Potential pitfalls and alternative approaches: Given the support of my consultants Drs. Esin Ozturk and Jing He **(Letters of Support)** and my own experiences with ATAC-seq and snRNA-seq in mouse brain, I do not anticipate major technical challenges to perform snATAC-seq on fresh rat brain. However, if low quality nuclei from human postmortem brain makes snATAC-seq infeasible, then I will pivot the aim to investigate the caudate, which has known roles in addiction[11,63,137] (**Figure 8**) and arguably contains most cell types of the NAc[41]. I have tested **Aim 2**'s feasibility on this near-complete caudate dataset, so pivoting to alternate approach will cause no delay. Performing snATAC-seq in the rat caudoputamen would still be innovative and relevant to SUD studies and, importantly, achieve the training goals.

Preparing Proposal

I identified Dr. Andreas Pfenning as my thesis mentor after completion of my two research rotations. During medical school, I met with Dr. Pfenning on a regular basis with the goal of constructing a research plan and preparing the necessary protocols, datasets, animals, reagents, and resources prior to my transition from medical to graduate school. Over the past several months, Dr. Pfenning and I discussed possible projects to create a proposal that combines my interests and strengths and identifies areas of growth to ultimately prepare me to become an independent researcher. I wrote all elements of this proposal, excluding letters of technical support and recommendation and the Sponsor and Co-sponsor Information Section, which was written by Dr. Pfenning and Dr. Logan. I then obtained feedback from Drs. Pfenning and Logan on all elements of this proposal who gave me impactful critiques that aided in the development of the final proposal. Specifically, I devised the proposed experiments and hypotheses, with the aid of critical feedback and suggestions from Dr. Pfenning and consultants about experimental design, data presentation and grantsmanship for the training plan. I would like to thank the reviewers of my initial proposal and have redesigned the research strategies with Dr. Pfenning and Logan to address their concerns. Dr. Pfenning has read and commented on this proposal in its entirety. This process has provided experience and a wealth of knowledge in grant writing that I will benefit from greatly as I continue to write grants, publications, and communicate science to a wider audience.

Accomplishing Proposal

From my prior research experience at the Lieber Institute for Brain Development and in my rotations in the Pfenning lab, I developed proficiency in neuroanatomy, preparing brain tissues for bulk tissue and single cell molecular biology assays required for experiments in **Aim 2a.** To obtain macaque brain single cell genomics datasets for **Aim 1a** and **Aim 2a**, I will continue to collaborate and consult with Dr. William Stauffer **(Letter of Collaboration)**, an expert in primate neurobiology who performs in snRNA-seq and snATAC-seq in macaque brain. In addition, I have consulted two postdocs Drs. Jing He and Esin Ozturk **(Letters of Support)**, the experts in single nucleus RNA-seq and single nucleus ATAC-seq molecular biology and non-human primate and rodent neurobiology. To collect and process human postmortem brain samples for **Aim 2a**, I am collaborating with Dr. David Lewis **(Letter of Collaboration)**, Chair of Psychiatry, an expert in schizophrenia neurobiology, and Director of the University of Pittsburgh Brain Tissue Donation Program, to identify neuropsychiatric control subjects without history of any substance use disorders for the proposed studies.

I have prior experience in genomics from my previous research experiences and have consulted two postdocs in the lab Dr. Irene Kaplow **(Letters of Support)**, an expert in comparative genomics and computational biology, for assistance in algorithms and cross-species comparative analyses of **Aims 1b** and **Aim 2b**. She will provide feedback on both rigorous and robust algorithm development and interpretation of genome conservation across primate and rodents. Furthermore, I will have support from my co-sponsor, Dr. Logan, an expert in mouse genetics, neuro-epigenetics, and addiction neurobiology, for assistance in interpreting the cell type-specific gene-regulatory mechanisms of addiction. Throughout my training, I will continue to have weekly group meetings as well as informal and formal one-on-one meetings with Dr. Pfenning where we discuss the data and algorithms and troubleshoot experimental and computational challenges. I will be primarily responsible for writing up the findings as a manuscript, and plan to present the discoveries at research conferences, such as Society of Neuroscience and the NIDA Genetic Consortium Meeting. Manuscripts and meeting presentations will initially be drafted by me with guidance from Dr. Pfenning and Dr. Logan, and my career advisor and department chair, Dr. Russell Schwartz. This diverse yet involved group of mentors consists of computational and neurobiologists at diverse stages of their careers who will provide me with outstanding academic and professional support.

Selection of Sponsor and Institution

As demonstrated by my past research experience and current career goals, I have a passion for and a long-standing dedication to a career in computational biology research and patient care. My careful selection of institution and primary research sponsor is a testament to my desire to pursue a career as a physician and computational neurobiologist.

Sponsor: Dr. Andreas R. Pfenning is the ideal mentor for an aspiring physician scientist. When selecting a research mentor, I sought out a well-established and funded Principal Investigator with a strong computational record, clear research goals to understand human health and disease, intimate knowledge of the MSTP structure and program goals, and a genuine desire to train creative interdisciplinary thinkers. Among Dr. Pfenning's many qualities as a mentor and leader are his scientific curiosity, creativity, and ability to communicate complex ideas clearly. As demonstrated by his excellent publication record, the Pfenning lab uses diverse computational and experimental methodologies to study fundamental questions about the epigenetic mechanisms encoded in the genome, including state-of-the-art machine learning and molecular biology tools. These complementary multifaceted approaches speak to Dr. Pfenning's rigor as a computational biologist. Furthermore, Dr. Pfenning employs a variety of strategies to aid his trainees in their development as well-rounded independent investigators—he encourages creative and rigorous scientific approaches, scientific communication, and professional development. He also never hesitates to help me build an informal mentorship network and has facilitated many introductions (in person as well as via email) with other PIs and potential collaborators. Dr. Pfenning is a highly qualified mentor who is clearly dedicated to training the next generation of computational biologists.

Co-sponsor: Dr. Ryan W. Logan is an outstanding scientist and mentor with an accomplished scientific record, having trained numerous graduate students. Furthermore, he serves as Internal Advisory Board member of the Jackson Laboratory Center for Systems Neurogenetics of Addiction. His laboratory has many years of experience using rodent models and human postmortem brains to study substance addiction behavior and genetics for which I will receive support and advice for the biological interpretations and consequence of **Aims 1c** and **2c** of this award **(Activities Planned Under this Award)**. Similar to Dr. Pfenning's mentorship, Dr. Logan, is readily available to provide academic and career advice. Given his expertise and mentorship experience, Dr. Logan is the perfect co-sponsor for this proposal and mentor for the development of my scientific career.

CMU-Pitt Computational Biology (CPCB) PhD Program: The CPCB comes from a collaborative network of dozens of training faculty spanning the Department of Computational Biology within the Carnegie Mellon University, School of Computer Science and the Department of Systems and Computational Biology within the University of Pittsburgh, School of Medicine. The CPCB innovative and rigorous interdisciplinary program has become one of the top computational biology programs in the country, as indicated by being chosen in 2005 by the Howard Hughes Medical Institute to receive one of only ten HHMI-NIBIB Interfaces Initiative awards. In 2008, it was also selected by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) to receive one of ten Interdisciplinary Training Grants.

Neuroscience Institute @ CMU/ Center for Neural Basis of Cognition (CNBC): The Neuroscience Institute and CNBC within it brings together faculty and students from across CMU to conduct multi-disciplinary work to advance the state of brain science. The CNBC, comprised of >150 faculty, leverages the strengths of CMU in cognitive and computational neuroscience and those of Pitt in basic and clinical neuroscience to support a coordinated cross-university research and educational program of international stature.

Institutions: The University of Pittsburgh (Pitt) and Carnegie Mellon University (CMU) Medical Scientist Training Program was truly the best fit for me. The main factors important to me when selecting these institutions were the availability of interdisciplinary neuroscience and computational biology research faculty and facilities, excellent clinical training faculty, and a well-established MSTP program with a strong history of alumni success, as demonstrated by residency matches and the proportion of alumni pursuing careers as physician scientists. This is not surprising, considering the MSTP curriculum is designed to provide the skills and training necessary for such a career path **(Additional Educational Information)**. For my graduate studies, I will complete two Longitudinal Clinical Clerkships (LCCs), both within Addiction Medicine. Upon completion of my PhD, I will enroll in the MSTP Clinical Reentry Elective, designed to facilitate the transition back to medical school. In addition, during my fourth year of medical school, I will have the option to complete the MSTP Postdoctoral Fellowship, a 5-month experience of nearly 100% dedicated research time (<u>not</u> included in time of covered support).

Training in the Responsible Conduct of Research

My proposed training plan includes courses, seminars, and professional development workshops that will enhance my understanding of the responsible conduct of research. The lessons and skills learned will aid my development into an ethical and thoughtful physician scientist. The format of my ethical training is as follows, in the order of completion: 1) medical school lectures and online coursework, 2) Medical Scientist Training Program (MSTP) ethics coursework, 3) MSTP workshops, and 4) graduate school ethics seminars and online coursework. **For each of these components, I will address the format, subject matter, level and form of faculty participation, duration of instruction, and frequency of instruction.** Over the course of the eight-year MSTP training, I will complete over 100 hours of ethics training.

1) University of Pittsburgh School of Medicine: Ethics, Law, and Professionalism Course in Patient, Physician, and Society block (completed 12/2017): This course met weekly for a 50-minute lecture from the course directors, and for a 50-minute small group discussion led by a School of Medicine faculty member. The course was 20 weeks in duration. Topics discussed included ethical recruitment and use of human subjects in research and ethical approaches to deliver patient care, among many others.

2) MSTP Professional Development 2 Course (completed 08/2018): In the summer of 2018, I completed this course as part of the MSTP curriculum. This course met eight times for 90-minute sessions that combined didactic lecture and small-group discussion led by the program director, Dr. Richard Steinman, and co-director, Dr. Russell Schwartz. Topics included rigor and reproducibility, falsification/fabrication of data, data documentation and ownership, mentor/mentee relationships, statistical rigor and appropriate scientific design of experiments.

3) MSTP Ethics for Medical Scientists (05/2020): I completed this course designed to be taken during the first spring of graduate school. This course consists of 4-5 two-hour long sessions. The goal of the multi-workshop style course is to provide a framework based on current methods and principles in ethics and to translate conceptual methodologies intro practical skills for evaluating ethical dilemmas.

4) MSTP Workshops (started 06/2017, ongoing): The MSTP program holds monthly gatherings for its entire student body. Each session is 90 minutes long and involves a combination of lecture and round-table discussion. Topics vary widely, but each session are initiated by student-identified needs in the training of tomorrow's physician scientists with appropriate faculty input and support. Each session includes ethics as a learning objective in research, medicine, and society.

5) Carnegie Mellon University CITI and IACUC Animal Research Online Classes (online completion 05/2020 and ongoing): Carnegie Mellon University IACUC provides a series of online training modules that are required for individuals performing research at the university as well as annual in person re-training of human and proper animal handling in biomedical research. I had to complete these online and in-person modules when I started in the MSTP in June of 2017 and I have re-certified the modules in 2020 and continue to do so as recommended by CMU's IACUC and animal facilities.

Role of Dr. Pfenning in Responsible Conduct of Research Training:

Throughout my graduate school training, I have had and will continue to have fruitful discussions with Dr. Pfenning regarding ethical issues in the conduct of research. In addition, the experience of performing research in his lab and our close mentor/mentee relationship will allow for robust training in the responsible conduct of research.

A. Research Support Available.

PFENNING ACTIVE:

1D1DA046585-01 PI: Pfenning 08/15/2018-06/30/2023 $300,000/year
NIH NIDA Interpreting the gene regulatory mechanisms underlying the predisposition to addictions disorders *The major goal of this project is to build a framework to study the function of both human and mouse brain enhancer regions *in vivo* to work towards deciphering biological mechanism underlying substance use disorders. This funded project is independent and complementary to the research aims proposed in this fellowship.

RFA-MH-19-135 PI: Stauffer, **Co-I: Pfenning** 07/01/2019-06/30/2024 $124,054/year
NIH NIMH A massive library of AAVs to target transcriptionally-defined primate cell types

LOGAN ACTIVE:
P50DA039841 PI**:** McClung, **Co-I: Logan** 10/1/2016-9/1/2021 $250,000/year
NIH NIDA Center for Systems Neurogenetics of Addiction

R33DA041872 PI: Logan 4/1/2018-3/31/2021 $370,000/year
NIH NIDA Generating novel mouse tools to investigate brain region and cell-type specific circadian molecular mechanisms of reward and motivation

R01HL150432 PI: Logan 9/23/2019-8/31/22 $300,000/year
NIH NHLBI Cell-type specific role of circadian-dependent transcription in fentanyl-induced synaptic and behavioral plasticity

R01 DA051390 PI: Logan 1/01/20-01/01/2024 $500,000/year
NIH NIDA Molecular rhythm alterations in human postmortem brain associated with opioid use disorder

B. Sponsor's/Co-Sponsor's Previous Fellows/Trainees

Dr. Andreas Pfenning's Trainees	**Current Position**
Meaghan Kennedy (master's student)	Doctoral candidate, University of North Carolina (Chapel Hill)
Sarah Hsu (master's student)	Computational Biologist, Massachusetts General Hospital
Nitinram Velraj (master's student)	Data Scientist, Astrazeneca
Siddharth Annaldasula (pre-doctoral student)	Fulbright Scholar, Master's student, Max Planck Institute

Dr. Ryan Logan's Trainees	**Current Position**
Zhuguang Huo (Co-mentored doctoral student)	Assistant Professor of Biostatistics, University of Florida
Kelly Cahill (Co-mentored master's student)	Systems Analyst
Puja J. Parekh (Co-mentored doctoral student)	Post-doctoral fellow at Weill Cornell Medicine
Xiyu Zhu (Co-mentored predoctoral student)	Doctoral candidate, University of Pittsburgh

C. Training Plan, Environment, Research Facilities

The primary goal of this proposal to train BaDoi to be an accomplished physician scientist at the intersection of machine learning, genomics, and neuroscience. With advances in high-throughput sequencing and single cell approaches, the need for broader training in computational techniques is widely recognized. Less recognized, but just as critical, is the need to train computational biologists to have deeper knowledge of specific biological domains. Making a transformative scientific advance doesn't just require the ability to develop a new algorithm or technique, but also enough insight into the biology to know the right questions to ask and to know how to interpret the output of those computational methods.

BaDoi's training prior to Pfenning laboratory provides a solid foundation in key areas. In his undergraduate education, he took courses related both neurobiology and computational biology from leaders in the field (including Steven Salzberg in computational biology). During his undergraduate/postbaccalaureate experiences, he was co-mentored by a computational biologist (Andrew Jaffe) and a neurobiologist (Brady Maher). These experiences prior to his thesis research, especially the research on Autism, demonstrate a commitment to a career in addressing mental health by combining computational genomics and neuroscience. From the perspective of techniques, BaDoi's ability to conduct bulk tissue RNA-Seq and that data also provide a solid foundation to conduct the proposed research. To build upon these experiences and skills, BaDoi's training through my laboratory will focus on two key areas: machine learning for computational genomics, the cell types and neural circuits underlying addiction.

The cornerstone of BaDoi's education is his experience in the Pfenning laboratory. The laboratory develops computational methods as well as conducts high-throughput genomic experiments. BaDoi will participate in weekly broad lab meetings with two presenting researchers. In addition, there are two smaller group meetings to get more targeted feedback. In a weekly experimental genomics meeting that I attend, 3-5 lab members discuss research plans troubleshoot any difficulties with their experiments. Similarly, in weekly computational meetings I attend, 4-6 lab members discuss progress in applying and/or developing machine learning and other methods for genomics. In addition to attending both of these, BaDoi and I will hold regular 1-1 meetings to discuss research updates and plan new experiments. BaDoi will have full access to the animals, equipment, facilities, reagents, and computing resources required to complete his proposed aims.

Machine learning for computational genomics. The training in machine learning for computational genomics will build upon BaDoi's prior training in computational biology. It will provide a foundation for making scientific advances by using cutting edge computational techniques to extract knowledge out of the complex high-throughput genomic data collected from the brain regions of the reward system **(Aim 1b and 2b)**.

Training opportunities. The Carnegie Mellon School of Computer Science and University of Pittsburgh School of Medicine share a Ph.D. program (Joint CMU-Pitt Ph.D. Program in Computational Biology, CPCB). This program leverages Carnegie Mellon's strength in machine learning, the University of Pittsburgh's strength in translation and both institutions strength in genomics. Also within the Carnegie Mellon School of Computer Science is the Machine Learning Department, providing further opportunity training. Every student in this program, including BaDoi, takes Ph.D. level machine learning at Carnegie Mellon School of Computer Science (10-701). That course is a co-requisite of Ph.D. level computational genomics (02-701). In addition, we have identified the machine learning course, probabilistic graphical models (10-708) as being relevant to the algorithm design component of the training. During his training experience, BaDoi will attend the weekly computational biology seminar as well as the weekly machine learning seminars that focus on applications to science. The CPCB graduate program has integrated monthly professional development activities into its curriculum including opportunities for graduate student presentations, grant writing mentorship, and online professional networking.

Research environment. The Computational Biology Department at Carnegie Mellon provides a fantastic research environment for machine learning applied to computational genomics. This includes Dr. Ziv Bar-Joseph, Dr. Eric Xing, Dr. Jian Ma, and Dr. Seyoung Kim. Additionally, the new department head Dr. Russell Schwartz (BaDoi's MSTP Career Advisor) and affiliated professor, Dr. Joel McManus, are integrating machine learning techniques with experimental approaches. BaDoi has full access to the resources of the Computational Biology Department at Carnegie Mellon and the Pfenning laboratory. The Carnegie Mellon Computational Biology Department has a sizable shared cluster facility (50 nodes), which include a large memory (512GB) node, and both CPU and GPU resources. The Pfenning laboratory itself has dedicated CPU, GPU nodes, and a large memory (512GB) node. These are more than sufficient to analyze most datasets collected in our laboratory and the laboratories of our collaborators. To scale our algorithmic approach and tune parameters, we currently have XSEDE resources through the Pittsburgh Supercomputing Center (Grant MCB190162 "Investigating the Evolution, Spatial Profiles and Gene Therapy Access Points of Neuronal Cell Populations Using Single Cell Sequencing and Analysis Approaches" **(Facilities and Resources)**.

Training and career goals. When computational approaches to molecular biology were still in their infancy, most of the field took an interdisciplinary approach: a computer scientist would collaborate with an expert in an area of biological sciences to solve a particular problem. Even today, a large proportion of computational biologists develop methods that are applied to wide range of biological problems. However, as both computation and genomic science advances, the training of transdisciplinary scientists that are expert in computational approaches as well the sub-discipline they are applying computational approaches to will be key. As a physician-scientist, training in machine learning and computational genomics will allow BaDoi to recognize the limitations of the data analysis that others have conducted in addiction. Furthermore, he will be able to address those limitations by developing powerful new approaches that are tailored specifically to the challenges of studying the brain and addiction.

Skills and Techniques. As the field of applying machine learning to computational biology is quite broad, BaDoi will focus on learning techniques that are specifically relevant to single cell biology and epigenetics. He will learn how to process and carefully analyze single nucleus RNA-Seq and single nucleus ATAC-Seq datasets. It has been a pleasure to mentor him thus far on preliminary analysis of the single nucleus RNA-Seq and how to connect those results to genome-wide association studies **(Research Strategy)**. These techniques are crucial for disentangling the cellular heterogeneity of the brain and how those cell types are involved in addiction. The

ability to translate findings between rodent models and humans requires substantial knowledge in comparative genomics. In collaboration with postdoctoral fellows in the group (such as Irene Kaplow, **Letter of Support**), I will train BaDoi in comparative genomics and epigenetics. The machine learning techniques we are applying to comparative genomics include the ability to train support vector machines and convolutional neural network models of orthologous regions of genome sequence between rodents and primates. Finally, BaDoi will receive training in graphical models more broadly, which will allow him to develop new techniques to model how gene expression varies across complex combinations of brain regions, cell types, species, and conditions, including different stages of the addiction process.

The cell types and neural circuits underlying addiction. Training the biology of addiction will provide a foundation for career in applying computational genomics and machine learning to understand the neurobiology of addiction and other brain disorders. At the molecular level, this involves conducting genomic experiments on specific cell types in the brain. More broadly, it involves building an understanding of those cell types connect to each other in neural circuits to create behavior.

Training opportunities. As formal training, BaDoi has identified the neuroscience course, Advanced Cellular Neuroscience (03-762) taught by a basal ganglia neurocircuit expert Dr. Aryn Gittis as relevant to understanding the basic cell types and neural circuits of the reward system. The Pfenning laboratory takes part in monthly research presentations given by neuroscience graduate students and post-docs in neurobiology laboratories of the Mellon Institute Building at Carnegie Mellon. In addition to attending these meetings, BaDoi will present his research finding at several points throughout his Ph.D. His first presentation on addiction genetics was on June 5th, 2020. Furthermore, BaDoi will participate in weekly meetings with the Stauffer group as a part of a funded collaboration ("A massive library of AAVs to target transcriptionally-defined primate cell types"). Here, BaDoi will be able to receive a world-class education and expert feedback on the neural circuitry of the reward system in primate models. Finally, BaDoi will present work-in-progress meetings in his co-sponsor Dr. Logan's lab and receive critical feedback on the neural circuitry of the reward system in rat and mouse, especially integrating findings from genetic experiments to map heritability of drug addiction behaviors. This feedback will be especially critical during Year 1 and 2 of this award to relate findings from **Aims 1c** and **2c** to the cell types and neural circuits underlying addiction. As a faculty advisor for several trainees in the Translational Neuroscience Program at the University of Pittsburgh, Dr. Logan is primed to advise BaDoi on translating these findings from **Aims 1 and 2** to human neurobiology of addiction.

Research environment. The research environment for neuroscience at Carnegie Mellon is excellent. The Pfenning laboratory forms a strong collaborative community with neurobiology faculty (Sandy Kuhlman, Aryn Gittis, Alison Barth, Eric Yttri, and Kate Hong) in the department of Biological Sciences. The Pfenning laboratory's space is just around the corner and across the hall from space shared by Aryn Gittis, Sandy Kuhlman, and Alison Barth. In parallel to the close ties in computational biology between Carnegie Mellon and University Pittsburgh, there the close ties in neuroscience as well. The Center for the Neural Basis of Cognition, joint between Carnegie Mellon and University of Pittsburgh, is an internationally renowned force in neuroscience research. The center has provided a platform for cross-institute collaborations, including joint projects between the Pfenning laboratory and the Stauffer and Logan labs. In collaboration with the Logan lab, BaDoi was involved in feasibility experiments and analyzed bulk ATAC-seq in cortex, dorsal striatum, and nucleus accumbens of two mouse strains with distinctly unique addiction phenotypes. Within the Pfenning laboratory itself, BaDoi's will have access to the required experimental infrastructure. We regularly conduct ATAC-Seq and RNA-seq and have established protocols and reagents available for fresh and frozen tissue (several of which BaDoi has piloted) as well as the equipment to carry out the experiments. For single cell experiments **(Aim 2a)**, the Pfenning laboratory hosts and maintains a 10x Chromium machine for single cell genomics. BaDoi has already successfully conducted experiments using this technology with help from collaborators in the Stauffer group. To receive quick feedback for experimental optimization, we also host and maintain an Illumina Mi-Seq machine. The Pfenning laboratory also has an animal colony maintained by Carnegie Mellon that BaDoi has the training and access to use for these experiments. Additional equipment and resources are detailed in Equipment and Facilities and other Resources.

Training and career goals. A major challenge in the study of psychiatric and neurological disorders is linking underlying genetic and molecular mechanisms to addiction behaviors observed in a clinic or a laboratory. Recently, single cell and cell and cell type-specific experiments provide an avenue to accomplish this more robustly. As BaDoi's preliminary analysis shows, genetic variants often have cell type-specific effects, which will influence particular neural circuits, which in turn, can lead to behavioral differences **(Research Strategy)**. Thus, neural circuits have the potential to bridge the gap between cell type-specific mechanisms of addiction and

behavior, with neural circuits. With this in mind, a deep understand of addiction circuitry will provide BaDoi with a foundation to connect genomic experiments and computational analysis he conducts to the clinic. Furthermore, to be able to accurately model the nuances and confounds of the cell types in the reward system, the scientists developing the computational approach should be able to directly interpret their results in the context of the field. Although BaDoi has prior experiences in relating cell type-specific genomics from rodent models of human neurodevelopmental disorders **(Other Research Experiences)**, he will receive additional training from his co-sponsor Dr. Logan. BaDoi will learn the technical and experimental designs of the genetically diverse and phenotypically heterogeneous mice used by the Center for Systems Neurogenetics of Addiction (CSNA) to map mouse loci associate with addiction behaviors and interpret enrichment experiments proposed in **Aim 2c**. He will meet one-on-one with Dr. Logan to review the addiction circuits literature and present work-in-progress updates to the Logan lab and incorporate critical feedback on implications of results from **Aim 2**. Thus, despite having little formal training in the rodent models and neural circuits of addiction, BaDoi has identified experts who have agreed to formally train him to learn the nuances of the field and provide resources from the CSNA to accomplish **Aim 2** of his proposal.

Skills and Techniques. Complementary to BaDoi's broader training on neural circuitry, he will continue to develop as experimental genomic scientist. Dr. William Stauffer **(Letter of Support)**, Dr. Leah Byrne, and myself, along with others in our laboratories (Jing He, Esin Ozturk, **Letters of Support)**, will train BaDoi to conduct single nucleus RNA-Seq and ATAC-Seq. His successful preliminary experiments presented **(Research Strategy)**, were primarily through my mentorship, but involved numerous helpful suggests from Dr. William Stauffer, Leah Byrne, and Jing He. These skills will build upon his undergraduate/postbaccalaureate foundation in conducting RNA-Seq and ATAC-Seq in bulk tissue samples as well as his preliminary research.

Mentorship team and roles. Due to BaDoi's focus of his proposal on computation and human genetics, he and I have thoughtfully included supportive expert mentors in his thesis committee and training program to guarantee he has a foundation in the design and theory behind his research aims. BaDoi's thesis committee includes myself, his co-sponsor Dr. Ryan Logan (addiction neurobiology), Dr. Dennis Kostka (single cell genomics algorithms), and Dr. Kathryn Roeder (statistical and human genetics). BaDoi will receive mentorship from Dr. Roeder, a leading expert in statistical methods and human genetics specifically in neuropsychiatric diseases and disorders. She lends vast expertise that will ensure thorough analysis of human GWAS risk variants, which will be key to BaDoi's aims.

Thesis Committee Member (NRSA Role)	Title
Dr. Andreas Pfenning, PhD (Sponsor, CMU)	Assistant Professor, Department of Computational Biology
Dr. Ryan Logan, PhD (Co-sponsor)	Assistant Professor, Department of Psychiatry
Dr. Dennis Kostka, PhD (comp algorithms, Pitt)	Associate Professor, Department of Developmental Biology
Dr. Kathryn Roeder, PhD (statistics and human genetics, CMU)	UPMC Professor of Statistics and Life Sciences, Departments of Statistics and Computational Biology

Clinical Training

The Longitudinal Clinical Clerkship (LCC) comprises 20-week clinical rotations for one half-day per week to provide 1) direct experience balancing both clinical and scientific work and 2) develop valuable clinical skills in BaDoi's field of interest, addiction medicine, during a scientifically rigorous part of his overall training. The LCC will provide BaDoi with the clinical context for his science research as he begins to think about integrating his research and clinical work, a crucial aspect of his planned career as a physician scientist performing translational research. The Department of Psychiatry at the University of Pittsburgh Medical Center currently ranks 7th nationally for residency training, a tribute to the quality of clinical training at the University of Pittsburgh. Although BaDoi has yet to begin his LCC, he intends to start his first LCC with Dr. Antoine Douaihy. Dr. Douaihy is not only an expert in addiction psychiatry but he is also a highly accomplished clinician who manages patients with substance use disorders. As an academic psychiatrist, Dr. Douaihy would be a fantastic role model and mentor for BaDoi, who aspires to become an addiction medicine clinician. Dr. Douaihy, therefore, would serve as an excellent mentor for bridging BaDoi's scientific and clinical interests as well as contributing to the development of his clinical skills in substance use management.

D. Number of Fellows/Trainees to be Supervised During the Fellowship

PI Pfenning plans to support 1 postdoctoral fellows during BaDoi's proposed NRSA fellowship period.

PI Pfenning plans to support 4 pre-doctoral trainees during this period, (3 students not including the applicant).

PI Logan plans to support 2 pre-doctoral students and 1 post-doc during this period.

E. Applicant's Qualifications and Potential for a Research Career

Sponsor: Andreas R. Pfenning, PhD; Assistant Professor of Computational Biology

When BaDoi joined my laboratory, he already had substantial undergraduate experience in using RNA-Seq to measure levels of gene expression in the brain across the genome. This undergraduate research was published in *Nature Neuroscience*, one of the top journals in the field. Impressively, he had substantial experience in conducting the experiments, in the pipelines to process the data, and in the statistical techniques to analyze the results. To build his skills in genomics, neurobiology, and analysis, he focused his rotation on measuring open chromatin by conducted the ATAC-Seq assay on mouse brain tissue. We are in the process of testing the hypothesis that the epigenomic state of cells breaks down during the aging process. BaDoi's computational and experimental skills helped to provide the foundation for a grant on the subject that was funded by the AFAR foundation: (https://www.afar.org/grantees/years/2018-recipients/).

For his thesis project, BaDoi is combining his medical training and computational skills to tackle addiction, an enormous problem that our society faces. The study of addiction has led to many insights on how specific neural circuits mediate associated behaviors. In parallel, new genetic studies are identifying hundreds of locations across the human and mouse genome that are associated with different aspects of addiction. BaDoi's project is essentially to connect the genetics to the neural circuits by combining new genomic experiments on the key brain regions with new machine learning models. By studying how naturally occurring mutations influence the predisposition to addiction, we hope to develop treatments that mimic those effects.

Although he has only been in the laboratory for less than a year, he has already accomplished an extraordinary amount. He has identified that different components of addiction (initiation of the behavior vs. cessation) are associated with different brain regions and different cell types within those brain regions. I gave a well-received talk at the NIDA genetics meeting in January based on these findings and submitted this work as co-first author for review at the *Journal of Neuroscience*. Furthermore, he has contributed to our collaboration with the Logan lab to link differential gene expression in the nucleus accumbens of subjects with opioid use disorder with to a number of substance use and risky behavior GWAS loci, which is currently in review. In parallel, he has conducted a difficult experimental procedure – single nucleus RNA-Seq – to follow up on those computational predictions.

BaDoi's ambition, flexibility, and focus are incredible. In our experimental meetings, he often chimes in with knowledge of the latest (sometimes unpublished) genomic techniques to measure gene expression or the epigenome in small numbers of cells. He has the mindset of a problem solver when it comes to designing new experiments. Similarly, he provides valuable feedback and direction on the statistics as well. On numerous occasions, he has explained important statistical concepts to undergraduates and graduate students in the laboratory. For example, in a meeting on aging epigenomics, he explained to an undergraduate how an interaction term could capture different rates of change with age across different segments of the population.

Co-Sponsor: Ryan W. Logan, PhD; Assistant Professor of Psychiatry

It is my pleasure to write this letter of support for the F30 application for BaDoi as well as confirm my willingness to serve as a co-sponsor on this exciting project. BaDoi has taken a perplexing clinical observation and used it to generate an interesting hypothesis that he has proposed to systemically address in a series of parallel but complementary approaches. Importantly, he will elucidate cell type-specific gene regulatory mechanisms conserved between primate and rodent. Results from each approach will not only inform the results of the other but also add to the existing literature in each of these areas. Furthermore, understanding the underpinnings of genetic manipulations in powerful rodent models of addiction is a critical step forward to improving our understanding of substance use disorder. A better mechanistic understanding of how to translate findings to human addiction neurobiology will be bring much insight to the field.

BaDoi has been able to make a strong impression and set high expectations for himself in the short period of time that he has been in the Pfenning Lab. The quality of the publications he contributed to his undergraduate and post-baccalaureate research speak to his abilities and potential as a scientist. He started this proposal process with a number of very good ideas and was able to engage from the start in discussions about how best to focus these ideas into tractable aims. He would go back to the literature and suggest alternative approaches when a clear path forward was not always obvious. He was always responsive to feedback, and in the end was

able to generate an outstanding project that will not only serve as an excellent training vehicle but, as noted above, will yield insight into the neuronal subtypes and gene regulators underlying addiction genetic variation.

As his co-sponsor, I will be actively involved providing support and guidance as he works his way through the proposed experiments as well as toward his career goal career goal to become a proficient scientist in the addiction genetics field. My years of research and mentorship experience will enable me to provide an excellent complement to the mentoring provided by his primary sponsor Dr. Andreas Pfenning.

As an independent investigator, I have successfully mentored 1 master's student, 1 doctoral student, and currently have 1 post-doctoral fellow in the laboratory. I am an active member of the Translational Neuroscience Program (TNP) and actively participate in collaborative lab and journal club meetings. Furthermore, I actively participate in training graduate students as a mentor and through supportive roles; my lab provides conceptual and methodological expertise on circadian rhythms, genetics, and addiction for several graduate trainees and post-doctoral fellows in TNP. BaDoi's scientific interests and goals in this proposal align quite well with my expertise and laboratory's research endeavors. The active and productive collaborations I have established with clinicians and MD/PhD faculty in the Department of Psychiatry, such as Dr. David Lewis, Dr. Zachary Freyberg, and Dr. Dan Buysee, should serve as an excellent backdrop for BaDoi's endeavors as he begins his career as a clinician scientist. Furthermore, my expertise in rodent models of addiction and genetic experiments to map high-quality mouse loci to differences in addiction behavior will directly benefit BaDoi's scientific and technical training. BaDoi has outlined a rigorous approach to his research proposal which will enable him to add to our understanding of conserved gene regulatory mechanisms of molecularly defined cell types of the reward system.

While I appreciate the tremendous hurdles facing clinicians in today's shifting healthcare environment, to say nothing of the additional hurdles faced by clinician scientists with additional constraints on time and money, I am particularly optimistic that BaDoi will be one of the few to succeed in this challenging environment. He came to the MSTP program with a strong commitment to research, having already made significant contributions to science. Having already mastered several powerful techniques, he clearly picks things up quickly and is able to apply these techniques to achieve publishable results.

BaDoi is just the type of person we need working on his long-term goal of marrying computational and genomic approaches to understand the implications of human genetics on neuropsychiatric conditions. While his proposal is admittedly ambitious, I believe it is also entirely feasible, both because of the speed and efficiency at which BaDoi has demonstrated he is able to work, and because all of the techniques to be employed are up and running, if not in Dr. Pfenning's lab, in one of the labs of his and BaDoi's collaborators. All of this, together with a focused training plan, make for an excellent training potential. Moreover, BaDoi has established a mentoring team that will ensure the successful completion of the experiments proposed and help establish a foundation upon which BaDoi will be able to build a successful career as a clinician scientist. With all this taking place within the Computational Biology PhD Program and CMU Neuroscience Institute, housed in institutions clearly committed to the success of their trainees, this is the ideal place for him to complete this training.

I look forward to working with BaDoi on this exciting project and hope the reviewers agree that this project will not only serve as an outstanding training vehicle for an exceptionally promising young scientist but will also generate data important to our understanding of several fundamental processes.

I thank the reviewers for their time and consideration.

Sincerely,

Ryan Logan, PhD

Assistant Professor of Psychiatry

University of Pittsburgh

April 8, 2020

David A. Lewis, MD
Distinguished Professor of Psychiatry and Neuroscience
Thomas Detre Professor of Academic Psychiatry
Chair, Department of Psychiatry

Medical Director and Director of Research
UPMC Western Psychiatric Hospital

UPMC Western Psychiatric Hospital
3811 O'Hara Street
Pittsburgh, PA 15213-2593
Phone: 412-246-6010
Fax: 412-246-6780
lewisda@upmc.edu

BaDoi Phan
PhD Student, Department of Computational Biology, Carnegie Mellon University MSTP Student, University of Pittsburgh School of Medicine
Pittsburgh, PA 15213

Re: NRSA Support Letter

Dear BaDoi,

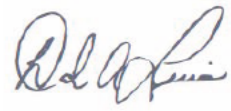
I am happy to serve as a collaborator on your F30 project with Dr. Pfenning and Dr. Logan, "Integrating primate-rodent cell types and epigenomics to identify conservation in substance addiction". The set of experiments you propose examining the gene marker and enhancer profiles of conserved cell populations between human, macaque, rat, and mouse, to in turn identify the cell type-specific contributors to substance use disorder heritable genetics is a novel and exciting extension of our research. In addition, your proposal has obvious clinical relevance to translate findings in model organisms towards human disease states. Lastly, I have been a close collaborator with your sponsors, especially Dr. Logan, and I know you will receive excellent training and mentorship working with them.

As you know, I am Chair of the Department of Psychiatry at the University of Pittsburgh, Director of the Translational Neuroscience Program and Director of the Brain Tissue Donation Program. Our bank of postmortem human brains is widely regarded as composed of specimens of very high quality with extensive clinical characterization. The results from multiple other studies using samples from our bank align with your goals to further apply single-nucleus genomics to expand our understanding of the conserved molecular profiles in the reward system. Despite the well-described anatomical approaches to investigate cell types of human brains, comparatively few studies have focused on the cross-species conservation using these single-cell gene expression and epigenetic profiles.

I will work with you in the selection of subjects comprised of subjects without any history of neuropsychiatric diagnoses or substance use history. The designated brain regions for the study (dorsolateral prefrontal cortex, caudate, putamen, and the nucleus accumbens) will be carefully dissected and provided to your laboratory for further processing to conduct high-resolution single-nuclei gene expression and open chromatin experiments. We will also provide all necessary information on each subject, including demographic features, life history derived from psychological autopsy findings, neuropathology exams, and brain toxicology studies. It is my understanding that the funds required to support these activities will be provided by your mentors.

Furthermore, my own career integrates basic science work with clinical practice focused on studying and caring for patients with schizophrenia and other psychiatric disorders. I will be happy to support your training as a physician scientist by providing advice and guidance on this career trajectory, as I know you plan to pursue a similar path combining basic neuroscience research and clinical practice.

Sincerely,



David A. Lewis, MD

University of Pittsburgh

April 8, 2020

William R. Stauffer
Department of Neurobiology
School of Medicine

4066 Biomedical Science Tower 3
3501 Fifth Ave
Pittsburgh,PA 15261
Tel: (412)-383-6563
Fax: (412)-648-1441

BaDoi Phan
PhD Student, Department of Computational Biology,
Carnegie Mellon University
MSTP Student, University of Pittsburgh School of Medicine
Pittsburgh, PA 15213

Dear BaDoi,

I am writing to express my enthusiastic support for your F30 fellowship application, entitled 'Integrating rodent-primate cell types and epigenomics to identify conservation in substance addiction.' This work, done with your PhD mentor Andreas Pfenning, has the potential to provide transformative insights into the molecular control of function and dysfunction in neural reward systems.

As you know, Andreas Pfenning and I have been collaborating for the past two years to classify cell types and identify cell type-specific regulatory elements in nonhuman primate (NHP) brain. Together, we have a BRAIN-Initative funded grant to support this work, and we recently submitted a manuscript that uses single nucleus RNA-Seq (snRNA-Seq) to characterize novel cell types in the Rhesus macaque striatum. Andreas is an excellent scientist and an excellent mentor. Moreover, it is clear to me – from your ongoing participation in joint collaboration meetings and from your excellent publication record – that you are an experienced and ambitious scientist-trainee with great potential. You have a firm grasp of the molecular biology and computational literature relevant to your project, and your insights into my research questions have been valuable.

Your project, using single nucleus RNA-seq and ATAC-seq in rodents and primates to identify conserved cellular and epigenetic programs in the reward system, is incredibly exciting. This project represents a novel application of machine learning to the neurobiology of addiction. This approach has significant promise in translating knowledge gained form rodent and NHP studies to human addiction. As you know, I have a longstanding interest in understanding the neurophysiological computations related to rewards, learning, and decision making. I am excited to assist your research and to benefit from it. I think your approach will identify cell type specific regulatory elements that will be valuable for systems neuroscience research and future translational applications.

My lab continues to do single nucleus RNA Seq and ATAC Seq in primates. I will provide you with access to single cell data for your computational analyses to find conserved cell types and enhancers between rodents and primates. I think it will provide us excellent insights into primate specific innovations that enable sophisticated cognition and behavior. Do not hesitate to contact me with any questions related to data acquisition or access, I am happy to help.

Sincerely,

William R. Stauffer, PhD
Assistant Professor, Department of Neurobiology
Principal Investigator, Systems Neuroscience Center
Training Faculty, Center for Neuroscience, and
Center for the Neural Basis of Cognition
University of Pittsburgh Brain Institute

University of Pittsburgh

Bilge Esin Ozturk
Postdoctoral Associate
Department of Ophthalmology,
University of Pittsburgh School of Medicine
3501, 5th Ave, BST3-10051, Pittsburgh, PA 15213
Tel: (412) 726-3782
E-mail: esin.ozturk@pitt.edu

March 30, 2020

Re: NRSA Support Letter

Dear BaDoi,

I am happy to provide advice to help support your use of single nucleus genomics technology for your F30 project, "Integrating rodent-primate cell types and epigenomics to identify conservation in substance addiction". I believe that the findings from this project will be of great help to the field and also have a huge impact on your training. Your proposed work is very similar to my prior work in collaboration with your sponsor, Andreas Pfenning. Your approach using single nucleus RNA-seq and single nucleus ATAC-seq in neural tissues across species seems feasible to collect with the network of support available to you, including me, and I will happily help you troubleshoot your assays as you begin to collect postmortem human brain tissue in particular.

As a molecular biologist and as a postdoctoral fellow working in a gene therapy lab, I have significant experience in single cell RNA-Seq, single nucleus RNA-seq, single nucleus ATAC-seq, as well as working with fresh/frozen tissue obtained from various species including human, non-human primates (NHPs), mouse, etc. As you know, I have been working in different projects to create single-cell transcriptome maps of primate brain and retinal cells and identifying NHP cell type specific enhancer/promoter constructs. I believe I have the expertise to help you identifying conserved cell types and marker genes for human SUD risk variants and I will be happy to assist you and provide any advice or input you need.

Since we first met, we have met many times to talk about your interest in translational neuroscience and applying computational techniques to learn more about the brain. I strongly believe that your determination, creativity and experience will enable you to successfully achieve your goals in this exciting project. I look forward to our continued collaboration and I will be more than happy to provide any support that you will need.

Sincerely,

Bilge Esin Ozturk, PhD
Department of Ophthalmology, School of Medicine
University of Pittsburgh

University of Pittsburgh

Department of Neurobiology
Systems Neuroscience Center
Center for the Neural Basis of Cognition

3501 fifth ave, BST3 4078
Pittsburgh, PA 15213
Tel: (412)613-4163
E-mail: jih118@pitt.edu

Jing He
Post-Doctoral Fellow

April 5, 2020

Re: NRSA Support Letter

Dear BaDoi,

I would be very happy to be a consultant on your project with Dr. Pfenning, titled "Integrating rodent-primate cell types and epigenomics to identify conservation in substance addiction". The project not only integrates genomic data that are collected from four species, and the expected results will yield important information regarding the conserved gene regulation of the reward system.

As you know, I have been conducting single cell RNA-seq, single nucleus RNA-seq, and single nucleus ATAC-seq as a molecular neurobiologist and postdoctoral researcher in Dr. Stauffer's group. We demonstrated these assays in cortex and striatum of macaque, marmoset, and rat etc. I am interested in cell types of the reward pathway and am very willing to advise you on single nucleus experiments to augment these data that we have already collected. Please feel free to reach out for any questions you might have as you grow your expertise as a molecular biologist in brain genomics.

Please do not hesitate to contact me if you have any questions or need more guidance on single-nucleus RNA-seq and ATAC-seq. Good luck with your application!

Yours sincerely,

Jing He, PhD
Department of Neurobiology
Systems Neuroscience Center
Center for the Neural Basis of Cognition
University of Pittsburgh

Carnegie Mellon University

Department of Computational Biology
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213-3890

March 31, 2020
Irene M. Kaplow
Gates-Hillman Building, Room 7703
5000 Forbes Avenue
Pittsburgh, PA 15217
ikaplow@andrew.cmu.edu

Dear BaDoi Phan,

I would be delighted to help teach you how to train machine learning models in a way that best leverages regulatory genomics data from multiple species. I have extensive experience training machine learning models using regulatory genomics data, interpreting them, and using them to make predictions in new contexts, both as a Computer Science Ph.D. student co-advised by Anshul Kundaje and Hunter Fraser at Stanford University and as a Lane Postdoctoral Fellow advised by Andreas Pfenning at Carnegie Mellon University. Specifically, I have trained support vector machines and convolutional neural networks to predict brain regulatory element ortholog activity across hundreds of species. Thus, I have extensive experience in machine learning and specifically in using it to trace the evolution of regulatory elements.

In addition to having extensive machine learning experience, I also have substantial experience teaching machine learning to others. As a Ph.D. student, I served as a teaching assistant for Stanford's graduate level courses in machine learning and probabilistic graphical models, which involved mentoring teams of students on final projects involving applying the methods from the class to real-world problems. I also served as an instructor at Ben-Gurion University's Center for Evolutionary Genomics and Medicine's Genomics Workshop, where I taught experimental biologists how to train convolutional neural networks using regulatory genomics data. In addition, I have mentored eleven students in regulatory genomics research projects.

I am excited about teaching you how to use machine learning to leverage multi-species regulatory genomics data because you are a fast and enthusiastic learner. I have already begun teaching you about using convolutional neural networks for genomics, and I was impressed with how quickly you were able learn and use what you learned to develop new ideas. I look forward to helping you to apply my skills to better understand transcriptional regulatory mechanisms underlying substance use disorders.

Sincerely,

Irene M. Kaplow
Lane Postdoctoral Fellow

Additional Educational Information (by Richard Steinman MD PhD, Director MSTP, Steinman@pitt.edu)

A. University of Pittsburgh-Carnegie Mellon University MSTP Structure. MSTP students in Pittsburgh complete a MSTP-specific enrichment curriculum beyond the standard courses in medical and graduate school. This consists of 3 summer research rotations, 3 summer professional development courses, a 3-semester weekly journal club featuring research papers consistent with the coincident SOM curriculum, a 4-week case-based ethics course, a monthly program-wide workshop, a 40-week longitudinal clinical clerkship (1/2 day/week) during the graduate years, a 11 day Junior Hospitalist service and yearly special events such as the two-day MSTP Scientific Retreat.

B. Laboratory Research Rotations. Research rotations begin the summer prior to the start of medical school. In addition to developing manuscripts and presenting at scientific meetings based on their rotation results, all students turn in a written scientific report that is reviewed by MSTP leadership and present their work at the annual MSTP Scientific Retreat. The choice of thesis laboratories by students is informed by their rotation history and by discussion with their individual Career Advisors (who follow them longitudinally in the program).

C. Professional Development. Students take three successive 10-week long Professional Development Courses during summers prior to starting graduate school. The first course (PD1) focuses on scientific writing and introduces students to biomedical software and to key methods used by different disciplines to approach scientific problems. The PD2 course focuses on scientific design and career development strategies, with particular emphasis on reproducibility and biostatistics. The PD3 course focuses on grant review and writing.

D. Training in Reproducibility in Science. The PD2 course focuses on optimizing reproducibility of findings, to power experiments, and analyze data with appropriate statistical testing. Topics for classes include problems arising from non-reproducible work, optimal experimental and reagent documentation and handling, the ARRIVE guidelines for animal work, measurement validity and sources of error, robust hypothesis testing, and a series of sessions on biostatistics including customized problem solving tied to student data.

E. Biomedical and ethical expertise. During MS1 and 2 years, students build biomedical knowledge through a 3-semester MSTP literature review course in which students present papers after formal consultation with local faculty experts in the field of that paper. During the G1 year of graduate school, MSTP students take a month-long, weekly, case-based research ethics course. Throughout both medical school and graduate school, all MSTP students meet monthly for student-arranged seminars that pose scientific, logistical, clinical and/or ethical dilemmas. These workshops are presented by students and/or guest faculty experts.

F. Clinical and Research Integration. This is a central focus to better model the physician scientist career.

F.1 Clinical Activities During the Research Years. Prior to starting graduate school, all MSTP students complete 8 weeks of required clinical core clerkships. This front-loads requirements once students re-enter medical school post thesis and enables research engagement in MS3 and 4. MSTP students are required to complete a (credited) minimum of two 20-week long Longitudinal Clinical Clerkships during graduate school. For each LCC, students spend a half day per week with a clinician scientist receive one-on-one clinical mentoring by a clinician scientist in an area of interest chosen by the student with guidance from the MSTP LCC director, Paul Monga, MD. Student objectives for the LCC and write-ups at the end are reviewed by MSTP leadership.

F.2 Transition from Graduate to Clinical Years. After the student's thesis defense but prior to returning to medical school, students take the MSTP required Junior Hospitalist Service, also known as the LCC3. A master clinician mentors the returning students for 11 days as they examine, discuss, diagnose and plan treatment for surrogate patients presenting with common outpatient or inpatient ailments.

F.3 Research During Clinical Years. Our students continue their research focus after re-entry to medical school generally in four ways: (1) MS3 and MS4 students continue to plan and execute MSTP Workshops that feature research topics and research challenges to be discussed with MSTP peers. (2) Students complete formal reflective and goal-oriented self-assessment evaluations during twice-yearly Career Advisor meetings. (3) Students average 2.8 new publications during the MS3 and MS4 years (at least one first authored), averaging 5-7 papers upon graduation. (4) Most students elect to take 1-2 Research Elective months during their MS4 year to extend findings of thesis work and/or to build skillsets in a translational area. Another novel feature of our MSTP, the Postdoctoral Fellowship, provides support for 5 months of postdoctoral research prior to residency for MSTP students graduating in December (25% of graduates in recent years). Applications address research hypotheses and aims, career development aims, planned deliverables, mentor fit and intellectual goals.

G. Monitoring and Evaluating Student Progress. Prior to matriculation, the Program Director assigns each new student a Career Advisor based on matching research interests who help orient and guide the students throughout their careers. Most of a trainee's time in the graduate program is spent in research training under the guidance of their research mentors, program leadership, and eventually their doctoral dissertation committee. To customize advice and resource allocation, all MSTP students complete and share *individual development*

plans with the Director and with their Career Advisor. The form allows students to identify specific skills that they want to develop, set technical, intellectual and professional goals, and identify how goals will be achieved and measured. Resources to reach goals and obstacles that could compromise success are enumerated and discussed. Progress toward goals is regularly reviewed with the Advisor and new goals are set.

H. Career Counseling. To better reflect the student's educational experience to prospective residency programs, the MSTP creates an executive summary which describes student evaluations, honors, presentations and participation in the combined degree training, rewarding students who altruistically give their time and demonstrate prowess in working in groups. Six months to one year before completing their doctoral program, students meet with the Program Director and the Career Advisory committee to discuss postgraduate training, residencies, fellowships, and faculty positions and non-academic based positions. Many of the faculty are MD/PhDs and are capable of participating in career planning for third- and fourth-year medical students.

I. Program Duration and Outcomes. Over the past 6 years, our time from enrollment to graduation has averaged 7.6 years (8.1 years in the prior 5-year period). The Pittsburgh MSTP has 170 alumnae. 89% of graduates from the past 15 years are in the academic pipeline (either still in training or in academic positions). Senior MSTPs in 2018 averaged 9.6 papers with 4.5 first-authored (median: 10 total, 4.5 first-authored).

J: F30 applicant, BaDoi Phan

J.1 Student Information. BaDoi matriculated into the MSTP in June 2017 (rotation pre-MS1 year) and is in his G1 year in the program as a graduate student in the Joint Carnegie Mellon-University of Pittsburgh PhD Program in Computational Biology (CPCB). BaDoi entered the CPCB Graduate program in 2019. He chose Dr. Andreas Pfenning (in whose laboratory BaDoi had previously rotated) for his thesis advisor, where BaDoi investigates computational models to explore regulatory elements in brain cell types and tissues and their contributions to addictive behaviors and neuropsychiatric disorders and aging. Dr. Pfenning has PhD and postdoctoral work in computational biology. Dr. Pfenning was selected as is expert in computational biology and neuroscience.

J.2 Coursework Completed. BaDoi completed his MS1 and MS2 coursework and passed Step 1 of the Boards in April 2019. He has completed all three MSTP Professional Development courses and laboratory rotations, and the RBMK courses. Prior to beginning graduate work, BaDoi also completed two 4-week clinical rotations in Pediatric Inpatient Medicine and Obstetrics/Gynecology Medicine.

J.3 Awards BaDoi received a Travel Grant from the Pitt Graduate Student Assoc. for work with Dr. Pfenning.

J.4 Graduate program coursework. The coursework for the CPCB consists of a total of 4 core courses, 5 graduate-level elective courses. The core courses of the CPCB are *Machine Learning* (completed), *Computational Structural Biology* (completed), *Computational Genomics* (completed), and *Cell & Systems Modelin*g (completed). MSTP students are excused from 4 elective courses (Laboratory methods, Life Sciences Elective, Open Elective, and CPCB Ethics). Thus, BaDoi has only 1 elective requirements in graduate school: *Probabilistic Graphical Models*. He plans to take *Cellular Neuroscience* offered by the Neuroscience Institute. Students in the CPCB also participate in Journal Club, Research Seminars, and Scientific Writing courses, where they critically evaluate and are exposed to primary research and key figures in the field of computational biology. MSTP students in CPCB TA for one graduate course, which BaDoi has completed for *Genomics and Epigenetics of the Brain* with Dr. Pfenning. The CPCB has a number of program milestones, and the timeline of these is adjusted for MSTP students to allow for expeditious completion. The first of these is the Thesis Proposal that is expected to be defended no later than the end of the 3rd semester in the CPCB (around 1-2 years sooner than the regular PhD students). The Thesis Proposal Defense consists of the preparation and defense of a doctoral thesis proposal in front of a committee of at least 4 faculty (3 from training program with at least 1 each from Carnegie Mellon University and the University of Pittsburgh, and 1 external). BaDoi has formed his committee from experts in computational biology, genetics, neurobiology and will have proposed by December 2020.

J.5 Milestones till completion of training. BaDoi is on track to form his committee and propose his thesis in Fall 2020. BaDoi has strong preliminary data gathered to form his thesis. After completing this final milestone, BaDoi will be a formal PhD candidate in the CPCB. He will then have regular committee meetings at biannual intervals, and he will be expected to complete his degree within approximately 2.5 years of his proposal. BaDoi is currently on track to defend his PhD and return to medical school in May of 2023. This would put BaDoi on track to complete his remaining medical school clerkships and graduate from our program December 2024 or May 2025, respectively. Thus, for the terms of the fellowship proposed, BaDoi plans to complete an additional 30 months of research, and 18-24 months of clinical work. Should BaDoi elect the route in which he graduates medical school in December 2024, he will then undertake the 5-month MSTP Postdoctoral Fellowship, described above. The post-doctoral fellowship <u>is not included</u> in the time of covered support requested in the current application.

PHS Human Subjects and Clinical Trials Information

OMB Number: 0925-0001

Expiration Date: 02/28/2023

Use of Human Specimens and/or Data

Does any of the proposed research in the application involve human specimens and/or data *	❍ Yes ● No
Provide an explanation for any use of human specimens and/or data not considered to be human subjects research.	
Are Human Subjects Involved	❍ Yes ● No
Is the Project Exempt from Federal regulations?	❍ Yes ❍ No
Exemption Number	❑ 1 ❑ 2 ❑ 3 ❑ 4 ❑ 5 ❑ 6 ❑ 7 ❑ 8
Other Requested Information	

Vertebrate Animals

A. Description of Procedures

All procedures have been approved in the IACUC protocol TR201900003 by the Carnegie Mellon University Institutional Animal Care and Use Committee (IACUC) review board.

Fresh brain dissection: Prior to decapitation, rats will be anesthetized with isoflurane until a lack of pedal withdrawal upon noxious pinch. Animals will be quickly decapitated. The brains will be quickly removed and sectioned in artificial cerebrospinal fluid on a vibrating microtome.

Breakdown of new animal experimentation needed to complete aims.

Aim	Experiment	Species / Strain	Number of Animals
2a	Single-nucleus ATAC-seq	Rat / Brown Norway (male, female)	3, 3

B. Justification of species

The rat nervous systems have been used by the addiction/reward neuroscience community to model the essential components of human addiction and reward. The use of rat aims to test this long-held assumption that the cellular and molecular profiles of the rat nucleus accumbens (NAc) is functionally and molecularly conserved. Widely used drug self-administration paradigms in rats recapitulate a number of behaviors characterized in human, some not observed in mice. Experimentalists remain undecided on whether lack of addiction behaviors in mice translates to inferiority as a model for addiction, so behavioral studies for substance use addiction are still conducted in rats. Inclusion of rat in this study will highly benefit interpretation of results from rats vs. mouse drug addiction studies performed at the cellular and molecular level and the relevance to the human condition.

Selection of species, strain, sex, and age of animals

For all experiments, the developmental equivalents to 20-40 year old human adults in rat (8-18 month) of both sex will be used. Age and sex will be used as covariables for all analyses.

Justification for the number of animals used

Aim 2a seek to profile cellular populations within the NAc. A power analysis aiming to capture a minimum of 300 nuclei of any rare cell populations that are least 5% of all nuclei requires a target capture rate of 6,000 nuclei per tissue per animal. Using an N=3 per sex will produce a minimum of 1,800 nuclei per species over 4 biological replicates of this rare cell population. This will power discovery of cell type-specific genes/enhancers within the NAc while covarying for known variables contributing to variability: sex, age, and animal (as a random variable).

C. Minimization of pain and distress

Description of veterinary care

The Mellon Institute Central Vivarium (MICV) is a fully accredited facility. Rats are currently housed in a barrier facility, which is in the same building in which our lab is located. All rats in these studies will be bred within the MICV or ordered from Jackson Laboratories. Animals maintained in accordance with applicable portions of the Animal Welfare Act and the NRC Guide for the Care and Use of Laboratory Animals. Animal research at Mellon Institute and other Schools at CMU is overseen by the MICV. The expert veterinary staff of MICV ensure the implementation of a humane animal care and use program by providing a range of services to the university's biomedical research community. All animals are housed in specific pathogen free (SPF) environments and are monitored daily by the MICV staff for infection or disease. If animal care issues arise, these issues will be immediately treated and resolved through interaction between laboratory personnel and the MICV staff.

Methods of euthanasia

Adult rat euthanasia will be performed under deep anesthesia. These methods are consistent with the recommendations of the 2020 Report of the American Veterinary Medical Association Panel on Euthanasia and will only be performed by trained personnel. The proposed euthanasia procedures have already been approved by the CMU IACUC.

RESOURCE SHARING PLAN:
I am committed to the rapid dissemination of knowledge generated from the work in this proposal and applying the **FAIR data principles** to disseminate my work. Data generated during the training period will be shared through posters and/or presentations at the CNBC, CPCB, and MSTP work-in-progress meetings or annual retreats, as well as national and international meetings such as the Society for Neuroscience and NIDA Genetics and Epigenetics Meeting.

Our goal is to publish our data by completing experiments outlined in this application to provide valuable insight into the functional role of conserved cell types and their gene regulators. User-friendly, detailed protocols, animals, and resources will be documented on the **Pfenning lab website** and the scientific community website **protocols.io**. All manuscripts will contain complete methodological details and link to the digital protocol to enable others to fully replicate the experimental protocols. The complete genomic data will be made available at the time of publication, and manuscripts will aim to be open-source or available on **PubMedCentral**.

Raw genomic data at the multiple stages will be made available as follows: Sequenced reads will be deposited to **Gene Expression Omnibus**, **Sequenced Read Archive**, or a similar public archive. Both unprocessed summarized counts as well as batch-corrected, normalized counts will be made available for download on the lab website or **figshare.com** to facilitate wide-spread community use of these data. Count data will be made in widely used single cell **Seurat** and **Anndata** objects for fast integration into analyses pipelines. Accompanying rich metadata to describe the sample batch, sex, age, and other variables will be included to provide data transparency.

Code to generate intermediate steps and final analyses will be published on **github.com** along with **iPython** notebooks walking through from raw data to reproducible analyses and figures. Code to run the implemented **nsPGM** framework with be published on a separate **github.com** repository, with **iPython** vignettes to install, run, and analyze outputs of the nsPGM with either snRNA-seq or snATAC-seq data.