

Washington University in St. Louis

Washington University Open Scholarship

Arts & Sciences Electronic Theses and
Dissertations

Arts & Sciences

Spring 5-15-2020

Transcription Factor-Mediated Epigenetic Regulation in the Healthy Brain and Neurological Disease

Alexander J. Cammack
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds



Part of the [Genetics Commons](#), and the [Neuroscience and Neurobiology Commons](#)

Recommended Citation

Cammack, Alexander J., "Transcription Factor-Mediated Epigenetic Regulation in the Healthy Brain and Neurological Disease" (2020). *Arts & Sciences Electronic Theses and Dissertations*. 2170.
https://openscholarship.wustl.edu/art_sci_etds/2170

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Neurosciences

Dissertation Examination Committee:

Timothy M. Miller, Chairperson

Joseph D. Dougherty

Harrison W. Gabel

Robi D. Mitra

Erik S. Musiek

Conrad C. Weihl

Transcription Factor-Mediated Epigenetic Regulation
in the Healthy Brain and Neurological Disease

by

Alexander James Cammack

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2020
St. Louis, Missouri

© 2020, Alexander Cammack

Table of Contents

List of Figures	iv
List of Tables	v
Acknowledgements.....	vi
Abstract of the Dissertation	xiii
CHAPTER 1	1
Transcription factors and enhancers	2
Epilepsy and acute seizure: an epigenetics perspective	8
Epigenetic contributions to amyotrophic lateral sclerosis	11
Summary of the dissertation	18
Bibliography	22
CHAPTER 2	38
Preface.....	39
Abstract.....	40
Introduction.....	41
Methods.....	45
Results.....	60
Discussion.....	82
Acknowledgments.....	86
Author contributions	87
Declaration of interests	87
Bibliography	88
Supplemental figures	97
CHAPTER 3	108
Preface.....	109
Abstract.....	110
Introduction.....	111
Methods.....	113
Results.....	117
Discussion.....	122
Acknowledgements.....	124
Author contributions	124

Bibliography	125
CHAPTER 4	130
Preface.....	131
Abstract.....	132
Introduction.....	133
Methods.....	137
Results.....	145
Discussion.....	155
Acknowledgements.....	158
Author contributions	158
Bibliography	159
CHAPTER 5	165
Preface.....	166
Abstract.....	167
Introduction.....	169
Methods.....	171
Results.....	178
Discussion.....	187
Acknowledgements.....	192
Author contributions	193
Declaration of interests	194
Bibliography	195
CHAPTER 6	203
Summary of the dissertation	204
Advances, limitations, and future directions.....	206
Impact, innovation, and final thoughts	220
Bibliography	221
Curriculum Vitae	230

List of Figures

Chapter 2

Main figures

Figure 2.1 Co-AAV9 intracranial injection efficiently delivers calling cards to the cortex	63
Figure 2.2 Unfused hypPB-directed calling cards insertions identify active enhancers and super enhancers in the brain	67
Figure 2.3 FLEX calling cards system generates cell type-specific RE profiles.....	72
Figure 2.4 Fusion of SP1 to hypPB in FLEX calling cards system records SP1 occupancy	74
Figure 2.5 Longitudinal SP1 profiling reports integrated record of SP1 binding.....	78
Figure 2.6 AAV::hypPB Frontflip reduces Cre-independent transposition <i>in vitro</i> and <i>in vivo</i>	81

Supplemental figures

Figure 2.1S AAV9 intracranial injection induces hypPB expression in the cortex.....	97
Figure 2.2S Comparison of neonatal and adult AAV delivery of traditional DNA calling card and SRT calling card systems.....	99
Figure 2.3S AAV SRT calling cards system does not induce excess degeneration or behavioral/developmental deficits.....	100
Figure 2.4S <i>In vitro</i> and <i>in vivo</i> sensitivity of unfused hypPB calling card libraries for active REs.....	101
Figure 2.5S Supplemental to: FLEX calling cards system generates cell type-specific RE profiles.....	103
Figure 2.6S Functional validation of enhancer activity at P7 for GFAP::Cre enriched unfused hypPB peaks	104
Figure 2.7S Candidate astrocyte enhancers identified with FLEX calling cards direct cell type-specific expression in astrocytes by P21	106
Figure 2.8S Gating procedure for flow cytometry in cells with BrokenHeart-derived TdTomato fluorescence	107

Chapter 3

Figure 3.1 Seizure severity does not correlate with mouse sex, genotype, or weight.....	118
Figure 3.2 Retrospective analysis of longitudinally-recorded BRD4 occupancy identifies vulnerability and resiliency loci that are predictive of seizure outcome	121

Chapter 4

Figure 4.1 Differentially expressed miRNAs in cervical spinal cord of <i>C9orf72</i> transgenic mice	146
Figure 4.2 BRD4-bound enhancer profiling in <i>C9orf72</i> transgenic mice.....	148
Figure 4.3 H3K27ac ChIP-seq in postmortem C9ALS, sALS, and Con lumbar spinal cord	150
Figure 4.4 ALS risk-associated genetic variants fall in active spinal cord enhancers.....	154

Chapter 5

Figure 5.1 C9ALS natural history: descriptive characteristics	180
Figure 5.2 C9ALS natural history: measures of disease progression.....	182
Figure 5.3 Relationship of G ₄ C ₂ repeat size and age.....	183
Figure 5.4 poly(GP) in C9ALS CSF and size in autopsy tissue	186

List of Tables

Chapter 4

Table 4.1 List of epigenomic profiling assays used in C9ALS mice and humans 136

Table 4.2 Clinical and demographic information about C9ALS, sALS, and Con patient tissues used in this study..... 144

Chapter 5

Table 5.1 Clinical and demographic information about prospective C9ALS cohort 179

Acknowledgements

There are many people that deserve as much credit for this PhD as I. These are the people that kept me grounded during the good times and supported me during the bad. I couldn't have made it here without them.

I'll start with the ones who were my motivation for coming to graduate school to study neurodegeneration in the first place: my grandparents. On my dad's side, this includes Maggie and Eugene Cammack (Grandmaggie and Granddad), and on my mom's, Helen and Bill Orchid (Mamaw and Papaw). Grandmaggie died when I was only 8 years old, yet I still have fond memories of playing Scrabble and watching Wheel of Fortune with her. She was a bedrock for our family. Granddad lived until 2014, though his last few years were beset with Alzheimer's disease. His Alzheimer's symptoms began to appear near the time I was beginning college, so while my family witnessed gradual declines in his memory, I saw noticeable, more drastic changes when I would visit home over breaks and holidays. I always felt bad that I couldn't be there more often to support my family. So, when it came time to choose my next career steps, it felt right to enter a neuroscience program and try to fight the disease for generations to come. By this time, however, it was not only my paternal grandfather that was affected by dementia; Papaw had it too. Papaw and Mamaw were married for 66 years before he died in 2018. When I was born, I became a major focus of Papaw's and Mamaw's lives and continued to visit them often throughout my middle and high school years. Seeing Papaw reach the end stages of his life was really difficult for me and was compounded by my living far away from home. This further cemented my intent to study neurodegenerative disease. Finally, I would be remiss if I failed to mention Naomi Cammack, who married Granddad after Grandmaggie passed and really filled the role of "grandmother" to me throughout my teenage years. Naomi passed away in 2018.

In all of the places I have lived, I've been fortunate to have amazing friends that have supported me and accompanied me through the best experiences of my life. My first "best friend" that I can remember was Kenji Thompson, my neighbor on Barryknoll Lane from 1st through 5th grades in Port Arthur, Texas. He taught me what friendship was like. My high school buddies, Eric, Adrian, and Chad, we've stuck together through it all. I miss our high school days. Playing in the band, hanging out at each other's houses, playing Call of Duty and drinking Mountain Dew until 5:00am, and maybe or maybe not partaking in some illicit activities involving toilet paper...those were some good times! Even though we've split apart to follow our own careers, we stay in close contact and visit each other often, for which I'm truly grateful. I am also thankful for my friends at Wesley United Methodist Church in Nederland, Texas, where I spent many of my formative years and thoroughly enjoyed playing in praise bands.

In college in New Orleans, I picked up another awesome set of lifelong friends - too many to name individually (but I'll try anyway, because this is my dissertation damn it, and there's no page limit to the acknowledgements section). I'd like to especially thank my TUMB friends, Eleanor, Laura, Ari, Griff, Rachel, Ian, Zak, Olivia, Mendel, Mark, Barry, Ray, Pete, Kyle, Brendan, Catherine, Martin, Richard, Matt VH, all of Green Wave Brass Band/Funky Pelican Pleasure Club, and of course, the TUMB trombone section, Liz, JT, and Kate. Special shout out to my beat bud Amy. And last but certainly not least, my best friend and lifetime member of the TUMB, Chris "Cooch" Cuccio, who's been by my side since day 1 and continues to be to this day. While I miss our days of late-night Rally's fries and South Park episodes, it's been even more fun to be your friend as we've grown older. I'm grateful that we are able to visit each other at least a few times each year and look forward to continuing that tradition in the future! Plus, who would have thought Lu's best friend would be a scraggly, old, ugly-but-cute

Elmo dog. Turns out, they are perfect for each other. Oh, and I guess there's one other TUMB member worth mentioning. But I'll get to her later.

In my 6 years here in St. Louis, I've only further expanded this friend network. Of course, I have to start with my first friend here, my roommate of 3 years, and one of my best friends to this day, Ben Seitzman. It was so fortunate to find you during our interviews and revisits and set up what turned out to be a great roommate situation. I always appreciated you bringing me into your friend group with Adam and crew early on and making me feel at home in this new city. And of course, Lu will always appreciate the playtime she was able to have with lil' Purrkinje. You're destined for great things, and I'll be proud to be watching when you achieve them. Next, special shout outs to Jonathan and Steph. Jonathan, when we met for the first time at Neuro retreat in 2014, I knew we'd be great friends. I was proven right when we became first baggage buddies in SLUA leagues and eventually best friends beyond that. And Steph, you've been awesome to hang out with too since we met and found out we lived in the same apartment complex. So many great memories (and some memories that only exist because pictures were taken to document them...coughcoughMardiGrascoughcough...). Wishing you both the best in your next adventure in Boston. Krissy and Andrew, and of course, Zooey "Beans" Hays, it was so fun hanging out with you two when you were still in St. Louis, and we are really glad we were able to visit you in Raleigh-Durham last year! We can't wait to make it back out there for another visit soon (and finally meet Ramsey!). My cohort here in the Neuro program has also been a huge support group for me, especially in the first couple of years - Brian, Tom, Ben, Kael, and Katherine especially, thanks for your friendship through the years. Kael and Manda, both of you have become best friends of mine and Melissa's. I'm so glad we were able to live down the street from you in the Grove. So many Harry Potter parties, U.R.B.

tastings, O'Shay's trivia nights, Handlebar evenings, cigars around the fire, hiking trips, and just general awesome hang out sessions. I know Melissa and I both really appreciate your friendships and the times we've had together. Speaking of trivia, how about that Dw/oB team though?! Ted, Tirth, Ben, Kathleen, Andrew, Anna, Nicole, Jonathan, Steph, Kael, Manda, and the rest of you champions, I salute you. But actually, weekly trivia with you all has been one of my longest standing and most fun hobbies in St. Louis. Thanks for keeping me sane. Lastly, I want to thank the entire Ultimate community in St. Louis. I've made a ton of friends playing frisbee, both in leagues, and of course, through p'ArchD. I really feel like I was able to be a part of this tight knit community while I was here, and I'll miss playing with you all every day. Special shout outs to Jonathan Lin, Jeff "BagelShoe" Hanson, Mitch Lindsey, Ciara Murphy, and Alex Holehouse, with whom I've been fortunate to play on many teams and whose company I have truly enjoyed.

And finally, to end my friend acknowledgements, I'll thank my labmates-turned-best-friends from the Miller lab. Mariah, you were my rotation mentor and are probably the biggest scientific influence in my life thus far. You're one of the smartest people I know, and everyone knows you're going to be a famous P.I. one day. Even more frustrating is how freaking nice and loving you are to your friends and family. And to top it off, you also have the world's nicest, funniest guy as a husband. Like geez dude, leave some good qualities for the rest of us! I'm kidding of course ☺. Kelly, you really are probably the singular funniest person I've ever met and such a genuine, caring human being. Mariah is lucky to have you, and Melissa and I are fortunate to call you both our friends. Next up: Wade! Oh Wade, oh Wade. Where to start? Wade was my Neuro program "buddy" that led me around to my interviews as a little prospective graduate student. And I say little, because relative to Wade, I am quite small. I mean this both literally (Wade was a linebacker in college) and figuratively (Wade has probably the biggest

personality of anyone I know). I never would have thought he would end up as one of my best friends, but alas, life presents itself in mysterious ways! Wade, you've always been a great backboard off which to bounce ideas and have always been supportive of me and given me advice, both in life and in science, when I've needed it. Melissa and I love getting to see you and Zoe every chance we can. Finally, Kathleen and Andrew. Andrew, our time together has been a blast. Thank you for teaching me your brewing and distilling ways, oh great master of ethanol. Your smoked meats are tasty, and your friendship, even tastier. Or something like that. And then there's Kathleen. Your upward mobility in friendship rankings is astounding; from administrative to executive friend in just a few short years! ☺ You know you're my best friend in the world. There's no way I could possibly recap what you mean to me in this section, so I'm not going to try. I think you know anyway. You and Andrew have been integral in mine and Melissa's lives. I don't think we would have made it through this journey without you. We will miss you both so much when we move on.

Additionally, I'd like to thank all of those at Tulane, LSU, and Wash. U. that have mentored me scientifically. Firstly, I'd like to thank Drs. Mimi Sammarco and Ed Grabczyk at Tulane and LSU, respectively, for introducing me to science research as an undergraduate and setting me on a path to success. I'd also like to thank my committee, Drs. Erik Musiek, Rob Mitra, Joe Dougherty, Chris Weihl, and Harrison Gabel. Harrison, I appreciate you jumping in last minute and providing your feedback for my defense. Erik, I really enjoyed rotating in your lab as a first-year student and have always valued your input in committee meetings. Chris, you've been a really supportive committee chair and collaborator these past few years, and I have always appreciated your critical but fair views on my science. And Rob and Joe, both of you have essentially been co-mentors to me. Your willingness to provide your time and energy to

a student that is in another lab is just an example of your selfless characters and fantastic mentorship styles. I've benefitted greatly from your guidance throughout the years and will always be grateful to both of you for your support. And finally, and most importantly, thank you to Dr. Tim Miller, whose lab I have thoroughly enjoyed being a part, and whose scientific and professional mentorship have shaped me as scientist. Tim, you've given me the freedom to explore and the guidance to be successful. I am the researcher I am today because of your mentorship.

I want to thank my family for being my most important support group throughout my life. Their unconditional love and steady encouragement have allowed me to grow and become the person I am today. First, the Myers's, Marilyn, Gary, Jamie, and Chris. You four are like a second family to me, and I think that's been especially important while growing up as an only child. In all of my best memories, you four are all there. Next, my mom. Mom, thank you for raising me and supporting me throughout the years. You've always been there for me and made sure I've had what I needed to be successful. As I've gotten older, our relationship has only grown. Melissa and I will always be thankful for the fun times we've been able to spend every year with you and Gary. And finally, my dad. Dad, you're the best man I know, and I am so thankful that you found Christina and are now truly happy again. You've raised me into the person I am today, and I live my life trying to be the man you would want me to be. Everything I know and every value I have comes from you. I am proud to be your son.

Now, some may say it's silly to devote an entire paragraph to an animal. I agree. So I'll devote two. Lulu, my pup, you are seriously the best dog and best friend a guy could ask for. You'll never read this, but I know you know how much you mean to me.

Paragraph number 2 for Lu. You are a beautiful puppy girl who has been there for me through so much of my life. I know you won't live forever. But I intend to cherish every single day I have with you on this Earth. I love you pup.

Oh right, now about that other TUMB friend I mentioned earlier...okay maybe she's more than just a friend ☺. Melissa, you are my world. You're the one that gets me through each day. These past 6 years together have been the best 6 years of my life, and it's all thanks to you. Every memory we have made and continue to make together is a blessing. Thank you for uprooting your life and coming to St. Louis to be with me. It's not a sacrifice I will ever take for granted. I am beyond excited to continue our journey together and experience life with you in the future. I love you.

My time as a graduate student has now come full circle. I arrived as a bright-eyed and bushy-tailed 22-year-old fresh out of undergrad, made it through the 3rd and 4th year doldrums where no experiments were working and there was no light at the end of the tunnel, and have now come out with even more ambition to continue to push forward our knowledge of neurodegenerative disease. I look back at this time in my life, both the struggles and the successes, and see how this experience has shaped me as scientist and as a person. I am sure now that science research is the right path for me. But more importantly, I also now recognize the power of a support group. Family, friends, labmates, pets, and even the friendly bartenders that always knew my name - you have all played enormous roles in getting me to where I am today. And for that, I am truly thankful.

Alexander J. Cammack

Washington University

May 2020

Abstract of the Dissertation

Transcription factor-mediated epigenetic regulation
in the healthy brain and neurological disease

by

Alexander James Cammack

Doctor of Philosophy in Biology & Biomedical Sciences

Neurosciences

Washington University in St. Louis, 2020

Professor Timothy M. Miller, M.D., Ph.D., Chairperson

Proper cellular development and function is a complex process established by elaborate gene expression networks. These networks are regulated by epigenetic processes, which alter chromatin states and coordinate the binding of transcription factors (TFs) to regulatory elements (REs), such as enhancers, across the genome to facilitate gene expression. It follows then that a major experimental effort is to profile and understand the binding patterns of TFs to REs in various cellular types and contexts. Critically however, current TF profiling techniques are limited in their abilities to profile TF occupancy in targeted cellular populations and temporal windows, hindering investigations into epigenetic control in complex, multicellular systems, such as the brain. This dissertation focuses on two related areas: firstly, the design of new tools for profiling TF genome

occupancy in the mouse brain in specific cellular populations and time periods, and secondly, investigating TF-mediated mechanisms of disease pathogenesis in animal models.

In Chapter 2, we describe the development of a novel, viral-mediated method, termed adeno-associated virus (AAV) calling cards, for profiling binding sites of TFs across the genome in the live mouse brain. The AAV calling cards approach allows unique access to TF occupancy information that is inaccessible with other existing techniques, including cell type specificity (through Cre-mediated conditional expression) and historical binding (through longitudinal occupancy recording).

Then, in Chapters 3 and 4, we apply this new technique to mouse models to investigate epigenetic misregulation in disease. Previous studies have demonstrated that a large portion of genetic variation associated with cellular dysfunction or disease exists in TF-bound enhancers, demonstrating the criticality of proper TF binding in maintaining cellular homeostasis. However, whether these elements are misregulated more broadly in disease contexts is unclear. In Chapter 3, we apply AAV calling cards to a model of acute seizure and uncover aberrant epigenetic regulation which is predictive of phenotypic outcomes. Particularly important in this study is the ability of AAV calling cards to record and integrate historical TF binding information, allowing linkage of antecedent epigenetic events to eventual seizure outcomes. Here, we longitudinally recorded prodromal enhancer activity to identify loci which are predictive of seizure severity.

Next, in Chapter 4, we investigate epigenetic regulation in animal models and postmortem tissues from individuals with amyotrophic lateral sclerosis (ALS). In this study, we focus on a subset of ALS caused by a large hexanucleotide (G₄C₂) repeat expansion in the gene *chromosome 9 open reading frame 72 (C9orf72)*, which is the most common genetic cause of ALS (C9ALS). Utilizing AAV calling cards as well as other established epigenomic profiling techniques, we

observe broad epigenetic misregulation both in C9ALS mouse models and human tissues at the transcriptional and translational levels. Importantly, the C9ALS mouse models used in this study do not develop motor neuron degeneration or ALS-like phenotypes and were profiled at an early age, suggesting that these changes occur early in the disease process and are likely driven by *C9orf72*-related pathologic species, such as dipeptide repeat proteins (DPRs).

Finally, in Chapter 5 we investigate the characteristic properties of *C9orf72*-specific pathologies, including DPRs, in human C9ALS. We probed size and abundance of DNA expansions and DPRs in blood, cerebrospinal fluid, and postmortem tissues from C9ALS and sporadic ALS (sALS) individuals and identified novel correlations of C9ALS patient pathologies with clinical and demographic data. Moving forward, these data will facilitate mechanistic studies and clinical trials aimed at reducing or altering C9ALS pathologies in the central nervous system (CNS).

In summary, the body of work detailed here extends our knowledge of TFs in both the healthy and diseased central nervous system (CNS), providing new insights into the role of epigenetic regulation in disease pathogenesis. Further, the establishment of AAV calling cards as a widely applicable epigenomic tool will empower innovative new studies in a variety of tissue and model systems.

CHAPTER 1

Introduction into epigenetics in the brain in health and disease

Transcription factors and enhancers

Transcription factors in the control of gene expression

The “central dogma of molecular biology,” first posited by Francis Crick over half a century ago (Crick 1958, 1970), states that genetic information flows from the master template of the cell, DNA; to a temporary, versatile copy, RNA; to the final workhorse, protein. While this general concept continues to be a guiding principle in the fields of genetics, genomics, and molecular biology, it is now clear that this process involves complex regulatory mechanisms by which a cell determines which genes to express and not express, many of which act at the DNA to RNA step (Chen et al. 2017). Indeed, of the estimated 20,000-25,000 protein coding genes in the human genome (The ENCODE Project Consortium 2004), the average human cell expresses only a 10,000-20,000 gene subset as RNA (Alberts, Johnson, and Lewis 2002), leading to the vast cellular diversity that creates a complex organism. Collectively, we term these regulatory processes “epigenetics,” and the genome-wide study of such mechanisms “epigenomics.”

Epigenetic mechanisms are multifaceted and include, but are likely not limited to: post-translational modifications of the histone octamer, a set of proteins around which DNA is wrapped and compacted; chemical changes to the DNA itself, namely methylation of CpG dinucleotides; higher order changes to chromatin architecture, resulting in inter- and intrachromosomal DNA–DNA interactions; and regulation of DNA or messenger RNAs (mRNAs) by non-coding RNAs such as long noncoding RNAs (lncRNAs) or micro RNAs (miRNAs) (Allis and Jenuwein 2016; Chen et al. 2017). Regardless of the mechanism, the majority of these processes ultimately result in local changes to the structure of chromatin, which in turn allows selective accessibility of transcription machinery such as transcription factors (TFs) to DNA-based regulatory elements (REs) (Guertin and Lis 2010; John et al. 2011; Liu et

al. 2006; Neph et al. 2012; Spitz and Furlong 2012). It is this critical step of TFs binding to REs that ultimately determines the gene expression networks that will dictate the development, function, and identity of a cell. Similarly, it is thought that altered TF-RE binding may disrupt cellular function and may contribute to the pathogenesis of human diseases (Corradin and Scacheri 2014; Ernst et al. 2011; Hnisz et al. 2013; Maurano et al. 2012).

TFs bind to a variety of types of REs, including gene promoters as well as cis-acting elements in the non-coding space of the genome such as silencers, insulators, and enhancers (Spitz and Furlong 2012). Of these, enhancers play a leading role in transcription initiation and thus continue to be a major focus of epigenomic study. Enhancers are regions of non-coding genomic space which harbor short (6-12 base pairs (bp) (Spitz and Furlong 2012)) canonical TF binding motifs and thus act as platforms for TF occupancy. These elements are typically short in length (50-1500 bp) (Blackwood and Kadonaga 1998; Ernst et al. 2011)), though more recently a small population of large enhancers, termed super enhancers, has been identified and characterized (Whyte et al. 2013). Super enhancers, which can reach lengths orders of magnitude longer than typical enhancers, consist of multiple constitutive enhancers which act cooperatively to drive high expression of a select group of critical cell identity genes (e.g. Yamanaka TFs in embryonic stem cells (ESCs)) (Hnisz et al. 2013; Whyte et al. 2013). Accordingly, these elements harbor a large portion of TF and co-factor occupancy. One such co-factor is the bromodomain and extra-terminal domain (BET) protein, BRD4 (Hnisz et al. 2013), a “chromatin reader” which binds to enhancer-associated acetylated lysine residues on histone tails (Jung et al. 2014; Kanno et al. 2004, 2014). This BRD4 binding regulates gene transcription through recruitment of positive transcription elongation factor b (P-TEFb) to promoters of active genes in order to promote phosphorylation of RNA polymerase II (RNAPII) (LeRoy, Rickards, and Flint

2008; Mochizuki et al. 2008). Because BRD4 is constitutively expressed across the majority of cell types, its occupancy is now a common and reliable proxy for super enhancer activation (Hnisz et al. 2013).

It is estimated that there are hundreds of thousands of enhancer elements scattered throughout the genome (Pennacchio et al. 2013), however only a small subset, on the order of tens of thousands for typical enhancers and merely a few hundred for super enhancers (Creyghton et al. 2010; Heintzman et al. 2009), are actively TF-bound and regulating gene expression in any given cell. The mechanisms by which enhancer usage is governed are still unclear and likely result from a combination of factors (Spitz and Furlong 2012), including nucleosome density (Guertin and Lis 2010; John et al. 2011; Liu et al. 2006; Neph et al. 2012), local sequence context (White et al. 2013; Yanez-Cuna et al. 2012), and parallel expression and activity of co-factors and other TFs (Slattery et al. 2011). Further, despite often being located tens or even hundreds of kilobases (kB) away, TF-bound enhancers are able to enact tight control over expression of their associated genes. A variety of mechanisms have been proposed for this long-range association, including chromatin looping (Amano et al. 2009; Shlyueva, Stampfel, and Stark 2014) and, more recently, liquid-liquid phase separation (LLPS) of TFs and cofactors (Boijja et al. 2018; Hnisz et al. 2017; Sabari et al. 2018) and even chromatin itself (Gibson et al. 2019). In any case, it remains clear that the specificity with which TFs bind enhancers and regulate gene expression is intentional and critical for the proper enactment of developmental and homeostatic cellular programs. It follows then that a major focus of the epigenomics field is to attempt to profile and understand the binding patterns of TFs to REs across the genome in various cellular types and contexts.

Methods for genome-wide transcription factor and enhancer profiling

There exists a large and rapidly growing catalog of methods to profile TF binding and/or RE activity across the genome, each providing unique insights into TF-mediated epigenetic control. The most prominent and widely used method for profiling individual proteins is an antibody-based technique called Chromatin Immunoprecipitation followed by sequencing (ChIP-seq), which works by first crosslinking proteins to DNA and then isolating, sequencing, and mapping protein-bound DNA, allowing for the genome-wide visualization of TF and/or post-translationally modified histone occupancy. ChIP-seq, as well as its more recent successors such as Cleavage Under Targets and Release Using Nuclease (CUT&RUN) (Skene, Henikoff, and Henikoff 2018) and Tagmentation (CUT&Tag) (Kaya-Okur et al. 2019), have provided numerous insights into the cellular functions of TFs and have been thoroughly used to illuminate the diversity of TF binding amongst cell types (Ernst et al. 2011; Heintzman et al. 2009; Hnisz et al. 2013). Alternatively, individual TFs or co-factors may be mapped using TF-tagging techniques in which the protein of interest is fused to a DNA-modifying enzyme (Schmid, Durussel, and Laemmli 2004; van Steensel and Henikoff 2000; Wang et al. 2012; Zentner et al. 2015). These can vary in which protein is chosen as the tagging enzyme, though the most prominent among these are DamID (van Steensel and Henikoff 2000), in which the TF of interest is fused to the *E. coli* Dam methylase and TF binding is marked via local adenine methylation, and calling cards (Wang et al. 2011, 2012), which is a major focus of this dissertation and works by fusing the TF to the *hyperPiggyBac* (hypPB) transposase and recording TF occupancy through binding site-proximal transposon insertion. These methods require that the TF-tagging enzyme fusion protein be exogenously expressed, a feature which allows for temporal and spatial control of genome-wide TF profiling while negating the requirement for TF-specific antibodies

that can burden and limit the scope of TF profiling when using ChIP-seq. However, the nature of TF-tagging techniques does risk unintended modification or disruption of the local chromatin landscape and thus may result in toxicity in some contexts (Southall et al. 2013). Finally, there are a range of techniques for profiling chromatin state more generally, such as Assay for Transposase-Accessible Chromatin followed by sequencing (ATAC-seq) (Buenrostro et al. 2013) and Formaldehyde-Assisted Isolation of Regulatory Elements followed by sequencing (FAIRE-seq) (Giresi et al. 2007), which both profile accessibility and compaction, or “openness and closedness,” of genomic regions at the nucleotide level, as well as chromatin capture methods, such as high-throughput chromosome conformation capture (Hi-C) (Kempfer and Pombo 2019), which are aimed at understanding higher order chromatin architecture. The available options for profiling TFs and REs are numerous, and their unique abilities, when combined, have led to the discovery of important insights into TF-mediated regulation. Nonetheless, improvements and new approaches are being developed continually, each building upon the last and opening doors for novel studies in the field of epigenomics.

Transcription factors and enhancers in the healthy brain and neurological disease

Regulation of TF binding and enhancer activity is highly varied and specific amongst cell and tissue types. This is particularly true for super enhancers; one report which profiled super enhancers with H3K27ac ChIP-seq across 86 different human tissues and cell lines found that nearly half of all super enhancers were active only in one tissue and upwards of 70% in three or fewer (Hnisz et al. 2013). This is perhaps unsurprising, given the strikingly variant gene expression profiles observed between cells of different origin (The ENCODE Project Consortium 2012) and the tendency for super enhancers to regulate cell type-specific cell

identity genes (Hnisz et al. 2013; Whyte et al. 2013), but nonetheless highlights the criticality of enhancer-mediated regulation in determining cell fate and maintaining proper cellular function. This cell type-specific epigenetic regulation is also quite apparent in the brain, where a number of studies have begun to elucidate profound cell type-specificity of enhancers within neuronal and glial populations (Girdhar et al. 2018; Mo et al. 2015; Nott et al. 2019). Even subpopulations of morphologically similar, yet transcriptionally separate neurons display many distinct and non-overlapping patterns of TF binding and chromatin accessibility which have been shown to regulate genes associated with known cell type-specific processes such as activity-dependent transcription (Mo et al. 2015). Some of this variant enhancer activity may be driven by the selective expression of cell type-specific TFs (Mo et al. 2015).

Perhaps the best evidence demonstrating the importance of proper TF binding and enhancer activation for maintaining cellular homeostasis is the finding that active enhancers harbor the vast majority of disease-associated genetic risk (Corradin and Scacheri 2014; Ernst et al. 2011; Hnisz et al. 2013; Maurano et al. 2012). Critically, these risk-associated single nucleotide polymorphisms (SNPs) most commonly fall within the cell type-specific enhancers of the disease-associated cell or tissue type (Ernst et al. 2011; Hnisz et al. 2013). In the brain, it has been shown that active enhancers and super enhancers harbor risk variants for Alzheimer's disease (Hnisz et al. 2013; Marzi et al. 2018; Nativio et al. 2018; Nott et al. 2019), schizophrenia (Girdhar et al. 2017), and a host of other neurological disorders (Hannon et al. 2019). This result, which has been well-replicated for a number of different diseases both in and out of the central nervous system (CNS), strongly suggests that misregulation of enhancer-mediated epigenetic control causes cellular dysfunction that could be a driver of disease pathogenesis.

Epilepsy and acute seizure: an epigenetics perspective

Drug induced acute seizure as a model of epilepsy

Epilepsy is one of the most commonly observed neurological disorders, affecting millions of individuals in the United States alone and costing billions of dollars in direct medical costs (England et al. 2012). While a number of anti-epileptic drugs (AED) are commonly and successfully used to control seizure susceptibility in some epileptic patients (Rogawski 2006), at least one third of diagnosed individuals do not respond to treatment and continue to incur uncontrolled epileptic seizures (Devinsky et al. 2018), emphasizing the need for the development of new therapeutic strategies. And with a lifetime risk of 1 in 26 and an estimated 150,000 new cases diagnosed annually (England et al. 2012), this need is important and urgent for the treatment of this patient population.

The term epilepsy actually refers to a group of similar yet distinct clinical phenotypes which can be diversified by their clinical features such as age at onset, co-morbidities, and seizure type (Myers and Mefford 2015). However, the unifying feature of these variant disorders is the recurrence of unprovoked epileptic seizures. While the underlying molecular mechanisms for seizure are unclear and likely diverse, we now understand that this phenotype is due to an imbalance in inhibitory and excitatory conductances in the brain (Staley 2015), which can be modeled in animals by manipulating inhibitory or excitatory circuits to induce acute seizures (Löscher 2011). One commonly used method of inducing acute seizures in mouse models is through intraperitoneal (i.p.) injection of the GABA antagonist pentylenetetrazole (PTZ) (Löscher 2011; Rogawski 2006). Upon injection of PTZ, animals undergo convulsive, clonic seizures within minutes, although there is a high degree of variability in the severity of the seizure induced, ranging from small spasms to full clonic activity and even death (DeVos et al.

2013). While models of more chronic seizure through genetic or kindling means are arguably closer in phenotype to human epilepsy, a number of AEDs first found to be effective in acute seizure models have been successfully translated to the clinic (Löscher 2011; Rogawski 2006), supporting that these models are valid and important tools for understanding molecular mechanisms underlying epileptic seizure.

Genetics and epigenetics of seizure

Genetic factors are thought to have a major role in the pathogenesis of epilepsy, or epileptogenesis. Estimates of epilepsy cases with a genetic basis range from 70-80% of the total (Hildebrand et al. 2013; Myers and Mefford 2015), a figure derived from twin studies in which monozygotic twins have been found to have a much higher concordance rate with epilepsy than do dizygotic pairs (Kjeldsen et al. 2003). A number of genes have been implicated in the pathogenesis of epilepsy and/or epileptic seizure via human genetic studies, many of which encode proteins affecting membrane potential, such as ion channels (Myers and Mefford 2015). For this reason, many epileptic disorders are termed “channelopathies.” However, these epilepsy-causing mutations are rare and do not explain the underlying mechanism for the majority of epileptic cases.

Recent evidence suggests that epigenetic mechanisms may also play roles in susceptibility to seizure. Previous studies in both rodent and human epilepsy have observed changes in DNA methylation, histone acetylation, and miRNA expression upon seizure (Devinsky et al. 2018; Kobow and Blümcke 2018). Further, a number of TFs and co-factors have been shown to be sufficient to alter seizure severity when manipulated. Hippocampal expression and activity of a TF called neuron-restrictive silencer factor (NRSF) has been found to be

induced upon seizure, subsequently repressing a variety of target genes including ion channels and receptors (McClelland et al. 2014), whereas blocking NRSF binding is sufficient to rescue expression changes (McClelland et al. 2014) and attenuate the epileptic phenotype (McClelland et al. 2011). Moreover, the BET family proteins have also been implicated in seizure susceptibility, suggesting that enhancers may too be involved in epileptogenesis. Genetic variants in *Brd2* increase risk for juvenile myoclonic epilepsy (Pal et al. 2003), while *Brd2* null and heterozygous mice display lowered clonic seizure thresholds and disruptions in GABAergic signaling pathways (Qureshi and Mehler 2014). Meanwhile, the super enhancer-associated protein BRD4 has been shown to play a major role in the neuronal transcriptional response to firing. Inhibition or genetic knockdown of BRD4 in primary mouse cortical neurons is sufficient to block activity-dependent induction of immediate-early genes such as *Arc* and *c-Fos*, as well as a host of other genes associated with synapses and membrane excitability. Further, BRD4 inhibition *in vivo* in prior to induction of acute seizure via PTZ injection results in a drastically reduced peak seizure severity, suggesting that BRD4 is a mediator of genes critical to the propagation of neuronal activity and necessary for the occurrence of seizure (Korb et al. 2015). However, the target genes downstream of BRD4 regulation that are ultimately responsible for this susceptibility are still yet to be deciphered.

In Chapter 3 of this dissertation, we utilize newly developed epigenomic tools to elucidate the role of BRD4 in PTZ-induced seizure, in order to identify downstream vulnerability and susceptibility loci.

Epigenetic contributions to amyotrophic lateral sclerosis

Genetics and prevalence of amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis (ALS), also known as motor neuron disease or Lou Gehrig's disease, is a devastating neurodegenerative disorder characterized by the progressive degeneration of upper and lower motor neurons (MNs) in the brain and spinal cord (Taylor, Brown, and Cleveland 2016). This MN loss manifests clinically as a rapid deterioration of muscle mass, inability to move, and ultimately death typically within 3-5 years after symptom onset, often due to respiratory failure (Rowland and Shneider 2001). ALS is a relatively rare disorder, with a prevalence of just 1-2 per 100,000, however, remarkably, the lifetime risk of developing ALS is estimated to be as high as 1:400 (Johnston et al. 2006); notably, the rapidly progressive nature of ALS prevents the accumulation of affected individuals in any given population. Sadly, since the initial descriptions of the disease by Charcot in the mid 19th century (Goetz 2000), relatively little progress has been made in terms of ALS therapy, with only two drugs FDA-approved for the treatment of ALS, each with very modest effects on disease progression (3-6 month survival extension) (Abe et al. 2014; Bensimon et al. 1994). Thus, the need to develop new targeted therapies for ALS is critical, which will be enabled through first understanding the biological mechanisms underlying MN death.

The majority (~90%) of ALS cases are singleton (sALS), however about 10% of individuals with ALS do have a family history of disease and are thus collectively termed “familial” ALS (fALS). Human genetic studies over the past 30 years have identified a number of gene mutations causative for ALS, beginning in 1993 with mutations in the *superoxide dismutase 1 (SOD1)* gene (Rosen et al. 1993). In total, these known mutations now account for nearly two thirds of fALS and a small portion (~11%) of sALS cases as well (Renton, Chiò, and

Traynor 2014). The majority of ALS-associated proteins fall within two categories: protein quality control (e.g. VCP, TBK1, OPTN, UBQLN2) or RNA processing (e.g. TDP-43, FUS, hnRNPA1/2B1) (Ghasemi and Brown 2018; Renton et al. 2014). However, more recently much of the field has shifted its focus to a more unique mutation, a hexanucleotide (GGGGCC, or G₄C₂) repeat expansion in the first intronic region of the *chromosome 9 open reading frame 72*, or *C9orf72*, gene, which causes 38% and 6% of fALS and sALS, respectively, and is now the most common genetic cause of ALS (abbreviated C9ALS) (DeJesus-Hernandez et al. 2011; Majounie et al. 2012; Renton et al. 2011). Additionally, the same mutation causes 25% and 6% of familial and sporadic frontotemporal dementia (C9FTD) (Majounie et al. 2012). This expansion is typically small in non-ALS individuals (<30 repeats), however the disease associated expansion ranges from hundreds to thousands of consecutive repeats (van Blitterswijk et al. 2013; Dols-Icardo et al. 2014; Gijselinck et al. 2016; Nordin et al. 2015) which are somatically unstable and can vary greatly in size even within individual (Nordin et al. 2015).

Molecular mechanisms of C9ALS

Because the *C9orf72* mutation is a GC-rich expansion rather than a typical point mutation, it was not initially identified by human genetic studies, and it was not until 2011 that these mutations were discovered and subsequently published in landmark studies by two independent groups (DeJesus-Hernandez et al. 2011; Renton et al. 2011). Since this time, a huge amount of progress has been made in understanding the biological consequences of having this large expansion (Gendron and Petrucelli 2017; Gitler and Tsuiji 2016). It is primarily hypothesized that *C9orf72* expansions induce toxicity through one or a combination of three

distinct loss and gain of function mechanisms, each initially based on studies of related repeat expansion disorders.

The first proposed mechanism is a loss of normal C9ORF72 function, due to repeat-induced transcription reduction of the *C9orf72* gene. Indeed, presence of the expansion does silence the repeat-containing allele, resulting in a ~50% reduction in total *C9orf72* mRNA (DeJesus-Hernandez et al. 2011; Renton et al. 2011). However, because loss of the C9ORF72 gene product does not result in neurodegeneration in mice (Burberry et al. 2016; O'Rourke et al. 2016) and no known coding mutations within *C9orf72* have been found to associate with the human disease (Harms et al. 2013), a solely loss of function mechanism is considered unlikely.

The remaining two proposed mechanisms are both based on gain of function toxicity. The *C9orf72* expansion is bidirectionally transcribed, and although the expansion is intronic, these transcripts are stabilized and aggregated into formations called RNA foci inside the nuclei of affected cellular populations (DeJesus-Hernandez et al. 2011). The presence of these foci causes sequestration of critical RNA binding proteins (Donnelly et al. 2013; Lee et al. 2013), which is the first of the two proposed gain of function toxicities. Moreover, these repetitive transcripts make their way to the cytoplasm where they are translated into protein through repeat-associated non-ATG (RAN) translation, a process wherein translation may initiate in the absence of a start codon. This produces 5 separate species of dipeptide repeat proteins (DPRs): poly(GA) and poly(GR) from the sense strand, poly(PA) and poly(PR) from the antisense, and poly(GP) from both (Ash et al. 2013; Gendron et al. 2013; Mori et al. 2013; Zu et al. 2013). These DPRs are specific to *C9orf72* expansion carriers, and all five species have been shown to exist in patient tissues, forming intracellular and intranuclear aggregates throughout the brain and spinal cord (Ash et al. 2013; Gendron et al. 2013; Mackenzie et al. 2013, 2015; Mori et al. 2013; Zu et al.

2013). Further, multiple separate studies have demonstrated that poly(GR) and poly(PR) are sufficient for degeneration in isolated neurons (Wen et al. 2014), *Drosophila* (Mizielinska et al. 2014), and even mice (Hao et al. 2019; Zhang et al. 2019), while a number of poly(GA)-mediated toxicities have also been demonstrated *in vitro* and *in vivo* (May et al. 2014; Nguyen et al. 2019; Nihei et al. 2019; Schludi et al. 2017; Zhang et al. 2014), thus leading to the hypothesis that the presence of these aberrant proteins could be a gain of function driver of disease.

Some studies have attempted to determine the relative contributions of DPRs versus RNA foci to disease pathogenesis (Arzberger et al. 2018; Mizielinska et al. 2014; Moens et al. 2018; Swinnen et al. 2018), however such studies are complicated by potential effects of overexpression and lack of animals models accurately depicting C9ALS phenotypes. Intriguingly however, while the relative toxicities driven by RNA foci and DPRs are still unclear, it has been shown that downstream affected pathways driven by these two pathologies often converge (e.g. disruptions in nucleocytoplasmic transport (Boeynaems et al. 2016; Freibaum et al. 2015; Jovičić et al. 2015; Zhang et al. 2015)). Thus, while what ultimately drives pathogenesis in C9ALS is still unclear, it is becoming increasingly accepted that these two pathologies are likely involved.

Evidence for epigenetic contributions to ALS

Emerging results suggest that epigenetic mechanisms may contribute to ALS pathogenesis. In particular, recent studies have pointed to both transcriptional (chromatin dynamics) and translational (miRNAs) components of epigenetic control as potentially compromised in the disease state.

Firstly, a coalescence of evidence supports that chromatin dynamics and enhancer activity may be altered in ALS. Using a targeted toxicity screen in *Drosophila*, one group

identified the chromatin modifier CHD1 as being impaired by TDP-43, resulting in disrupted nucleosomal dynamics, increased stress granule formation, and reduction of important stress-induced heat shock genes (Berson et al. 2017). Further, a recent study demonstrated that DPRs co-localize to regions of heterochromatin in both mouse *C9orf72* models and in C9ALS/FTD postmortem cortical tissue and induce alterations in chromatin methylation, suggesting that *C9orf72*-derived pathologies can directly alter the chromatin state (Zhang et al. 2019). In mutant *FUS* transgenic mouse models, global histone hypoacetylation has been observed, while reversal of these changes with pharmacological inhibitors of histone deacetylases (HDACs) has been shown to slow disease progression and reverse aberrant metabolic changes (Rossaert et al. 2019). Finally, one study observed an increase in ALS-associated genetic variation in active enhancers profiled in postmortem entorhinal cortex samples, supporting that epigenetic misregulation may increase ALS risk (Hannon et al. 2019).

While the mechanisms through which chromatin dynamics may be disrupted remain unclear, indirect lines of evidence suggest that altered LLPS of chromatin-associated proteins could be a factor. LLPS is a phenomenon which has been heavily implicated as a possible driver of ALS-related pathologies. A number of ALS-associated proteins (e.g. TDP-43, FUS, hnRNPA1) have been shown to undergo LLPS *in vitro* and in live cells, forming liquid-like condensates that naturally transition between liquid, gel-like, and solid states (Boeynaems et al. 2017; Molliex et al. 2015; Patel et al. 2015). However, disease-linked mutations in these proteins can drive these condensates more frequently into the solid state and ultimately into amyloid-like aggregates reminiscent of those observed in patient tissues (Molliex et al. 2015; Patel et al. 2015). Moreover, poly(GR) and poly(PR) have also been shown to undergo LLPS (Boeynaems et al. 2017), during which these aberrantly expressed DPRs associate with and re-localize to

other phase-separated subcellular bodies, such as nucleoli, and in turn alter their natural LLPS dynamics (Lee et al. 2016). Interestingly, one recent study observed that poly(PR) disrupts the LLPS dynamics of heterochromatin protein 1 (HP1), subsequently co-localizing to regions of closed chromatin and inducing alterations in histone H3 methylation status, demonstrating that DPRs can directly affect chromatin organization through alteration of phase separation (Zhang et al. 2019). Extending this further, it is also now understood that TFs, nucleosomes, and co-factors such as BRD4 also utilize LLPS to maintain or change chromatin architecture (Boija et al. 2018; Gibson et al. 2019; Hnisz et al. 2017; Sabari et al. 2018); thus while no evidence has yet been shown to suggest that DPRs directly affect TFs, it is possible that similar LLPS disruptions may exist and affect enhancer activity in the context of C9ALS.

Epigenetic misregulation at the translational level through miRNA-mediated repression may also contribute to ALS pathogenesis (Hoye et al. 2017, 2018; Koval et al. 2013; Reichenstein et al. 2019; Varcianna et al. 2019). A number of studies, including from our group, have shown that miRNA levels are altered in ALS mouse models (Hoye et al. 2017; Koval et al. 2013), patient-derived induced pluripotent stem cell (iPSC) models (Varcianna et al. 2019), and patient tissues (Hoye et al. 2017; Reichenstein et al. 2019). These changes can result in neuronal and glial dysfunction and have been linked to mechanisms involved in ALS pathogenesis (Amin et al. 2015; Hoye et al. 2017, 2018; Koval et al. 2013; Reichenstein et al. 2019; Varcianna et al. 2019). However, whether miRNA-mediated regulation is similarly altered in C9ALS is unclear.

This dissertation explores and tests the concept that epigenetic disruption is an upstream event in C9ALS pathogenesis. The work described in Chapter 4 details application of epigenomic techniques to both mouse models and postmortem human CNS samples of C9ALS

to determine whether *C9orf72*-related pathologies induce alterations in miRNA and enhancer activity.

Summary of the dissertation

This dissertation focuses on the mapping of TF binding across the genome in the brain to better understand TF-mediated epigenetic regulation in healthy and diseased conditions. The work within first describes the development of a novel TF and enhancer profiling technique, termed AAV calling cards, followed by its application to two models of neurological disease. This method, the development of which is described in Chapter 2, allows access to unique TF and co-factor occupancy information inaccessible with previously existing techniques such as ChIP-seq or other antibody-based methodologies. These include cell type-specificity of TF binding between neuronal and glial cell types, as well as integrated TF occupancy recording over time, both of which are thoroughly tested and validated in the second half of Chapter 2. The work describing AAV calling cards is part of a manuscript which, as of this writing, is currently available as a pre-print on *BioRxiv* (Cammack, Moudgil, et al. 2019) and has been recently accepted for publication at *PNAS*. Following this is the application of this technique to two independent neurological disease models, both of which function as test cases of AAV calling cards utilization for means of novel information gathering as well as elucidative studies into the roles of TFs and enhancers in disease contexts.

Chapter 3 describes use of AAV calling cards to investigate the role of BRD4 in susceptibility to or protection from acute seizure. In this study, BRD4 binding is recorded genome wide in the mouse cortex across a 4-week period of time prior to induction of acute seizure via PTZ injection. By subsequently comparing mice that incurred mild seizures to those that progressed to a more severe phenotype, BRD4-bound vulnerability and susceptibility loci are identified. Beyond elucidating a potential role for BRD4-mediated enhancer regulation in acute seizure, this study also demonstrates the utility of longitudinal recording offered by AAV

calling cards. Indeed, because BRD4 binding was recorded before rather than after seizure induction, it can be interpreted that the observed differences between mild and severe seizure-receiving animals are present at seizure onset, rather than a resultant effect of the seizure itself.

In Chapter 4, we test the hypothesis that misregulated epigenetic control, including miRNA and BRD4-bound enhancer activity, is a defining feature of C9ALS that may contribute to pathogenesis. This work first details profiling of miRNAs and enhancers in spinal cord samples from C9ALS mouse models carrying multiple copies of the human expansion mutation (Liu et al. 2016; O'Rourke et al. 2015). These mice exhibit pathologic features reminiscent of the human disease, however, in our colony, do not display weakness or other disease-related phenotypes; thus, these mice represent “pre-onset” C9ALS and can be used to assess pathologically-driven biology without the confounding variables of neuroinflammation or neurodegeneration associated with end-stage ALS. By profiling global miRNA levels in the spinal cords of these animals with microarray, we find a number of miRNAs to be transcriptionally misregulated at an early timepoint. This is followed by application of BRD4-directed AAV calling cards to the cortex of one of these models. Here, a number of differential BRD4 binding sites are observed between C9ALS animals and their non-transgenic controls, suggesting that *C9orf72*-related pathologies disrupt BRD4 binding. To extend this to the human condition, enhancers are subsequently profiled with H3K27ac ChIP-seq in postmortem lumbar spinal cord samples from C9ALS, sALS, and non-ALS control individuals. Again, differential enhancer activity was observed between ALS and non-ALS samples which correlated between C9ALS and sALS. Lastly, we demonstrate that these human spinal cord enhancers harbor known ALS risk-associated SNPs, suggesting that misregulation of these regions may increase risk of disease.

Finally, in Chapter 5, we investigate further the properties of DNA expansions and DPRs in human C9ALS. We first probed size of the G₄C₂ DNA repeat in blood samples from >100 C9ALS and non-ALS mutation carriers via Southern blotting and found that repeat size correlates positively with age and age at disease onset. We extended this study to carriers of large but non-expanded repeats (13-25 repeats), where a similar relationship with age was observed. Together these results support that *C9orf72* repeat expansions may be dynamic and able to change size over time. We then asked whether size of the DNA expansion was related to that of resultant DPRs. To do so, we probed DPR size in human autopsy CNS tissue via a novel combination of size-exclusion chromatography and poly(GP) immunoassay and observed that DPRs are generally quite large and often reach lengths similar to that of their DNA counterparts. Interestingly, we also found that poly(GP) abundance in the cerebrospinal fluid (CSF) was negatively correlated with DNA expansion size, suggesting that presence of larger expansions reduces DPR load, possibly through reduced G₄C₂ RNA. This work is published in *Neurology* in conjunction with a large, prospective natural history study of C9ALS in which clinical features and biomarkers are thoroughly examined in this patient population (Cammack, Atassi, et al. 2019).

The body of work detailed here attempts to extend our knowledge of TFs and enhancers in the brain, both in healthy and diseased conditions. Through the development of a new epigenomic tool and its application to disease models, we have provided new insights into the potential role of enhancer-mediated regulation in disease pathogenesis. Further, the unique utilities offered by AAV calling cards will be equally relevant for the study of other neurological conditions, as well as during neurodevelopment. While there remains much to do to fully

understand the biology of TFs and enhancers in the brain, the contributions of this dissertation will pave the way for new epigenetic insights and discoveries in the future.

Bibliography

- Abe, Koji et al. 2014. “Confirmatory Double-Blind, Parallel-Group, Placebo-Controlled Study of Efficacy and Safety of Edaravone (MCI-186) in Amyotrophic Lateral Sclerosis Patients.” *Amyotrophic lateral sclerosis & frontotemporal degeneration* 15(7–8): 610–17.
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4266079&tool=pmcentrez&rendertype=abstract>.
- Alberts, B, A Johnson, and J Lewis. 2002. “An Overview of Gene Control.” In *Molecular Biology of the Cell. 4th Edition.*, , Available from: <https://www.ncbi.nlm.nih.gov/books>.
- Allis, C. David, and Thomas Jenuwein. 2016. “The Molecular Hallmarks of Epigenetic Control.” *Nature Reviews Genetics* 17(8): 487–500. <http://dx.doi.org/10.1038/nrg.2016.59>.
- Amano, Takanori et al. 2009. “Chromosomal Dynamics at the Shh Locus: Limb Bud-Specific Differential Regulation of Competence and Active Transcription.” *Developmental Cell* 16(1): 47–57. <http://dx.doi.org/10.1016/j.devcel.2008.11.011>.
- Amin, Neal D et al. 2015. “Loss of Motoneuron-Specific MicroRNA-218 Causes Systemic Neuromuscular Failure.” *Science* 350(6267): 1525–29.
- Arzberger, Thomas et al. 2018. “RNA versus Protein Toxicity in C9orf72 ALS/FTLD.” *Acta Neuropathologica* 135(3): 475–79. <http://link.springer.com/10.1007/s00401-018-1823-1>.
- Ash, Peter E a et al. 2013. “Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to C9FTD/ALS.” *Neuron* 77(4): 639–46.
<http://dx.doi.org/10.1016/j.neuron.2013.02.004>.
- Bensimon, G., L. Lacomblez, V. Meininger, and The ALS/Riluzole Study Group. 1994. “A Controlled Trial of Riluzole in Amyotrophic Lateral Sclerosis.” *New England Journal of Medicine* 330(9): 585–91.

- Berson, Amit et al. 2017. “TDP-43 Promotes Neurodegeneration by Impairing Chromatin Remodeling.” *Current Biology*: 3579–90.
- Blackwood, Elizabeth M., and James T. Kadonaga. 1998. “Going the Distance: A Current View of Enhancer Action.” *Science* 281(5373): 60–63.
- van Blitterswijk, Marka et al. 2013. “Association between Repeat Sizes and Clinical and Pathological Characteristics in Carriers of C9ORF72 Repeat Expansions (Xpansize-72): A Cross-Sectional Cohort Study.” *The Lancet Neurology* 12(10): 978–88.
[http://dx.doi.org/10.1016/S1474-4422\(13\)70210-2](http://dx.doi.org/10.1016/S1474-4422(13)70210-2).
- Boeynaems, Steven et al. 2016. “Drosophila Screen Connects Nuclear Transport Genes to DPR Pathology in C9ALS/FTD.” *Scientific Reports* 6(August 2015): 20877.
<http://www.nature.com/articles/srep20877>.
- . 2017. “Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics.” *Molecular Cell* 65(6): 1044-1055.e5.
<http://linkinghub.elsevier.com/retrieve/pii/S1097276517301284>.
- Boija, Ann et al. 2018. “Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains.” *Cell*: 1–14.
<https://linkinghub.elsevier.com/retrieve/pii/S0092867418313989>.
- Buenrostro, Jason D. et al. 2013. “Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position.” *Nature Methods* 10(12): 1213–18.
- Burberry, Aaron et al. 2016. “Loss-of-Function Mutations in the C9ORF72 Mouse Ortholog Cause Fatal Autoimmune Disease.” *Science Translational Medicine* 8(347).
- Cammack, Alexander J., Arnav Moudgil, et al. 2019. “A Viral Toolkit for Recording

- Transcription Factor-DNA Interactions in Live Mouse Tissues.” *BioRxiv [Preprint]*
(<https://doi.org/10.1101/538504>).
- Cammack, Alexander J., Nazem Atassi, et al. 2019. “Prospective Natural History Study of C9orf72 ALS Clinical Characteristics and Biomarkers.” *Neurology* 93(17): e1605–17.
- Chen, Zhen et al. 2017. “Epigenetic Regulation: A New Frontier for Biomedical Engineers.” *Annual Review of Biomedical Engineering* 19: 195–219.
- Consortium, International Human Genome Sequencing. 2004. “Finishing the Euchromatic Sequence of the Human Genome.” *Nature* 431(21 October): 931–45.
<http://www.nature.com/articles/nature03001>.
- Consortium, The ENCODE Project. 2012. “An Integrated Encyclopedia of DNA Elements in the Human Genome.” *Nature* 489(7414): 57–74.
- Corradin, Olivia, and Peter C. Scacheri. 2014. “Enhancer Variants: Evaluating Functions in Common Disease.” *Genome Medicine* 6(10): 1–14.
- Creyghton, M. P. et al. 2010. “Histone H3K27ac Separates Active from Poised Enhancers and Predicts Developmental State.” *Proceedings of the National Academy of Sciences* 107(50): 21931–36. <http://www.pnas.org/cgi/doi/10.1073/pnas.1016071107>.
- Crick, Francis H. 1958. “On Protein Synthesis.” *Symposia of the Society for Experimental Biology*: 138–63.
http://s3.amazonaws.com/academia.edu.documents/31353481/Symp_Soc_Exp_Biol_1958_Crick_on_protein_synthesis.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1493406505&Signature=Pj68wjQq8HUhOBMD3niX5IHiQpM%253D&response-content-disposition=inline%253B%20filena.
- . 1970. “Central Dogma of Molecular Biology.” *Nature* 228(August 8): 561–63.

- DeJesus-Hernandez, Mariely et al. 2011. "Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS." *Neuron* 72(2): 245–56. <http://dx.doi.org/10.1016/j.neuron.2011.09.011>.
- Devinsky, Orrin et al. 2018. "Epilepsy." *Nature Reviews Disease Primers* 3(18024).
- DeVos, Sarah L. et al. 2013. "Antisense Reduction of Tau in Adult Mice Protects against Seizures." *Journal of Neuroscience* 33(31): 12887–97.
- Dols-Icardo, Oriol et al. 2014. "Characterization of the Repeat Expansion Size in C9orf72 in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia." *Human Molecular Genetics* 23(3): 749–54.
- Donnelly, Christopher J. et al. 2013. "RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention." *Neuron* 80(2): 415–28. <http://dx.doi.org/10.1016/j.neuron.2013.10.015>.
- England, Mary Jane, Catharyn T. Liverman, Andrea M. Schultz, and Larisa M. Strawbridge. 2012. "Epilepsy Across the Spectrum: Promoting Health and Understanding." *Washington DC: National Academies Press*. <https://www.ncbi.nlm.nih.gov/books/NBK100596/>.
- Ernst, Jason et al. 2011. "Mapping and Analysis of Chromatin State Dynamics in Nine Human Cell Types." *Nature* 473(7345): 43–49. <http://dx.doi.org/10.1038/nature09906>.
- Freibaum, Brian D. et al. 2015. "GGGGCC Repeat Expansion in C9orf72 Compromises Nucleocytoplasmic Transport." *Nature*. <http://www.nature.com/doi/10.1038/nature14974>.
- Gendron, Tania F. et al. 2013. "Antisense Transcripts of the Expanded C9ORF72 Hexanucleotide Repeat Form Nuclear RNA Foci and Undergo Repeat-Associated Non-ATG Translation in C9FTD/ALS." *Acta Neuropathologica* 126(6): 829–44.

- Gendron, Tania F, and Leonard Petrucelli. 2017. "Disease Mechanisms of C9ORF72 Repeat Expansions." *Cold Spring Harbor Laboratory Press*: 1–22.
- Ghasemi, Mehdi, and Robert H. Brown. 2018. "Genetics of Amyotrophic Lateral Sclerosis." *Cold Spring Harbor Laboratory Press* 8(a024125).
- Gibson, Bryan A et al. 2019. "Organization of Chromatin by Intrinsic and Regulated Phase Separation." *Cell* 179(2): 470-484.e21. <http://www.ncbi.nlm.nih.gov/pubmed/31543265>.
- Gijselinck, I. et al. 2016. "The C9orf72 Repeat Size Correlates with Onset Age of Disease, DNA Methylation and Transcriptional Downregulation of the Promoter." *Molecular Psychiatry* 21(8): 1112–24.
- Girdhar, Kiran et al. 2017. "Cell-Specific Histone Modification Maps Link Schizophrenia Risk to the Neuronal Epigenome." *Nature Neuroscience*. <http://dx.doi.org/10.1038/s41593-018-0187-0>.
- . 2018. "Cell-Specific Histone Modification Maps in the Human Frontal Lobe Link Schizophrenia Risk to the Neuronal Epigenome." *Nature Neuroscience* 21(8): 1126–36. <http://dx.doi.org/10.1038/s41593-018-0187-0>.
- Giresi, Paul G. et al. 2007. "FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) Isolates Active Regulatory Elements from Human Chromatin." *Genome Research* 17(6): 877–85.
- Gitler, Aaron D., and Hitomi Tsuiji. 2016. "There Has Been an Awakening: Emerging Mechanisms of C9orf72 Mutations in FTD/ALS." *Brain Research* 1647: 19–29. <http://linkinghub.elsevier.com/retrieve/pii/S0006899316301974>.
- Goetz, Christopher G. 2000. "Amyotrophic Lateral Sclerosis: Early Contributions of Jean-Martin Charcot." *Muscle & Nerve* (March): 336–43.

- Guertin, Michael J., and John T. Lis. 2010. "Chromatin Landscape Dictates HSF Binding to Target DNA Elements." *PLoS Genetics* 6(9).
- Hannon, Eilis, Sarah J Marzi, Leonard S Schalkwyk, and Jonathan Mill. 2019. "Genetic Risk Variants for Brain Disorders Are Enriched in Cortical H3K27ac Domains." *Molecular Brain* 12(7): 1–6.
- Hao, Zongbing et al. 2019. "Motor Dysfunction and Neurodegeneration in a C9orf72 Mouse Line Expressing Poly-PR." *Nature Communications* 10(1): 2906.
<http://www.nature.com/articles/s41467-019-10956-w>.
- Harms, Matthew B. et al. 2013. "Lack of C9ORF72 Coding Mutations Supports a Gain of Function for Repeat Expansions in Amyotrophic Lateral Sclerosis." *Neurobiology of Aging* 34(9): 2234.e13–2234.e19. <http://dx.doi.org/10.1016/j.neurobiolaging.2013.03.006>.
- Heintzman, Nathaniel D. et al. 2009. "Histone Modifications at Human Enhancers Reflect Global Cell-Type-Specific Gene Expression." *Nature* 459(7243): 108–12.
<http://dx.doi.org/10.1038/nature07829>.
- Hildebrand, Michael S. et al. 2013. "Recent Advances in the Molecular Genetics of Epilepsy." *Journal of Medical Genetics* 50: 271–79.
- Hnisz, Denes et al. 2013. "Super-Enhancers in the Control of Cell Identity and Disease." *Cell* 155(4): 934–47. <http://linkinghub.elsevier.com/retrieve/pii/S0092867413012270>.
- . 2017. "A Phase Separation Model for Transcriptional Control." *Cell* 169: 13–23.
<http://dx.doi.org/10.1016/j.cell.2017.02.007>.
- Hoye, Mariah L. et al. 2017. "MicroRNA Profiling Reveals Marker of Motor Neuron Disease in ALS Models." *The Journal of Neuroscience* 37(22): 3582–16.
<http://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.3582-16.2017>.

- . 2018. “Motor Neuron-Derived MicroRNAs Cause Astrocyte Dysfunction in Amyotrophic Lateral Sclerosis.” *Brain* 141(9): 2561–75.
- John, Sam et al. 2011. “Chromatin Accessibility Pre-Determines Glucocorticoid Receptor Binding Patterns.” *Nature Genetics* 43(3): 264–68. <http://dx.doi.org/10.1038/ng.759>.
- Johnston, Clare A. et al. 2006. “Amyotrophic Lateral Sclerosis in an Urban Setting: A Population Based Study of Inner City London.” *Journal of Neurology* 253: 1642–43.
- Jovičić, Ana et al. 2015. “Modifiers of C9orf72 Dipeptide Repeat Toxicity Connect Nucleocytoplasmic Transport Defects to FTD/ALS.” *Nature Neuroscience* 18(9): 1226–29. <http://www.nature.com/doi/10.1038/nn.4085>.
- Jung, Marie et al. 2014. “Affinity Map of Bromodomain Protein 4 (BRD4) Interactions with the Histone H4 Tail and the Small Molecule Inhibitor JQ1.” *Journal of Biological Chemistry* 289(13): 9304–19.
- Kanno, Tomohiko et al. 2004. “Selective Recognition of Acetylated Histones by Bromodomain Proteins Visualized in Living Cells.” *Molecular Cell* 13(1): 33–43.
- . 2014. “BRD4 Assists Elongation of Both Coding and Enhancer RNAs by Interacting with Acetylated Histones.” *Nature Structural and Molecular Biology* 21(12): 1047–57. <http://dx.doi.org/10.1038/nsmb.2912>.
- Kaya-Okur, Hatice S. et al. 2019. “CUT&Tag for Efficient Epigenomic Profiling of Small Samples and Single Cells.” *Nature Communications* 10(1): 1–10.
- Kempfer, Rieke, and Ana Pombo. 2019. “Methods for Mapping 3D Chromosome Architecture.” *Nature Reviews Genetics*. <http://dx.doi.org/10.1038/s41576-019-0195-2>.
- Kjeldsen, Marianne Juel, Linda A. Corey, Kaare Christensen, and Mogens Laue Friis. 2003. “Epileptic Seizures and Syndromes in Twins: The Importance of Genetic Factors.” *Epilepsy*

Research 55: 137–46.

Kobow, K., and I. Blümcke. 2018. “Epigenetics in Epilepsy.” *Neuroscience Letters* 667: 40–46.

<https://doi.org/10.1016/j.neulet.2017.01.012>.

Korb, Erica et al. 2015. “BET Protein Brd4 Activates Transcription in Neurons and BET Inhibitor Jq1 Blocks Memory in Mice.” *Nature Neuroscience* 18(10): 1464–73.

Koval, Erica D. et al. 2013. “Method for Widespread MicroRNA-155 Inhibition Prolongs Survival in ALS-Model Mice.” *Human Molecular Genetics* 22(20): 4127–35.

Lee, Kyung-Ha et al. 2016. “C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles.” *Cell* 167(3): 774-788.e17.

<http://linkinghub.elsevier.com/retrieve/pii/S0092867416313836>.

Lee, Youn Bok et al. 2013. “Hexanucleotide Repeats in ALS/FTD Form Length-Dependent RNA Foci, Sequester RNA Binding Proteins, and Are Neurotoxic.” *Cell Reports* 5(5): 1178–86.

LeRoy, Gary, Brenden Rickards, and S. J. Flint. 2008. “The Double Bromodomain Proteins Brd2 and Brd3 Couple Histone Acetylation to Transcription.” *Molecular Cell* 30(1): 51–60.

Liu, Xiao et al. 2006. “Whole-Genome Comparison of Leu3 Binding in Vitro and in Vivo Reveals the Importance of Nucleosome Occupancy in Target Site Selection.” *Genome Research* 16: 1517–28.

Liu, Yuanjing et al. 2016. “C9orf72 BAC Mouse Model with Motor Deficits and Neurodegenerative Features of ALS/FTD.” *Neuron* 90: 1–14.

<http://dx.doi.org/10.1016/j.neuron.2016.04.005>.

Löscher, Wolfgang. 2011. “Critical Review of Current Animal Models of Seizures and Epilepsy Used in the Discovery and Development of New Antiepileptic Drugs.” *Seizure* 20(5): 359–

68. <http://dx.doi.org/10.1016/j.seizure.2011.01.003>.

Mackenzie, Ian R. et al. 2013. "Dipeptide Repeat Protein Pathology in C9ORF72 Mutation

Cases: Clinico-Pathological Correlations." *Acta Neuropathologica* 126(6): 859–79.

———. 2015. "Quantitative Analysis and Clinico-Pathological Correlations of Different

Dipeptide Repeat Protein Pathologies in C9ORF72 Mutation Carriers." *Acta*

Neuropathologica. <http://link.springer.com/10.1007/s00401-015-1476-2>.

Majounie, Elisa et al. 2012. "Frequency of the C9orf72 Hexanucleotide Repeat Expansion in

Patients with Amyotrophic Lateral Sclerosis and Frontotemporal Dementia: A Cross-

Sectional Study." *Lancet Neurology* 11: 323–30.

[http://www.thelancet.com/pdfs/journals/lanneur/PIIS1474-4422\(12\)70043-1.pdf](http://www.thelancet.com/pdfs/journals/lanneur/PIIS1474-4422(12)70043-1.pdf) (May 3,

2015).

Marzi, Sarah J et al. 2018. "A Histone Acetylome-Wide Association Study of Alzheimer's

Disease Identifies Disease-Associated H3K27ac Differences in the Entorhinal Cortex."

Nature Neuroscience 21: 1618–1627.

<http://biorxiv.org/lookup/doi/10.1101/183541%0Apapers3://publication/doi/10.1101/183541>

1.

Maurano, Matthew T. et al. 2012. "Systematic Localization of Common Disease-Associate

Variation in Regulatory DNA." *Science* 337(September): 1190–95.

May, Stephanie et al. 2014. "C9orf72 FTL/ALS-Associated Gly-Ala Dipeptide Repeat Proteins

Cause Neuronal Toxicity and Unc119 Sequestration." *Acta Neuropathologica*: 485–503.

McClelland, Shawn et al. 2011. "Neuron-Restrictive Silencer Factor-Mediated

Hyperpolarization-Activated Cyclic Nucleotide Gated Channelopathy in Experimental

Temporal Lobe Epilepsy." *Annals of Neurology* 70(3): 454–64.

- . 2014. “The Transcription Factor NRSF Contributes to Epileptogenesis by Selective Repression of a Subset of Target Genes.” *eLife* 3: e01267.
- Mizielinska, Sarah et al. 2014. “C9orf72 Repeat Expansions Cause Neurodegeneration in *Drosophila* through Arginine-Rich Proteins.” *Science* 345(9): 1192–95.
<http://www.ncbi.nlm.nih.gov/pubmed/19556136><http://www.ncbi.nlm.nih.gov/pubmed/25103406>.
- Mo, Alisa et al. 2015. “Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain.” *Neuron* 86(6): 1369–84. <http://dx.doi.org/10.1016/j.neuron.2015.05.018>.
- Mochizuki, Kazuki et al. 2008. “The Bromodomain Protein Brd4 Stimulates G1 Gene Transcription and Promotes Progression to S Phase.” *Journal of Biological Chemistry* 283(14): 9040–48.
- Moens, Thomas G. et al. 2018. “Sense and Antisense RNA Are Not Toxic in *Drosophila* Models of C9orf72-Associated ALS/FTD.” *Acta Neuropathologica*.
<http://link.springer.com/10.1007/s00401-017-1798-3>.
- Molliex, Amandine et al. 2015. “Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization.” *Cell* 163(1): 123–33.
<http://dx.doi.org/10.1016/j.cell.2015.09.015>.
- Mori, Kohji et al. 2013. “The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTL/ALS.” *Science* 339(March): 1335–39.
<http://www.sciencemag.org/content/suppl/2013/02/06/science.1232927.DC1.full> (May 7, 2015).
- Myers, Candace T, and Heather C Mefford. 2015. “Advancing Epilepsy Genetics in the Genomic Era.” *Genome medicine* 7(1): 91.

<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4549122&tool=pmcentrez&rendertype=abstract>.

- Nativio, Raffaella et al. 2018. “Dysregulation of the Epigenetic Landscape of Normal Aging in Alzheimer’s Disease.” *Nature Neuroscience* 21(4): 497–505.
- Neph, Shane et al. 2012. “An Expansive Human Regulatory Lexicon Encoded in Transcription Factor Footprints.” *Nature* 489(7414): 83–90. <http://dx.doi.org/10.1038/nature11212>.
- Nguyen, Lien et al. 2019. “Antibody Therapy Targeting RAN Proteins Rescues C9 ALS/FTD Phenotypes in C9orf72 Mouse Model.” *Neuron* 105(February 19): 1–18.
<https://linkinghub.elsevier.com/retrieve/pii/S0896627319309705>.
- Nihei, Yoshihiro et al. 2019. “Poly-glycine–Alanine Exacerbates C9orf72 Repeat Expansion-mediated DNA Damage via Sequestration of Phosphorylated ATM and Loss of Nuclear HnRNPA3.” *Acta Neuropathologica* 23 October(<https://doi.org/10.1007/s00401-019-02082-0>). <https://doi.org/10.1007/s00401-019-02082-0>.
- Nordin, A. et al. 2015. “Extensive Size Variability of the GGGGCC Expansion in C9orf72 in Both Neuronal and Non-Neuronal Tissues in 18 Patients with ALS or FTD.” *Human Molecular Genetics* 24(February): 1–10.
<http://www.hmg.oxfordjournals.org/cgi/doi/10.1093/hmg/ddv064>.
- Nott, Alexi et al. 2019. “Cell Type-Specific Enhancer-Promoter Connectivity Maps in the Human Brain and Disease Risk Association.” *Science* 10.1126/sc.
<https://www.biorxiv.org/content/10.1101/778183v1.full>.
- O’Rourke, J G et al. 2016. “C9orf72 Is Required for Proper Macrophage and Microglial Function in Mice.” *Science* 351(6279).
- O’Rourke, Jacqueline G. et al. 2015. “C9orf72 BAC Transgenic Mice Display Typical

- Pathologic Features of ALS/FTD.” *Neuron* 88(5): 892–901.
<http://dx.doi.org/10.1016/j.neuron.2015.10.027>.
- Pal, Deb K. et al. 2003. “BRD2 (RING3) Is a Probable Major Susceptibility Gene for Common Juvenile Myoclonic Epilepsy.” *American Journal of Human Genetics* 73(2): 261–70.
- Patel, Avinash et al. 2015. “A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation.” *Cell* 162(5): 1066–77.
<http://dx.doi.org/10.1016/j.cell.2015.07.047>.
- Pennacchio, Len A et al. 2013. “Enhancers: Five Essential Questions.” *Nature Reviews Genetics* 14(4): 288–95.
- Qureshi, Irfan A., and Mark F. Mehler. 2014. “Sex, Epilepsy, and Epigenetics.” *Neurobiology of Disease* 72(PB): 210–16. <http://dx.doi.org/10.1016/j.nbd.2014.06.019>.
- Reichenstein, Irit et al. 2019. “Human Genetics and Neuropathology Suggest a Link between MiR-218 and Amyotrophic Lateral Sclerosis Pathophysiology.” *Science Translational Medicine* 11(eaav5264; 18 December): 1–12.
- Renton, Alan E. et al. 2011. “A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD.” *Neuron* 72(2): 257–68.
- Renton, Alan E., Adriano Chiò, and Bryan J. Traynor. 2014. “State of Play in Amyotrophic Lateral Sclerosis Genetics.” *Nature Neuroscience* 17(1): 17–23.
<http://dx.doi.org/10.1038/nn.3584>.
- Rogawski, Michael A. 2006. “Diverse Mechanisms of Antiepileptic Drugs in the Development Pipeline.” *Epilepsy Research* 69(3): 273–94.
- Rosen, Daniel R et al. 1993. “Mutations in Cu/Zn Superoxide Dismutase Gene Are Associated with Familial Amyotrophic Lateral Sclerosis.” *Nature* 362(4 March): 59–62.

- Rossaert, Elisabeth et al. 2019. “Restoration of Histone Acetylation Ameliorates Disease and Metabolic Abnormalities in a FUS Mouse Model.” *Acta Neuropathologica Communications* 7(1): 1–19.
- Rowland, Lewis P, and Neil A Shneider. 2001. “Amyotrophic Lateral Sclerosis.” *New England Journal of Medicine* 344(22): 1688–1700.
- Sabari, Benjamin R. et al. 2018. “Coactivator Condensation at Super-Enhancers Links Phase Separation and Gene Control.” *Science* 361(6400): eaar3958.
<http://science.sciencemag.org/>.
- Schludi, Martin H. et al. 2017. “Spinal Poly-GA Inclusions in a C9orf72 Mouse Model Trigger Motor Deficits and Inflammation without Neuron Loss.” *Acta Neuropathologica*.
<http://link.springer.com/10.1007/s00401-017-1711-0>.
- Schmid, Manfred, Thérèse Durussel, and Ulrich K. Laemmli. 2004. “ChIC and ChEC: Genomic Mapping of Chromatin Proteins.” *Molecular Cell* 16(1): 147–57.
- Shlyueva, Daria, Gerald Stampfel, and Alexander Stark. 2014. “Transcriptional Enhancers: From Properties to Genome-Wide Predictions.” *Nature Reviews Genetics* 15(4): 272–86.
<http://dx.doi.org/10.1038/nrg3682>.
- Skene, Peter J., Jorja G. Henikoff, and Steven Henikoff. 2018. “Targeted in Situ Genome-Wide Profiling with High Efficiency for Low Cell Numbers.” *Nature Protocols* 13(5): 1006–19.
<http://dx.doi.org/10.1038/nprot.2018.015>.
- Slattery, Matthew et al. 2011. “Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins.” *Cell* 147(6): 1270–82.
<http://dx.doi.org/10.1016/j.cell.2011.10.053>.
- Southall, Tony D. et al. 2013. “Cell-Type-Specific Profiling of Gene Expression and Chromatin

- Binding without Cell Isolation: Assaying RNA Pol II Occupancy in Neural Stem Cells.”
Developmental Cell 26(1): 101–12. <http://dx.doi.org/10.1016/j.devcel.2013.05.020>.
- Spitz, François, and Eileen E M Furlong. 2012. “Transcription Factors: From Enhancer Binding to Developmental Control.” *Nature Reviews Genetics* 13(9): 613–26.
<http://dx.doi.org/10.1038/nrg3207>.
- Staley, Kevin. 2015. “Molecular Mechanisms of Epilepsy.” *Nature Neuroscience* 18(3): 367–72.
<http://www.nature.com/doi/10.1038/nn.3947>.
- van Steensel, Bas, and Steven Henikoff. 2000. “Identification of in Vivo DNA Targets of Chromatin Proteins Using Tethered Dam Methyltransferase.” *Nature Biotechnology* 18(April).
- Swinnen, Bart et al. 2018. “A Zebrafish Model for C9orf72 ALS Reveals RNA Toxicity as a Pathogenic Mechanism.” *Acta Neuropathologica* (0123456789).
<http://link.springer.com/10.1007/s00401-017-1796-5>.
- Taylor, J. Paul, Robert H. Brown, and Don W. Cleveland. 2016. “Decoding ALS: From Genes to Mechanism.” *Nature* 539(7628): 197–206.
<http://www.nature.com/doi/10.1038/nature20413>.
- Varcianna, André et al. 2019. “Micro-RNAs Secreted through Astrocyte-Derived Extracellular Vesicles Cause Neuronal Network Degeneration in C9orf72 ALS.” *EBioMedicine* 40: 626–35.
- Wang, Haoyi et al. 2011. “Calling Cards Enable Multiplexed Identification of the Genomic Targets of DNA-Binding Proteins.” *Genome Research* 21(5): 748–55.
- . 2012. “‘Calling Cards’ for DNA-Binding Proteins in Mammalian Cells.” *Genetics* 190(3): 941–49.

- Wen, Xinmei et al. 2014. “Antisense Proline-Arginine RAN Dipeptides Linked to C9ORF72 - ALS / FTD Form Toxic Nuclear Aggregates That Initiate In Vitro and In Vivo Neuronal Death.” *Neuron* 84(6): 1213–25. <http://dx.doi.org/10.1016/j.neuron.2014.12.010>.
- White, M. A., C. A. Myers, J. C. Corbo, and B. A. Cohen. 2013. “Massively Parallel in Vivo Enhancer Assay Reveals That Highly Local Features Determine the Cis-Regulatory Function of ChIP-Seq Peaks.” *Proceedings of the National Academy of Sciences* 110(29): 11952–57. <http://www.pnas.org/cgi/doi/10.1073/pnas.1307449110>.
- Whyte, Warren A. et al. 2013. “Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes.” *Cell* 153(2): 307–19. <http://linkinghub.elsevier.com/retrieve/pii/S0092867413003929>.
- Yanez-Cuna, J. Omar et al. 2012. “Uncovering Cis-Regulatory Sequence Requirements for Context-Specific Transcription Factor Binding.” *Genome Research*: 2018–30.
- Zentner, Gabriel E et al. 2015. “ChEC-Seq Kinetics Discriminates Transcription Factor Binding Sites by DNA Sequence and Shape in Vivo.” *Nature Communications* 6(8733). <http://dx.doi.org/10.1038/ncomms9733>.
- Zhang, Ke et al. 2015. “The C9orf72 Repeat Expansion Disrupts Nucleocytoplasmic Transport.” *Nature*. <http://www.nature.com/doi/10.1038/nature14973>.
- Zhang, Yong-Jie et al. 2014. “Aggregation-Prone C9FTD/ALS Poly(GA) RAN-Translated Proteins Cause Neurotoxicity by Inducing ER Stress.” *Acta Neuropathologica* 128(4): 505–24. <http://link.springer.com/10.1007/s00401-014-1336-5>.
- . 2019. “Heterochromatin Anomalies and Double-Stranded RNA Accumulation Underlie C9orf72 Poly(PR) Toxicity.” *Science* 363(6428): eaav2606. <http://science.sciencemag.org/content/363/6428/eaav2606/tab-pdf>.

Zu, Tao et al. 2013. "RAN Proteins and RNA Foci from Antisense Transcripts in C9ORF72 ALS and Frontotemporal Dementia." *Proceedings of the National Academy of Sciences of the United States of America* 110(51): E4968-77.

<http://www.pnas.org/content/110/51/E4968.full> (May 3, 2015).

CHAPTER 2

A viral toolkit for recording transcription factor-DNA interactions in live mouse tissues

Preface

This chapter contains contents from an *in press* manuscript:

A viral toolkit for recording transcription factor-DNA interactions in live mouse tissues

Alexander J. Cammack¹, Arnav Moudgil^{2,3,4}, Jiayang Chen³, Michael J. Vasek³, Mark Shabsovich¹, Katherine McCullough³, Allen Yen³, Tomas Lagunas³, Susan E. Maloney³, June He^{2,3}, Xuhua Chen^{2,3}, Misha Hooda¹, Michael N. Wilkinson^{2,3}, Timothy M. Miller¹, Robi D. Mitra^{2,3}, Joseph D. Dougherty^{3,5}

¹Washington University School of Medicine, Department of Neurology, St. Louis, MO

²Washington University School of Medicine, Edison Family Center for Genome Sciences and Systems Biology, St. Louis, MO

³Washington University School of Medicine, Department of Genetics, St. Louis, MO

⁴Washington University School of Medicine, Medical Scientist Training Program, St. Louis, MO

⁵Washington University School of Medicine, Department of Psychiatry, St. Louis, MO

Abstract

Transcription factors (TFs) enact precise regulation of gene expression through site-specific, genome-wide binding. Common methods for TF occupancy profiling, such as chromatin immunoprecipitation, are limited by requirement of TF-specific antibodies and provide only endpoint snapshots of TF binding. Alternatively, TF-tagging techniques, in which a TF is fused to a DNA-modifying enzyme that marks TF binding events across the genome as they occur, do not require TF-specific antibodies and offer the potential for unique applications, such as recording of TF occupancy over time and cell type-specificity through conditional expression of the TF-enzyme fusion. Here we create a viral toolkit for one such method, calling cards, and demonstrate that these reagents can be delivered to the live mouse brain and used to report TF occupancy. Further, we establish a Cre-dependent calling cards system and, in proof-of-principle experiments, show utility in defining cell type-specific TF profiles and recording and integrating TF binding events across time. This versatile approach will enable unique studies of TF-mediated gene regulation in live animal models.

Introduction

Proper cellular development and function is a complex process established by elaborate gene expression networks. These networks are fundamentally regulated by transcription factors (TF), which bind to regulatory elements (RE) across the genome and facilitate gene expression through a variety of mechanisms, including recruitment of transcriptional co-factors and modulation of chromatin state (Spitz and Furlong 2012). Extensive efforts to profile TF genome occupancy and identify active REs across the genome have highlighted the profound diversity of TF binding, providing important insights into TF-mediated gene regulation (Consortium 2012; Ernst et al. 2011; Heintzman et al. 2009; Mo et al. 2015). Further, a large portion of genetic variation associated with improper cellular function or disease has been shown to exist in TF-bound REs (Corradin and Scacheri 2014; Ernst et al. 2011; Girdhar et al. 2018; Hnisz et al. 2013; Maurano et al. 2012; Wells et al. 2015), demonstrating the criticality of proper TF binding in maintaining cellular homeostasis.

Several methods exist for profiling TF occupancy across the genome. Antibody-based techniques, such as chromatin immunoprecipitation followed by sequencing (ChIP-seq), and more recently Cleavage Under Targets and Release Using Nuclease (CUT&RUN)(Skene, Henikoff, and Henikoff 2018) or Tagmentation (CUT&Tag) (Kwon et al. 2017), are widely used and have provided numerous insights into the cellular functions of TFs (Consortium 2012; Ernst et al. 2011; Heintzman et al. 2009; Hnisz et al. 2013). Notably however, these methods require the availability and individual optimization of TF-specific antibodies, limiting the throughput and breadth of genome-wide TF profiling. Further, ChIP-seq provides only a snapshot of TF activity at the moment of cell lysis and thus may be inefficient at detecting transient or infrequent TF binding events. Finally, while robust for non-cell type-selective, tissue-level analyses, it is

often challenging to interpret ChIP-seq data obtained from complex tissues such as the brain, which is comprised of many different interconnected cell types. Because of this limitation, efforts have recently been made to modify ChIP-seq for cell type-specific use, either through physical nuclear sorting (Bonn et al. 2012; Girdhar et al. 2018), or conditional expression and subsequent isolation of tagged nuclei (Deal and Henikoff 2010; Mo et al. 2015) or chromatin-associated enzymes (Zhou et al. 2017) prior to ChIP. However, these methods thus far have been limited to highly abundant targets, such as histone modifications (Girdhar et al. 2018; Mo et al. 2015) and transcriptional coactivators (Zhou et al. 2017), and may be complicated by potential disassociation-related artifacts (van den Brink et al. 2017). Therefore, it is unclear if ChIP-seq is feasible from sorted or isolated nuclei for less abundant TFs.

An alternative approach is to record TF binding events by fusing the TF of interest to DNA-modifying enzymes (Schmid, Durussel, and Laemmli 2004; van Steensel and Henikoff 2000; Wang et al. 2012; Zentner et al. 2015). Prominent among these are two techniques: DNA adenine methylation identification (DamID) (van Steensel and Henikoff 2000), which records TF binding through local adenine methylation by an *E. coli* Dam methylase fused to a TF of interest, and calling cards (Wang et al. 2011, 2012), in which a TF is fused to a transposase enzyme and binding events are recorded through transposon insertion proximal to the TF binding site. Importantly, TF-tagging techniques do not require TF-specific antibodies and have the ability to record and integrate occupancy information across time (Mitchell et al. 2016), while requiring very little starting material (Southall et al. 2013). Further, these approaches offer the potential for cell type-specificity through conditional expression of the TF-enzyme fusion protein. In this way, DamID has been successfully implemented for cell type-specific profiling (van den Aamele, Krautz, and Brand 2019), primarily in *Drosophila* (Southall et al. 2013) but also with some

studies in cultured mammalian cells (Cheetham et al. 2018; Tosti et al. 2018; Vogel et al. 2006) and embryos (Tosti et al. 2018). Meanwhile, calling cards has also been successfully applied to yeast (Wang et al. 2008) and mammalian cell (Wang et al. 2012) model systems. However, neither of these methodologies has to date been implemented for TF recording in postnatal mammalian model systems, such as mice.

Here we adapt calling cards for *in vivo* use by delivering this system to the mouse brain via adeno-associated virus (AAV). This method, in the mold of traditional calling card technologies (Wang et al. 2012), works by first expressing the *hyperPiggyBac* (hypPB) transposase within a cell and providing donor transposons. HypPB inserts donor transposons at TTAA sites throughout the genome, leaving permanent marks, or calling cards, at these loci. These transposons can later be sequenced and mapped to the genome to record the history of hypPB localization across the genome. HypPB-mediated insertions can be used to assess TF binding in two ways: 1) hypPB may be fused to a TF of interest, so that the TF directs the insertion of transposons near its genomic binding sites (Wang et al. 2012), or 2) unfused hypPB directly interacts with the bromodomain and extra-terminal domain (BET) protein, BRD4, and directs transposon DNA into BRD4-associated genomic regions (Gogol-Doring et al. 2016; Yoshida et al. 2017), most prominently active super enhancers (Hnisz et al. 2013). We establish that calling card systems can be delivered to the mouse brain via AAV and that these components successfully record TF occupancy without the need for a TF-specific antibody. We then create a conditionally-expressed, Cre recombinase-dependent version of AAV calling cards, termed Flip-Excision, or FLEX, calling cards and demonstrate, as a proof-of-principle, the ability of this system to record cell type-specific TF occupancy profiles in the brain. Lastly, we provide evidence that through continued transposon insertion, FLEX calling cards can record and

integrate TF binding events over extended time periods following viral delivery, providing insights into transient TF activity that would be otherwise missed with endpoint measures such as ChIP-seq.

Methods

Animals

All animal practices and procedures were approved by the Washington University in St. Louis Institutional Animal Care and Use Committee (IACUC) in accordance with National Institutes of Health (NIH) guidelines. Transgenic mouse strains used in this study include Synapsin 1 (Syn1)::Cre (RRID:IMSR_JAX:003966) and glial fibrillary acidic protein (GFAP)::Cre (RRID:IMSR_JAX:024098). All mice were bred to the C57BL/6J background, with the exception of animals used for adult stereotactic injections, which were wild-type animals of the FVB/N6 background (RRID:IMSR_JAX:001800). At indicated endpoints, mice were anesthetized with isoflurane and perfused with 15ml of cold saline (PBS) prior to tissue collection. Animals in the BRD4 Syn1::Cre, BRD4 GFAP::Cre, and P28 SP1 Syn1::Cre cohorts received pentylenetetrazole-induced seizures immediately prior to sacrifice. Unless otherwise noted, brains were either dissected and flash frozen in liquid nitrogen (for molecular analyses) or fixed in 4% paraformaldehyde 24-48 hours, exchanged into 30% sucrose, and either directly frozen at -80°C or cryoprotected in O.C.T. (for immunofluorescence).

To collect newborn litters, pregnant females were monitored daily until giving birth. Newborn pups were injected within 24 hours of being found in the cage, i.e. 0-48 hours after birth; thus, these injections are labeled “P0-1.” One exception is AAV::hypPB Frontflip in **Fig 2.6**, for which 1/3 of both the GFAP::Cre(+) and GFAP::Cre(-) animals were P3-4 when injected.

Cell culture, transfections, and flow cytometry analysis

HEK293T and N2a cells used in this study were cultured in 1X DMEM with 10% fetal

bovine serum (FBS) and grown under standard conditions (37°C; 5% CO₂). Plasmid transfections in HEK293T cells were carried out with Fugene® 6 Transfection Reagent (Promega, Madison, WI, USA) with the manufacturer's protocol. Calling cards constructs were delivered to N2a cells via either Fugene 6 or Neon Electroporation (ThermoFisher #MPK10025) with the following settings: 1050V, 30ms, 2 pulses.

For evaluation of SRT-tdTomato fluorescence in **Fig 2.6B**, HEK293T cells were collected 96-hours post-transfection for flow cytometry analysis. The cells were lifted using 0.25% trypsin-EDTA (ThermoFisher 25200056) and washed once with PBS. Cell pellets were resuspended at 1x10⁶ cells/ml in PBS/0.5% BSA/2mM EDTA with 0.1µg/ml DAPI (Sigma-Aldrich D9542, St Louis, MO, USA) for 10 min for dead cell and debris exclusion from analysis. Cells were washed once more with PBS/0.5% BSA/2mM EDTA and analyzed on a CytoFLEX S (Beckman Coulter), with each sample being run at a low flow rate (10-30µl/min) for 3 min to collect approximately 20,000 events per condition. A violet 405nm laser with 450/45 bandpass filter was used to detect DAPI and a yellow-green 561nm laser with 585/42 bandpass filter was used to collect tdTomato fluorescence. Flow cytometry data was analyzed with CytExpert software (Beckman Coulter). The gating was performed as outlined in **Fig 2.8SA**.

Immunofluorescence and imaging

10µm-thick (for co-localization studies) or 40µm-thick (for imaging of AAV::BrokenHeart) fixed-frozen sagittal or coronal brain sections were washed with PBS and permeabilized with 0.1% Triton X-100 (Sigma-Aldrich, St. Louis, MO, USA). Non-specific binding was blocked with 5% normal donkey (Jackson ImmunoResearch, West Grove, PA,

USA) or goat (Vector Laboratories, Burlingame, CA, USA) serum for 30-60 minutes at room temperature. After blocking, slides were exposed to primary antibody overnight at 4°C, washed three times with PBS, and then incubated with secondary antibodies for 1 hour at room temperature. Nuclei were counterstained with DAPI (Sigma-Aldrich, St. Louis, MO, USA) and coverslips were applied with ProLong Gold Antifade (ThermoFisher, Waltham, MA, USA) or Fluoromount-G (SouthernBiotech, Birmingham, AL, USA) mounting media. Immunofluorescent images of brain sections were acquired with a Nikon A1Rsi or Zeiss LSM 700 confocal microscope and imported into ImageJ (v. 1.51s) for manual cell counts and quantification. For analyses of hypPB expression in various cell types, 5 mice were used, and co-localization was quantified in 2 cortical images from a single section per animal. Antibodies used for immunostaining included chicken anti-GFP (Aves Labs GFP-1020) at 1:1000 dilution, mouse rabbit anti-RFP at 1:400 or 1:500 dilution (Rockland 600-401-379), anti-NeuN at 1:100 dilution (Millipore-Sigma MAB377), rabbit anti-cMyc at 1:250 dilution (Sigma C3956), and goat anti-GFAP at 1:500 dilution (Abcam ab53554).

Cells transfected with AAV::hypPB FLEX for testing Cre-dependence were live imaged for TdTomato on a Leica DMI 3000B tissue culture microscope. All images were acquired with equal conditions and exposure times for direct comparison.

In situ hybridization and imaging

10µM-thick, 4% paraformaldehyde fixed-frozen sections were cut and slide-mounted. mRNA encoding for *hyperPiggyBac* (VF1-20268-01) was detected using a custom probe-set designed by Affymetrix (now ThermoFisher) using the Affymetrix ViewRNA ISH Tissue 1-Plex kit (ThermoFisher, QVT0050) and chromogenic signal amplification kit (ThermoFisher

VT0200) with the following modifications: Slides were immersed in 4% paraformaldehyde overnight at 4°C prior to *in situ* hybridization, then the baking, deparaffinization, and heat pretreatment steps were omitted (steps 1-3, 5) because sections were not embedded in paraffin. Slides were hybridized either with anti-hypPB or no-probe controls. Following the *in situ* labeling protocol, sections were labeled for 5 minutes with DAPI (1:20,000, Sigma D9542), washed with PBS, and a drop of prolong gold (ThermoFisher P36934) was added while applying the coverslip. Slides were then imaged at 20x magnification on a Zeiss LSM 700 confocal microscope using a Cy3 filterset to detect FastRed fluorescence.

Analysis of brain degeneration following viral injection

Mice were exposed to either the calling cards viruses (AAV::hypPB + AAV::SRT) or RFP-only virus (as a control) via P0-1 injection and sacrificed 28 days later for silver staining (a marker for cells irreversibly committed to cell death). Briefly, animals were heavily sedated and perfused with TRIS fixative in 4% paraformaldehyde followed by vibratome sectioning of brains at 75µM in the coronal plane. Every eighth section across the rostrocaudal extent of the brain was then silver stained as described previously (Noguchi, Nemmers, and Farber 2005). The number of degenerating neurons was quantified for each animal by a rater blind to treatment who counted the total number of silver positive neurons in the dorsal cortex of each section.

Mouse behavior, developmental milestones, and sensorimotor battery

Mouse behavior and development was monitored and compared between animals injected calling cards AAV reagents (AAV::hypPB and AAV::SRT) or AAV::RFP only. In addition to weight, which was measured at P8, P14, P25, tests were administered to assess attainment of

developmental milestones (P14 righting from back), anxiety related behavior (P25 1 hour behavior, recording time spent in the edge or center of cage), and balance/strength/coordination (P25/26 sensorimotor battery). Procedures were done as previously described (Grady et al. 2006; Wozniak et al. 2004), with two trials per animal. A break was allowed after the completion of the first set of test trials to avoid exhaustion effects, and the test order was reversed for the second trial for all animals. Walking initiation, ledge performance, platform performance, and pole performance were all administered at P25, while 60° inclined screen test, 90° inclined screen test, and inverted screen test were administered on P26. Prior to testing, mice were given a routine health check, and from this, two animals were excluded; one for runtiness and one for severe hydrocephaly likely derived from needle stick. Further, one litter of RFP-only animals was not able to complete all timepoints (cage flooding) and was thus also excluded. In total, 21 animals were included in the calling card group (11M/10F) and 24 in the RFP-only group (10M/14F), all of which were used for downstream analyses. Test administrators were blinded from treatment group identity during all testing.

Virus generation and injections

Transposase and donor transposon constructs were cloned into Cre-dependent (FLEX) or Cre-independent AAV transfer vectors and used for *in vitro* transfection or viral packaging. Plasmids were packaged into AAV by the Hope Center Viral Vectors Core at Washington University School of Medicine. For *in vivo* experiments involving P0-1 delivery, transposase and donor transposon viruses (mixed equally by volume) or undiluted RFP-only virus were intracranially injected into the cortex of postnatal day 0-1 (P0-1; 3 sites per hemisphere, 1 µl viral mix per site). For adult injections, viruses were delivered to P107 animals intraparenchymally

with stereotactic surgery, as previously described (DeVos and Miller 2013). Two sites were chosen for direct, unilateral cortical injection with coordinates relative to bregma of 1.25mm rostral; 1.5mm lateral; 0.55mm depth; and 1.06mm caudal; 1.5mm lateral; 0.55mm depth. 2 μ l of viral mix was delivered at a rate of 0.2 μ l/minute.

Viral titers (viral genomes per milliliter) were as follows:

AAV::hypPB	1.0x10 ¹³ - 1.1x10 ¹³ vg/ml
AAV::hypPB FLEX	8.0x10 ¹² - 1.0x10 ¹³ vg/ml
AAV::SP1(621C)-hypPB FLEX	1.0x10 ¹³ vg/ml
AAV::SRT	1.6x10 ¹³ - 2.2x10 ¹³ vg/ml
AAV::BrokenHeart FLEX	1.4 x 10 ¹³ vg/ml
AAV::RFP only	1.6x10 ¹³ vg/ml
AAV::hypPB Frontflip	8.0x10 ¹³ vg/ml

SRT and BrokenHeart library preparation, sequencing, and mapping

SRT libraries were prepared from cortex RNA samples. Prior to library preparation, cortex samples were dissected into 10 separate pieces, from which RNA was independently isolated with the manufacturer's protocol (Qiagen RNeasy kit, Germantown, MD, USA). This allows for identification of up to 10 independent insertion events into any TTAA site, given that these insertions occur in spatially separate samples. From these RNA samples, transposon sequencing libraries were generated with our bulk SRT protocol (Moudgil et al. 2019). In brief, RNA samples were first reverse transcribed, from which self-reporting transcripts, including flanking genomic sequences, were amplified via PCR. These amplicons were then tagged with universal Illumina sequencing overhangs, with separate indexes for libraries from each dissected

piece, allowing for 10 ‘barcodes’ per sample and sequenced on Illumina HiSeq 2500, NextSeq 500, or MiniSeq platforms (Illumina, San Diego, CA, USA).

BrokenHeart libraries were prepared from cortical DNA samples, as previously described (Wang et al. 2012). A set of 20 individually barcoded BrokenHeart transposons were pooled and packaged into AAV. Thus, independent insertions into the same TTAA can be uniquely identified via barcode, removing the need to dissect and process tissue samples in separate pools as with SRT. Extracted DNA was self-ligated, amplified with inverse PCR, and sequenced with the Illumina NextSeq 500 platform (Illumina, San Diego, CA, USA).

Sequencing reads obtained from SRT and BrokenHeart libraries were stringently filtered for features of true insertion events (presence of *piggyBac* terminal repeat sequence; intact sequencing adapters, barcodes, and indexes; and a TTAA site) and mapped to build version mm10 of the mouse genome with Novoalign 3 (Novocraft Technologies) (for SRT) or Bowtie2 (Langmead et al. 2009) (for BrokenHeart). Reads aligning to the same TTAA with separate barcodes were considered unique insertions, and all analyses in this report considered all unique insertions equally, independent of read depth. We used BEDtools intersect to count the number of insertions directed to introns, exons, 3'- and 5'-UTRs, and intergenic regions, using annotations from the HOMER package (Heinz et al. 2010) (**Fig 2.2SE**).

Significant insertion peak calling and motif discovery

Significantly enriched insertion peaks were identified via a count-based statistical comparison as previously described (Moudgil et al. 2019). In brief, this pipeline first segments the genome into blocks of constant insertion density. For each block, it calculates the p-value of insertion enrichment relative to a background model assuming uniformly distributed insertions.

A user-defined significance threshold is defined, and all blocks surpassing this threshold are considered “significantly enriched insertion peaks”. This “background-free” method for unbiased identification of all significantly enriched genomic regions in a single experimental sample, used here in **Fig 2.2**, is expected to identify all BRD4-bound regions within the parameters of the calling cards system.

Alternatively, we can define differentially-bound regions between two experimental samples, as was done in **Fig 2.3, Fig 2.6, Fig 2.5S, and Fig 2.6S** for astrocyte- or neuron-specific BRD4 peaks (using the combined insertion pools from three GFAP::Cre [n = 3,072,163 insertions] and three Syn1::Cre animals [n = 2,484,133 insertions]) and **Fig 2.4** for SP1 peaks over unfused hypPB. In this analysis, the pipeline again segments the genome into blocks, but then assigns a p-value to each block based on the differential enrichment between the two samples. As with the background-free pipeline, a user defined p-value threshold is chosen, below which all blocks are considered significantly enriched. Because of the natural BRD4-bias of hypPB, using this differential peak caller allows for identification of binding sites of specific TFs by selecting genomic locations with re-directed insertion density in a TF-hypPB profile compared to an unfused hypPB background.

TF motifs were identified with MEME-ChIP v4.11.2 motif discovery software (Bailey et al. 2009; Machanick and Bailey 2011) with `-zoops -meme-minw 6 -ccut 250`.

Defining enhancers and super enhancers

Since H3K27ac is a known marker of active enhancers (Creyghton et al. 2010; Heintzman et al. 2009) and super enhancers (Hnisz et al. 2013; Whyte et al. 2013), we utilized published P14 mouse cortex (Stroud et al. 2017) and N2a (Kwon et al. 2017) H3K27ac ChIP-Seq

datasets to define cortical and N2a enhancers super enhancers, respectively. As previously described (Hnisz et al. 2013; Whyte et al. 2013), we used the rank ordering of super enhancers (ROSE) v0.1 pipeline and the model-based analysis for ChIP-Seq (MACS) v1.4.1 peak finding algorithm (Zhang et al. 2008) with a p-value enrichment threshold of 10^{-9} to define enhancers and super enhancers. We then used the BEDtools suite (Quinlan and Hall 2010) to compare the coincidence of enhancers and super enhancers with our unfused hypPB calling cards insertion peaks (**Fig 2.2 and Fig 2.4SC-D**).

Additionally, for qualitative measures of histone modification enrichment at calling cards BRD4 insertion peaks (**Fig 2.2**), we used publicly available P0 mouse forebrain ChIP-Seq datasets from ENCODE (Quinlan and Hall 2010); specifically, H3K27ac (ENCSR094TTT), H3K4me1 (ENCSR465PLB), and H3K27me3 (ENCSR070MOK).

Super enhancer in vitro sensitivity and specificity

Sensitivity and specificity of calling cards peaks were assessed for super enhancer identification in N2a cells that were either transfected or electroporated with AAV::hypPB and AAV::SRT plasmids (**Fig 2.4S**). RNA was separately isolated from a total of 33 wells (i.e. barcodes) from 6-well plates, and SRTs were sequenced, generating 806,653 unique insertions, though of note, the majority (651,631) were derived from 12 barcodes that received plasmids via electroporation. 800,000 unique insertions were randomly selected from the total pool of 806,653 insertions, from which significantly enriched peaks were defined using our background-free peak calling method at a range of significance thresholds. These peaks were intersected (with BEDtools intersect) with known N2a super enhancers defined via a previously published N2a H3K27ac ChIP-seq dataset (Kwon et al. 2017), and sensitivity was defined as the percentage of

peaks intersecting super enhancers for each peak calling significance threshold. To then define specificity, we identified the “true negative” space of the genome, and assessed the percentage of true negative peaks intersected by calling cards peaks. To do this, we first identified any possible active enhancer region of the genome with MACS peak finding using a low-stringency significance threshold of 10^{-1} and subtracted these peaks from the mouse genome, creating a “true negative” genome. We then sampled peaks (with BEDtools shuffle) within this true negative genome of the same size distributions as the list of active super enhancers until we collected an average of 1X coverage across the genome. With a true negative space of 2,616,503,093 basepairs and a total super enhancer size of 23,143,876 basepairs, this required 114 random samplings, resulting in 85,158 true negative peaks. Finally, we intersected our calling card peaks with these true negative peaks, and specificity was defined as the percentage of true negative peaks not intersected by a calling card peak. Of note, we expect that unfused hypPB is driven to super enhancers via interaction with BRD4; thus, sensitivity and specificity measurements may be higher if compared to BRD4 occupancy rather than H3K27ac.

Analysis of enhancer- and promoter-associated gene expression

Gene expression has been shown to be preferentially regulated by proximal enhancer elements (Creyghton et al. 2010; Heintzman et al. 2009; Visel et al. 2009). Thus, since a cell type-specific mapping of enhancers to the genes they regulate is not available, we used proximity as an imperfect (Maurano et al. 2012) albeit widely used (Ernst et al. 2011) proxy. In our analyses of cell type-specific expression of genes near cell type-enriched BRD4 calling cards peaks (**Fig 2.3, Fig 2.6, and Fig 2.5S**), we first defined the nearest gene (or genes, if multiple intersected a calling cards peak) to each significant calling cards peak. These gene sets were then

filtered and the remaining genes were used for subsequent analyses. Gene sets were filtered as follows: 1) genes greater than 10,000 bases away from a differential insertion peak were removed, to eliminate low confidence gene-enhancer pairs, 2) genes near or overlapping multiple insertion peaks counted once, and 3) genes for which cell type-specific RNA expression data were unavailable in our comparison dataset were removed.

Unbiased cell type identification was completed with the Cell-type Specific Expression Analysis (CSEA) tool (Xu et al. 2014) (<http://genetics.wustl.edu/jdlab/csea-tool-2/>) using candidate gene sets near either GFAP::Cre enriched or Syn1::Cre enriched insertion peaks. For each set, we analyzed genes near the most enriched peaks for each cell type. For GFAP::Cre, this included 131 genes ($p < 10^{-21}$ for associated insertion peaks), of which 114 were present in CSEA reference sets and used for analysis. For Syn1:Cre, this included 123 genes ($p < 10^{-11}$), with 110 present in reference sets.

For comparison of SP1 binding and gene expression in **Fig 2.5A-B**, we utilized the mm10_knownCanonical gene set and mm10_TSS coordinates from the UCSC genes table. We defined promoter-proximal regions as ± 1000 bases from the TSS. We first filtered mm10_knownCanonical gene set to remove duplicates ($< 3\%$ of total genes) and then intersected gene coordinates with promoter proximal regions. After manually filtering to assign true promoters to each transcript (i.e. immediately upstream from TSS), we generated a list of unique promoter/gene combinations (24,528 unique genes) and compared insertion density and gene expression at these coordinates.

For comparison of P28 and P10 SP1 promoter insertions to RNA expression in **Fig 2.5C-G**, we utilized a previously published RNA-seq dataset (Lister et al. 2013) with RNA expression data available for week 1 (Wk1) and week 4 (Wk4), which correspond to $\sim P7$ and $\sim P28$,

respectively. Before assessing P28 or P10 insertion density at promoters, insertion profiles were downsampled such that each cohort had exactly 240,000 insertions per library (80,000 per mouse for P10, 3 mice; 60,000 per mouse for P28, 4 mice); thus, insertion totals could be directly compared without any normalization to library size. Further, this downsampling procedure eliminates the possibility that any given observed increase in insertion density at P28 was due to an overall increase in insertion total over time. We then calculated number of insertions at each unique promoter (using the list of unique promoter/gene combinations generated above) and removed any gene with no insertions at either timepoint (19,046 / 24,528 unique genes remaining). A pseudocount of 1 was added to promoter insertion totals for each gene at each timepoint prior to analysis. To eliminate noise due to low RNA expression and/or random low-frequency insertion events, we next removed any gene with <6 insertions combined between the P28 and P10 datasets (including the 2 pseudocounts) or <1 FPKM combined between Wk1 and Wk4 RNA-seq expression, leaving a final total of 4991 unique gene/promoter combinations which were used in subsequent analyses. This list of 4991 genes were divided into three categories, based on their RNA expression at Wk1 and Wk4: (1) early genes; $\log(\text{Wk4}/\text{Wk1 FPKM}) < -0.5$, (2) constitutive genes; $-0.5 < \log(\text{Wk4}/\text{Wk1 FPKM}) < 0.5$, and (3) late genes; $\log(\text{Wk4}/\text{Wk1 FPKM}) > 0.5$. Within these categories, SP1 occupancy was compared between the P28 and P10 cohorts.

Validation of astrocyte enhancer candidates with PALE (Stogsdill et al. 2017)

Candidate astrocyte-enriched enhancers were selected from the list of GFAP::Cre-enriched insertion peaks in **Fig 2.3** based on three criteria, which select for high-confidence astrocyte enhancers: 1) enhancer size (smaller than 3kB, for cloning purposes), 2) significant

GFAP::Cre enrichment (over Syn1::Cre), and 3) astrocyte-specific RNA expression of their nearest genes. These candidates were PCR amplified with primers listed in the table below and with an MluI overhang adapter (TGTAGGACGCGT) on either end, and cloned into the miniP-dsRed plasmid with MluI, upstream of the hsp68 minimal promoter. As a positive control, the canonical GFAP promoter (Sakers et al. 2017) was also cloned into this plasmid, in the same location.

To test efficacy of the candidates for enhancing dsRed expression, each plasmid was electroporated into lateral ventricle-proximal cells, along with a separate plasmid containing CFP driven by the canonical GFAP promoter. P0-1 pups were placed on a wet towel on wet ice for 10 minutes to anesthetize. Then plasmids were delivered to the lateral ventricle via intraventricular injection along with Fast Green FCF dye (Sigma, 2353-45-9), with coordinates approximately equidistant from the lambdoid suture and the eye, 2mm lateral to the sagittal suture, and 2mm depth to ensure lateral ventricle penetration. 1µl of DNA was delivered into one hemisphere for each mouse, at a concentration of 1µg/µl for GFAP::CFP and 0.5µg/µl for dsRed plasmids. Electroporation was induced for 5 square pulses, 50ms per pulse at 100V and 950ms inter-pulse intervals, sweeping the electrodes from the dorsal to lateral using ~25° angle intervals.

Pups were sacrificed at P7 or P21 and brains were collected and analyzed with immunohistochemistry for CFP and dsRed (CFP and dsRed stained with GFP and RFP antibodies, respectively, as indicated in methods). Individual astrocytes were imaged for intensity analysis in **Fig 2.3E-F**, **Fig 2.6SF,H**, and **Fig 2.7SA,C** using equivalent exposure settings. A region of interest (ROI) was defined around the astrocyte based on CFP only, and dsRed and CFP fluorescence was quantified within the ROI using ImageJ (v. 1.51s). This was done from 3 brains (34-42 cells) per condition at P7 and from 2 brains (28-32 cells) per condition

at P28 (except pGFAP, which only had 1 brain and 17 cells used in quantification at P28, as this was a positive control sample). The ratio of CFP:dsRed was calculated for each cell and averaged and compared across conditions for final assessment of dsRed enhancement. For P7 co-localization quantifications in **Fig 2.6SG**, dsRed(+) cells were manually counted from 3 brains (150 total cells) per condition, while P21 quantifications in **Fig 2.7SB** were taken from 2 brains (287-769 total cells) per condition (again, the positive control, pGFAP had 1 brain, 67 total cells used in quantifications).

eCandidate	Forward primer	Reverse primer
eRasa2	TTCATGAACTCTGTTACTAGTTTGT	TTTTAAACAGATGAGCTGGAGCC
eTaf4b	ATATTGGATCTCACTGGAGTTGC	TCAAAGTCTAGATTTAGGCATGAT
ePla2g7	ACAGAACAGACTCACATAAACTTGT	CCATTGTCACATCTAGTCATCAGT
eMms22l	GCTTATTTAAAATGAAAAGA	AAATTCCTTAAACCCCTG

Statistical analyses

Statistical tests were done with GraphPad Prism v8.1.2 and are detailed in figure legends. For box-and-whisker plots, central tendency line represents median, box represents 25th-75th percentiles, and whiskers represent minimum and maximum values.

Data availability

All raw and processed data is available through GEO accession GSE128493 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128493>) and the access token is eroromaodhabzkz. Figures containing raw data include: 2.2, 2.3, 2.4, 2.5, 2.6, 2.2S, 2.4S, 2.5S, and 2.6S. Supplemental tables containing information about significant peaks and figure-specific analyses are also available upon request.

Code availability

Calling card analysis software were developed previously (Moudgil et al. 2019) and are available upon request.

Results

Intracranial delivery of calling cards via AAV invokes widespread transposon insertion in the mouse cortex.

In order to perform transposon calling cards in mammalian cells, two basic components are required: the hypPB transposase (or a TF-hypPB fusion) and donor transposons (Wang et al. 2012). We sought to develop an *in vivo* method to efficiently deliver calling card components throughout the mouse brain to identify TF-associated sites. We first tested AAV as a means for calling card reagent delivery, as viral *piggyBac* delivery methods have been successful in other organ systems previously (Cooney, Singh, and Sinn 2015; Smith et al. 2015). We packaged a myc-tagged version of hypPB and donor transposons carrying TdTomato reporter genes into separate AAV serotype 9 (AAV9) vectors, which efficiently transduce neuron and astrocyte populations (Hammond et al. 2017), and intracranially injected these vectors into the cortices of postnatal day 0-1 (P0-1) mice. Animals were sacrificed at P21 for analysis (**Fig 2.1A**). We analyzed hypPB expression with *in situ* hybridization (**Fig 2.1SA**) and transposon-derived TdTomato immunofluorescence (**Fig 2.1B**) across the brain and observed widespread viral transduction in the neocortex, hippocampus, and inner brain structures. As expected with the AAV9 serotype (Hammond et al. 2017), the vast majority of transduced cell types were neurons and astrocytes (**Fig 2.1C-D**). These results demonstrate that calling card reagents can be efficiently delivered to the mouse brain by AAV.

Earlier implementations of the calling cards method (e.g. BrokenHeart) mapped transposon insertions by directly sequencing genomically-inserted transposon DNA (Wang et al. 2011, 2012) (**Fig 2.2SA-B**). However, our group recently developed a specialized calling cards donor transposon, termed a “self-reporting transposon” (SRT), which allows for amplification of

each insertion via RNA transcription and highly efficient mapping of transposition events through deep sequencing of transposon-derived RNA (Moudgil et al. 2019) (**Fig 2.2SA,C**). We first sought to directly compare traditional DNA calling cards (BrokenHeart) to RNA calling cards (SRT) in AAV systems. To do this, we intracranially injected P0-1 mice with AAV::hypPB and either AAV::BrokenHeart or AAV::SRT. At P21, we isolated DNA or RNA from cortex samples and generated and sequenced calling cards libraries (**Fig 2.2SB-C**). We found that SRT reads mapped much more consistently to the mouse genome than BrokenHeart, where the majority of alignment was to the original AAV episomal sequence. We then mapped transposon insertions from these reads. While insertions from the two methods reliably mapped to similar genomic locations, we were able to recover an order of magnitude greater total number of insertions from animals receiving AAV::SRT compared to those receiving AAV::BrokenHeart (**Fig 2.2SE-F**). In summary, we found that SRTs provide a much greater sensitivity for recovering insertion events from AAV than traditional DNA methods. Finally, we tested whether SRT calling cards could also be delivered efficiently via adult intraparenchymal stereotactical cortical injection to precise coordinates, as this is more standard for AAV delivery than P0-1 injection. Indeed, after delivery of AAV::hypPB and AAV::SRT to three P107 adult mice (euthanized one month later, at P136), we observed a near-equivalent read mapping rate (**Fig. 2.2SD**), insertion total (**Fig. 2.2SE**), and insertion localization (**Fig. 2.2SF**) as in P0-1 SRT delivery. Thus, AAV calling cards systems are functional *in vivo* and can be delivered to the mouse brain at various timepoints and to targeted locations.

AAV calling cards delivery to the mouse brain does not induce excess degeneration, weight loss, or behavioral/developmental defects.

An important question with all calling card technologies is whether continued transposon insertion is deleterious, particularly after long-term delivery to a living mammalian system such as the mouse brain. To address this question empirically, we intracranially delivered either calling card viruses (AAV::hypPB/AAV::SRT) or control viruses (RFP only) to P0-1 mouse pups, and then assessed neuronal degeneration, attainment of developmental milestones, anxiety related behavior, and balance/strength/coordination during the first 4 weeks of life. Four weeks after viral delivery, a small population of degenerating cells were found in the cortex, near the injection site, however there was no significant difference between calling card and RFP-only injected animals, suggesting that the observed degeneration was likely due to needle injury (**Fig 2.1E-F**). Likewise, no significant differences were observed in weight, righting from back, or edge/center dwelling, indicating that calling cards-injected animals develop normally and display no overt anxiety-like behavior (**Fig 2.1G and Fig 2.3SA-C**). Finally, in a sensorimotor battery, calling cards-injected animals displayed a largely normal phenotype, with only one test (60° inclined screen climbing) having a significant reduction compared to RFP-only controls. Further, these animals performed normally on the 90° inclined screen and inverted screen tests, which are even more difficult tests of balance and strength than the 60° inclined screen (**Fig 2.1H and Fig 2.3SD-I**). In summary, AAV calling card reagents did not result in excess degeneration, weight loss, or behavioral/developmental defects, suggesting that genomic transposon insertion does not introduce significant toxicity or deleterious effects to the animal.

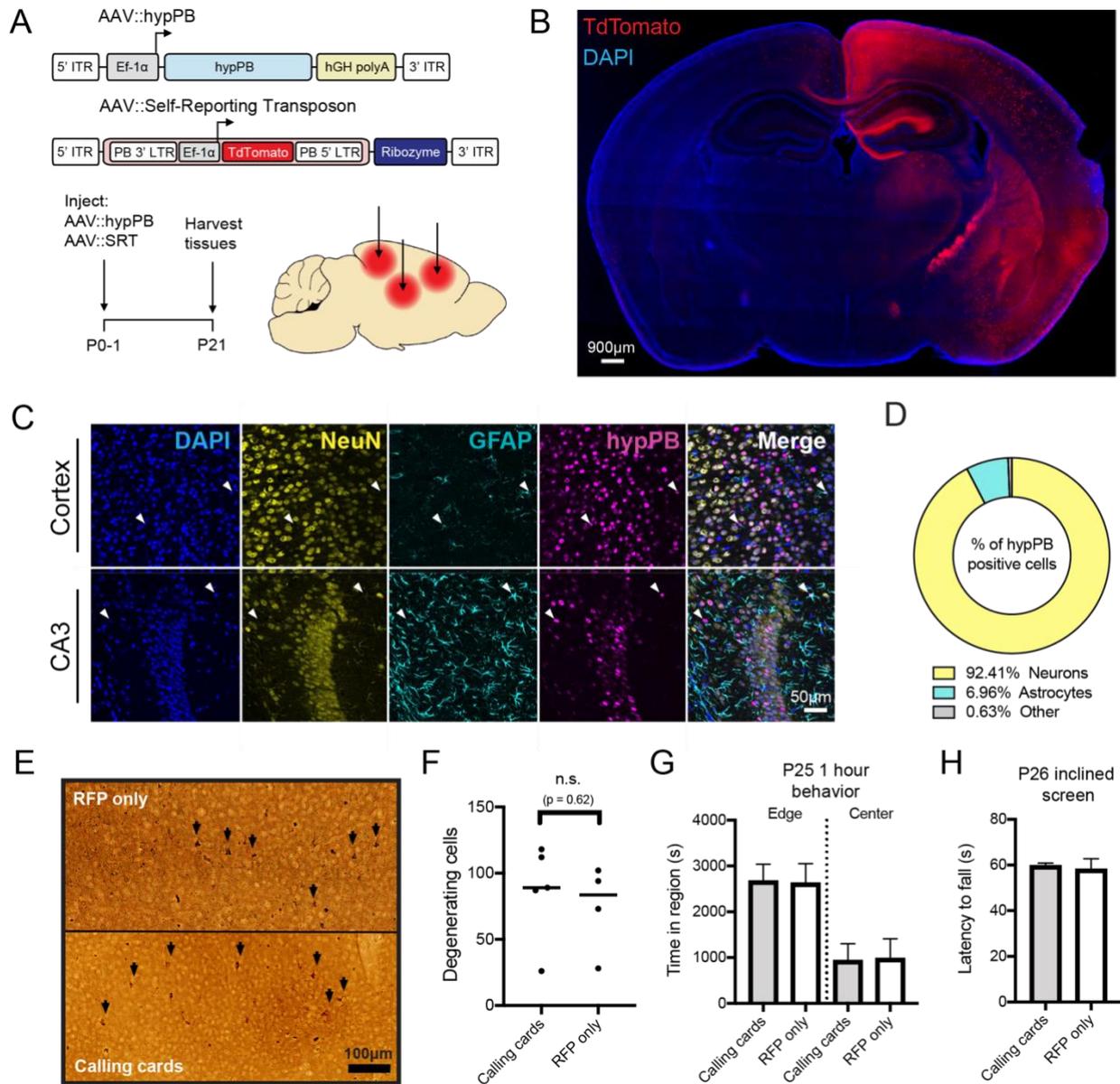


Figure 2.1. Co-AAV9 intracranial injection efficiently delivers calling cards to the cortex.

(A) Experimental paradigm and AAV constructs. Arrows represent approximate AAV injection sites. (B) Coronal section of brain injected unilaterally at P0-1 with AAV::hypPB and AAV::SRT, displaying widespread expression of SRT-derived TdTomato fluorescence throughout the brain (C,D) Co-immunofluorescence showing hypPB expression in neurons and astrocytes. (C) Representative images display co-localization of hypPB with neuronal (NeuN) and astrocyte (GFAP) markers in the cortex and hippocampus. HypPB is myc-tagged allowing for visualization with myc-specific antibodies. Arrowheads show examples of hypPB-positive astrocytes. (D) Majority of hypPB(+) cells transduced with AAV9 are NeuN(+) neurons and GFAP(+) astrocytes (n=1005 myc(+) cells, counted across cortical image fields from 5 mice). (E) Representative images of silver staining in dorsal cortex to screen for degenerating cells (black arrows) in mice intracranially injected at P0-1 with (top) red fluorescent protein (AAV::RFP only) or (bottom) calling cards viruses (AAV::hypPB and AAV::SRT) and sacrificed at P28. (F) Quantification of silver-positive cells in dorsal cortex revealed injection with either virus produces limited neurotoxicity that did not significantly differ between groups (two-tailed, unpaired Student's t-test, $p > 0.05$). (G-H) Mice injected at P0-1 with AAV calling cards (n=21) or control, RFP only (n=24) viruses displayed no significant differences in anxiety related behavior (center/edge dwelling; G) or motor skills (inclined screen test; H) relative to control. See Fig 2.3S for further behavioral and developmental assessments of these groups. All group comparisons were done with two-tailed, unpaired Student's t-test, with Bonferroni corrected $\alpha=0.05$ as a significance threshold (including all tests in Fig 3S).

Unfused hypPB delivered with AAV profiles active REs in the brain.

BRD4 acts as a “chromatin reader” by binding to acetylated lysine residues on histones (Jung et al. 2014; Kanno et al. 2004, 2014) and regulating gene transcription (LeRoy, Rickards, and Flint 2008; Mochizuki et al. 2008). Accordingly, BRD4 is strongly associated with active, TF-bound REs, most prominently super enhancers (Hnisz et al. 2013; Whyte et al. 2013). Importantly, BRD4 has a known affinity for the unfused hypPB protein (Gogol-Doring et al. 2016), and consequently unfused hypPB insertions are greatly enriched at known BRD4 binding sites (Gogol-Doring et al. 2016) such as super enhancers (Yoshida et al. 2017). Thus, we aimed to test the hypothesis that unfused hypPB transposon insertion can be used to identify active REs in the brain (**Fig 2.2A**).

We first analyzed the sensitivity and specificity of unfused hypPB insertions for identification of active super enhancers *in vitro* in neuroblastoma (N2a) cells. While N2a’s do not represent epigenetic regulation *in vivo* (only 20.9% of super enhancers are shared between N2a cells and P14 mouse cortex), this pure population allows for assessment of sensitivity and specificity without complex cellularity or serotype/transduction variability. To do this, we transfected these cells with hypPB and SRT plasmids, from which we collected a total of 806,653 insertions. We then downsampled this library into randomly selected pools of various insertion totals and used peak calling to identify regions of significantly enriched insertion density in each pool, at a range of significance thresholds. Using a previously published N2a H3K27ac ChIP-seq dataset (Kwon et al. 2017) to independently define active super enhancers in this population, we assayed sensitivity and specificity of calling card insertion peaks for identifying these regions. From this, we observed that calling card peaks are highly specific for active super enhancers across a range of sensitivities, with a high area under the receiver-

operator characteristic curve (0.82; **Fig 2.4SA**). Further, we observed a high sensitivity for super enhancer identification, even at low insertion totals (e.g. sensitivity of up to 0.8 from only 10,000 insertions), that increases steadily with increasing number of insertions (**Fig 2.4SB**). Thus, unfused hypPB calling card profiles can be used to identify active super enhancers *in vitro*.

We next asked whether AAV::hypPB could identify active REs, including super enhancers, in the brain. To do this, we combined all 3,732,694 insertions collected from two mice injected with AAV::hypPB and AAV::SRT at P0-1 and defined significantly enriched insertion peaks (7,031 peaks; $p < 10^{-30}$), which we predict to be BRD4-bound REs. We observed that insertion density at these peaks was highly correlated between the two animal replicates, indicating a high reproducibility of this method (**Fig 2.2B**). To assess whether insertion peaks represented active REs, we compared our calling card data to ENCODE ChIP-seq datasets (Hammond et al. 2017) of enhancer-associated histone modifications (Heintzman et al. 2009) in the developing mouse cortex. At the 7,031 significantly enriched insertion peaks, we observed a strong enrichment of active enhancer-associated histone modifications H3K27ac and H3K4me1 and a de-enrichment of the repressive mark H3K27me3 (**Fig 2.2C-F**). We then used a previously published (Stroud et al. 2017) H3K27ac ChIP-seq data from P14 mouse cortex to independently define active enhancers and super enhancers and asked whether calling card peaks significantly overlapped these regions. We observed that the majority of insertion peaks intersected with H3K27ac-defined enhancers, significantly higher than when insertion peak coordinates are randomized (**Fig 2.2G**). Similarly, calling card peak intersection with super enhancers is also significantly higher than chance (**Fig 2.2H**). As expected for a BRD4-mediated mechanism, unfused hypPB calling card profiles identify only a subset of all enhancers, but do intersect the majority of super enhancers (**Fig 2.2I-J**). Of note, the reference ChIP-seq dataset used for

enhancer and super enhancer identification encompasses REs from all cortical cell types, while calling card peaks are derived only from transduced cells (i.e. neurons and astrocytes), thus likely underrepresenting RE sensitivity. Further, this overlap analysis was performed using our standard, highly rigorous significance threshold for peak calling ($p=10^{-30}$), however we have also performed these analyses at a range of p-value thresholds to confirm the finding is robust to this parameter (**Fig 2.4SC-D**). Together, these data support that AAV calling card insertion profiles of unfused hypPB can be used to identify putative REs in the brain.

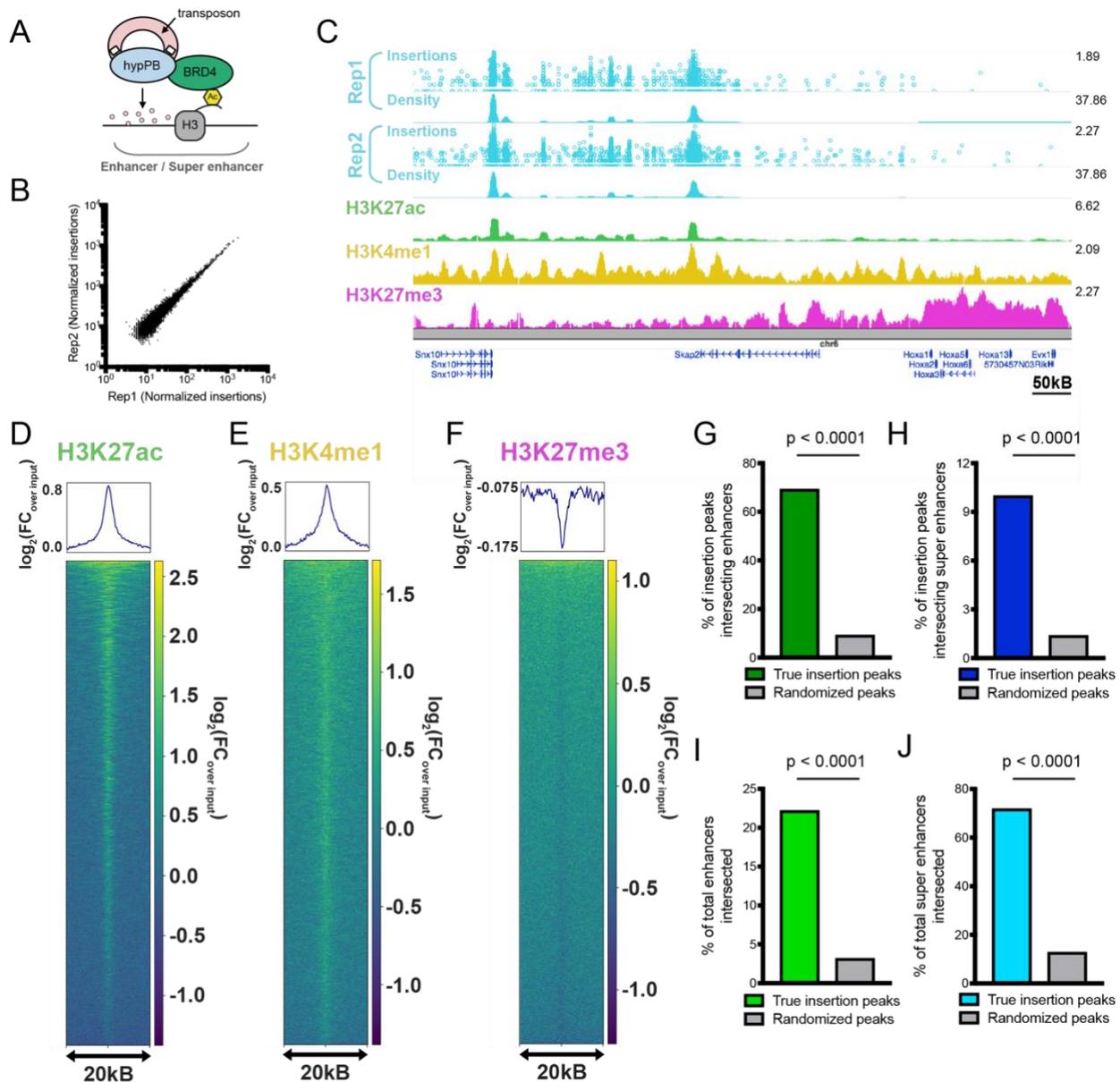


Figure 2.2. Unfused hypPB-directed calling cards insertions identify active enhancers and super enhancers in the brain.

(A) Unfused hypPB endogenously interacts with BRD4 and is redirected toward sites of BRD4 occupancy, i.e. enhancers and super enhancers. (B) Normalized insertion totals in two littermate C57BL/6J mice (Rep1 and Rep2) at 7031 significantly-enriched insertion peaks ($p < 10^{-30}$) displaying high correlation between replicates ($R = 0.994$). (C-F) Unfused hypPB-directed insertions are highly enriched for the active enhancer marks H3K27ac and H3K4me1 and depleted for suppressive mark H3K27me3. Representative image (C), heatmaps, and enrichment plots (D-F) of H3K27ac, H3K4me1, and H3K27me3 density at 7031 significantly-enriched insertion peaks in two littermate mice. In (C), top track of each insertion replicate displays unique insertions, where each circle = 1 unique insertion and y-axis represents number of reads supporting each insertion on a \log_{10} scale, and bottom track displays normalized local insertion density across the genome (insertions per million per kB). Y-axis of ChIP-seq data represents read depth with smoothing filter applied. Heatmaps and enrichment plots are centered on insertion peaks and extend 10kB in either direction. Relative enrichment quantifications displayed in $\log_2(\text{fold-change over ChIP-seq input})$. (G,H) Percentage of 7031 significantly-enriched insertion peaks with at least 1 basepair (bp) intersection with a H3K27ac-marked enhancer or super enhancer. Gray bar represents intersections after randomizing genomic coordinates of insertion peaks. χ^2 test with Yates correction: $p < 0.0001$. (I,J) Percentage of H3K27ac-marked enhancers and super enhancers with at least 1 bp intersection with a significantly-enriched insertion peak. χ^2 test with Yates correction: $p < 0.0001$.

FLEX calling cards system allows for cell type-specific profiling of REs in the brain.

Gene-based TF-tagging systems such as calling cards offer the potential for cell type-specific profiling through conditional expression of the TF-enzyme fusion. Thus, we generated a Cre-dependent calling cards system, termed FLEX calling cards, and tested the ability of this system to record cell type-specific RE activity or TF binding in complex tissues without isolation of the cell population of interest. In the FLEX system, the reverse complement of the hypPB or TF-hypPB gene is positioned downstream of a strong, ubiquitous promoter and is flanked by two sets of loxP sites. In the presence of Cre, the transposase gene is flipped to the correct orientation and is expressed. To confirm Cre-dependence of the FLEX system, we co-transfected the Cre-dependent hypPB plasmid, hypPB FLEX, into HEK293T cells along with the BrokenHeart reporter plasmid (**Fig 2.5SA**), which expresses TdTomato only after genomic transposon insertion (Qi, Michael Nathaniel Wilkinson, et al. 2017). 24-hours post-transfection, we observed BrokenHeart-derived TdTomato fluorescence only in cells that received both the FLEX calling card constructs and a Cre expression plasmid, demonstrating that this system is Cre-dependent (**Fig 2.5SA**).

As a proof of principle, we focused on two prominent and well-studied Cre-driver mouse lines, Syn1::Cre and GFAP::Cre, which direct expression to neurons and astrocytes, respectively. We packaged the hypPB FLEX plasmid into the AAV9 vector and intracranially co-injected it along with AAV::SRT into P0-1 mouse pups of either the Syn1::Cre or GFAP::Cre genotype, along with Cre(-) littermates as controls. We euthanized the animals at P28, isolated cortical RNA, and sequenced and mapped insertions across the genome (**Fig 2.3A**). Immediately apparent upon sacrifice was that brains of Syn1::Cre positive animals were noticeably more red than their negative littermates, even to the naked eye (**Fig 2.5SB**), a result of the transposition-

enhanced TdTomato reporter expression derived from AAV::SRT. This change in color was striking for Syn1::Cre brains, but not as apparent in GFAP::Cre animals, an observation that is consistent with the relative abundances of transduced neurons and astrocytes (**Fig 2.1C-D**). In Syn1::Cre brains, we analyzed TdTomato expression with immunofluorescence and noted a marked increase in neurons of Cre(+) animals but not Cre(-) littermates (**Fig 2.5SC**). We then sequenced insertions in Cre(+) and Cre(-) littermates from each line and observed a significant increase in insertion events in positive animals as compared to their negative littermates (**Fig 2.5SD**).

We next sought to test whether FLEX calling cards with unfused hypPB could identify cell type-specific REs. To do this, we identified insertion peaks that were differentially enriched in either Syn1::Cre over GFAP::Cre or GFAP::Cre over Syn1::Cre by a count-based statistical comparison (**Fig 2.3B**). We then asked whether genes near these differentially enriched peaks are more likely to be expressed in neurons or astrocytes, using a previously published and widely used cell type-specific RNA-seq dataset (Zhang et al. 2014). As predicted, we found that as our significance threshold for defining differentially enriched insertion peaks became more stringent, the RNA expression of proximal gene sets became more cell type-specific (**Fig 2.5SE-F**). At a stringent significance threshold of $p=10^{-7}$, we compared all nearest genes to Syn1::Cre or GFAP::Cre enriched insertion peaks, and found significant differences in astrocyte versus neuron expression in the expected directionalities (**Fig 2.3B and Fig 2.5SG**). Of note, these neuron and astrocyte RNA enrichments were observed despite using proximity as a means for enhancer-gene pairing, which while widely used for analyses such as these (Ernst et al. 2011), is likely only identifying the correct gene of interest in a subset of pairs (Maurano et al. 2012). Lastly, we inputted these gene sets into an unbiased cell type-identification tool (CSEA (Xu et al. 2014))

and successfully identified cortical astrocyte and neuron populations for genes near GFAP::Cre and Syn1::Cre enriched insertion peaks, respectively (**Fig 2.5SH-I**). Together, these data indicate that peaks derived from FLEX calling cards insertion profiles recorded by unfused hypPB represent cell type-specific REs responsible for driving expression of cell type-enriched genes.

Lastly, we sought to functionally validate the enhancer activity of a subset of the novel astrocyte-enriched REs by testing whether these regions could enhance the expression of a dsRed reporter gene in astrocytes *in vivo*. We chose 4 candidate astrocyte-enriched REs, based on their size, cell type-specific activity, and astrocyte/neuron RNA expression of their nearest genes (**Fig 2.6SA-E**). We then cloned these candidate REs upstream of the hsp68 minimal promoter driving a dsRed reporter gene. As a positive control, we also cloned the canonical GFAP promoter (pGFAP) (Sakers et al. 2017) into the same location upstream of hsp68::dsRed (**Fig 2.3C**). To test the functional enhancer activity of these REs *in vivo*, we delivered these plasmids, along with a separate plasmid carrying a CFP reporter under the GFAP promoter for astrocyte identification, via postnatal astrocyte labeling electroporation (PALE) (Stogsdill et al. 2017). At P7, mice were euthanized and brains were collected for immunohistochemistry. This method specifically targeted astrocytes in the cerebral cortex, noting that >96% of dsRed(+) cells in this region were also GFAP::CFP(+) (**Fig 2.6SF-G**). As expected, the positive control (pGFAP hsp68::dsRed) plasmid exhibited enhanced dsRed fluorescence in astrocytes, relative to a negative control plasmid carrying only hsp68::dsRed, that approached statistical significance ($p=0.055$; **Fig 2.3D-E and Fig 2.6SF,H**). We then quantified enhancer activity of our candidate REs and observed a significant enhancement of dsRed fluorescence for three of the four candidates (**Fig 2.3D-E and Fig 2.6SF,H**). Thus, these astrocyte-enriched REs display functional enhancer activity in astrocytes in the mouse brain at P7. Next, we repeated this experiment

allowing mice to age to P21 to allow further astrocyte maturation prior to euthanasia. To our surprise, at this timepoint we observed a change in localization of dsRed expression in brains receiving the minimal promoter hsp68::dsRed construct, with fewer GFAP::CFP(+) astrocytes and a new population of NeuN(+) neurons labeled with dsRed in the cortex (**Fig 2.7SA-B**). This suggests that the PALE method does deliver plasmids to neurons or neural progenitors in addition to astrocyte progenitors, but that expression via the hsp68 promoter in neurons does not arise until later in postnatal development. Strikingly however, in animals that received pGFAP hsp68::dsRed or any of the RE candidate plasmids, dsRed expression was contained to GFAP::CFP(+) astrocytes, suggesting that in addition to enhancing expression in astrocytes, these RE sequences also repress activity of hsp68 in neurons (**Fig 2.7SA-B**). Additionally, as in the P7 experiment, we observed increased dsRed expression within GFAP::CFP(+) astrocytes for all RE constructs relative to hsp68::dsRed only, of which two reached statistical significance (**Fig 2.7SC**). Indeed, even eMms22l, which was not yet active at P7, displays these functional enhancer phenotypes at P21. Overall, these data demonstrate that cell type-specific REs derived from FLEX calling cards functionally regulate cell type-specific gene expression in the brain.

Fusion of hypPB to the promoter-binding transcription factor SP1 records SP1 occupancy

A key feature of calling cards is the ability to record binding of a TF of interest using TF-hypPB fusions. To demonstrate calling card *in vivo* TF recording with AAV calling cards, we fused hypPB to a sequence-specific DNA-binding general TF, SP1, which binds to gene promoters and is involved in transcription (Beishline and Azizkhan-Clifford 2015; Li et al. 2004), and cloned this fusion gene into the FLEX calling cards system for cell type-specific use (**Fig 2.4A**). As full length SP1 is too large to be efficiently packaged into AAV, we instead used a truncated version of SP1 containing the C-terminal 621 amino acids, which includes the DNA-binding domain and has been shown to be sufficient to replicate sequence-specific binding of full length SP1 (Kadonaga et al. 1988). To test this system, we intracranially co-injected AAV::SP1(621C)-hypPB FLEX along with AAV::SRT into P0-1 mice of the Syn1::Cre line, sacrificed animals at P28, generated and sequenced SRT libraries from cortical RNA samples, and compared insertion profiles to that of unfused hypPB. Consistent with the affinity of SP1 for proximal promoters, we found that insertions were significantly enriched upstream of TSS, as compared to unfused hypPB insertion profiles (**Fig 2.4C**). Next, we defined differentially enriched insertion peaks in SP1(621C)-hypPB profiles over unfused hypPB ($p < 10^{-15}$) and found that the majority of significant enrichments occur in gene promoters (**Fig 2.4D-E**). Finally, at these SP1 peaks, we performed motif discovery and identified the canonical SP1 binding motif, GGGCGGGG (Wang et al. 2012) (**Fig 2.4F**). Thus, fusion of SP1 to hypPB and delivery via AAV identifies SP1 binding sites in the mouse brain.

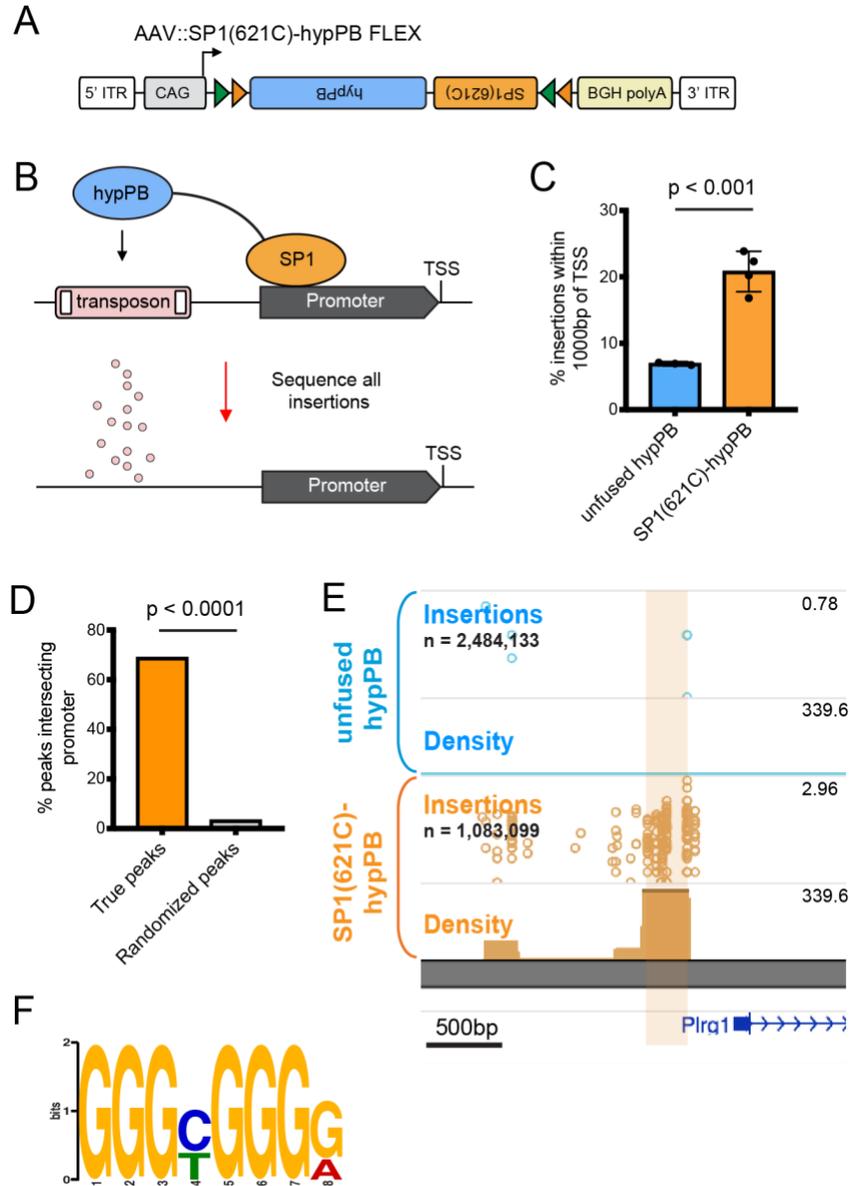


Figure 2.4. Fusion of SP1 to hypPB in AAV calling cards system records SP1 occupancy.

(A) Schematic of AAV::SP1(621C)-hypPB FLEX construct. (B) Fusion of the promoter-binding TF SP1 to hypPB directs insertions to promoter-proximal TTA sites. (C) Percentage of total insertions within 1000 basepairs (bp) of a transcription start site (TSS), displaying increased promoter-directed insertions upon SP1 fusion as compared to unfused hypPB ($n = 3-4$ mice per group; two-tailed, unpaired Student's t-test: $p < 0.001$, $t = 7.66$, $df = 5$, 95% C.I._{diff} [9.22 : 18.54]; unfused hypPB_{mean} = 6.9%, SP1(621C)-hypPB_{mean} = 20.8%). Error bars represent SD. Total SP1(621C)-hypPB insertions: 1,083,099. Total unfused hypPB insertions: 2,484,133. (D) Percentage of significant SP1 insertion peaks differentially enriched over unfused hypPB ($p < 10^{-15}$; 1596 intersecting out of 2316 total peaks) intersecting promoter-proximal regions (1000bp on either side of TSS) compared to randomized peak coordinates (78/2316). χ^2 test with Yates correction: $p < 0.0001$. (E) Representative insertion peak displaying significantly increased insertion density near the TSS of the *Plrg1* gene. (F) Highest information content motif in the sequences flanking the center of significantly enriched SP1(621C)-hypPB insertion peaks ($p < 10^{-15}$) is the canonical SP1 binding motif (GGGCGGGG; $p < 10^{-42}$).

FLEX calling cards provides historical TF binding information through longitudinal TF recording

An intriguing potential use of calling card technologies is in the recording of TF binding over an integrated period of time. Such a method, which is not possible with endpoint TF profiling methods such as ChIP-seq or CUT&Tag, could empower novel studies in which historical TF binding information would be useful, such as during cellular development or differentiation. Further, by integrating signal over time, longitudinal calling cards may report transient binding events which would be otherwise missed with endpoint-only measures.

To test whether FLEX calling cards could report integrated, historical TF occupancy, we asked whether we could recover transient SP1 promoter binding events and successfully read them out at a later date. Importantly, consistent with the known role of SP1 in regulating gene expression (Beishline and Azizkhan-Clifford 2015; Li et al. 2004; Wang et al. 2012), we observed that expression of genes genome-wide was on average correlated with the number of SP1-directed promoter insertions (**Fig 2.5A-B**). Thus, we predicted that should a gene be expressed early, but not late, in the lifetime of the animal, this transient event could be marked by SP1 binding and be recoverable via SP1 calling cards at a later timepoint.

To test this hypothesis, we intracranially co-injected AAV::SP1(621C)-hypPB and AAV::SRT into two separate cohorts of P0-1 mice. The first cohort was euthanized at P10, while the second cohort was allowed to continue to record SP1 occupancy until P28 (**Fig 2.5C**). This time period of postnatal brain development involves several key neurodevelopmental processes (Semple et al. 2013), including substantial hippocampal neurogenesis (Khalaf-Nazzal and Francis 2013) as well as glial and synaptic maturation (Semple et al. 2013), development of the extracellular matrix (Semple et al. 2013), and closing of critical periods (Semple et al. 2013),

which are accompanied by numerous changes in gene and protein expression (Laeremans et al. 2013). For these analyses, we utilized a previously-published cortical RNA-seq dataset (Lister et al. 2013) with postnatal timepoints of 1 week (Wk1; ~P7) and 4 weeks (Wk4; ~P28). From these expression data, we then derived and tested three separate predictions (**Fig 2.5C**). First, for genes expressing at Wk1 but not Wk4, (i.e. “early genes”) we would observe near-equivalent SP1 binding at promoters in both cohorts. Second, for constitutive genes that express equally at Wk1 and Wk4, we would observe continued integration of SP1 binding in the P28 cohort, resulting in increased SP1 insertion density at promoters. And third, for genes expressing only at Wk4 (i.e. “late genes”), we would observe SP1 promoter binding only in the P28 cohort. We defined early genes as having a $\log(\text{Wk4/Wk1 FPKM})$ of less than -0.5 ($n = 292$), late genes as greater than 0.5 ($n = 285$), and all genes in the middle as constitutive ($n = 4413$; **Fig 2.5D-E**). Indeed, at the promoters of these gene sets, we observed SP1 promoter occupancy to be consistent with our three predictions (**Fig 2.5E**; example early and late genes, *Idh1* and *Gjb6*, displayed in **Fig 2.5F** and **Fig 2.5G**, respectively). Importantly, at the promoters of early genes, we observed near-equivalent binding in the P10 and P28 cohorts (mean P28:P10 SP1 promoter insertion ratio = 0.98), despite these genes only being transiently expressed; thus, this system is capable of permanently recording transient TF binding events for retrospective read out at a later date. Together, these data support that TF-hypPB fusions integrate signal over time and provide a historical, integrated picture of TF occupancy.

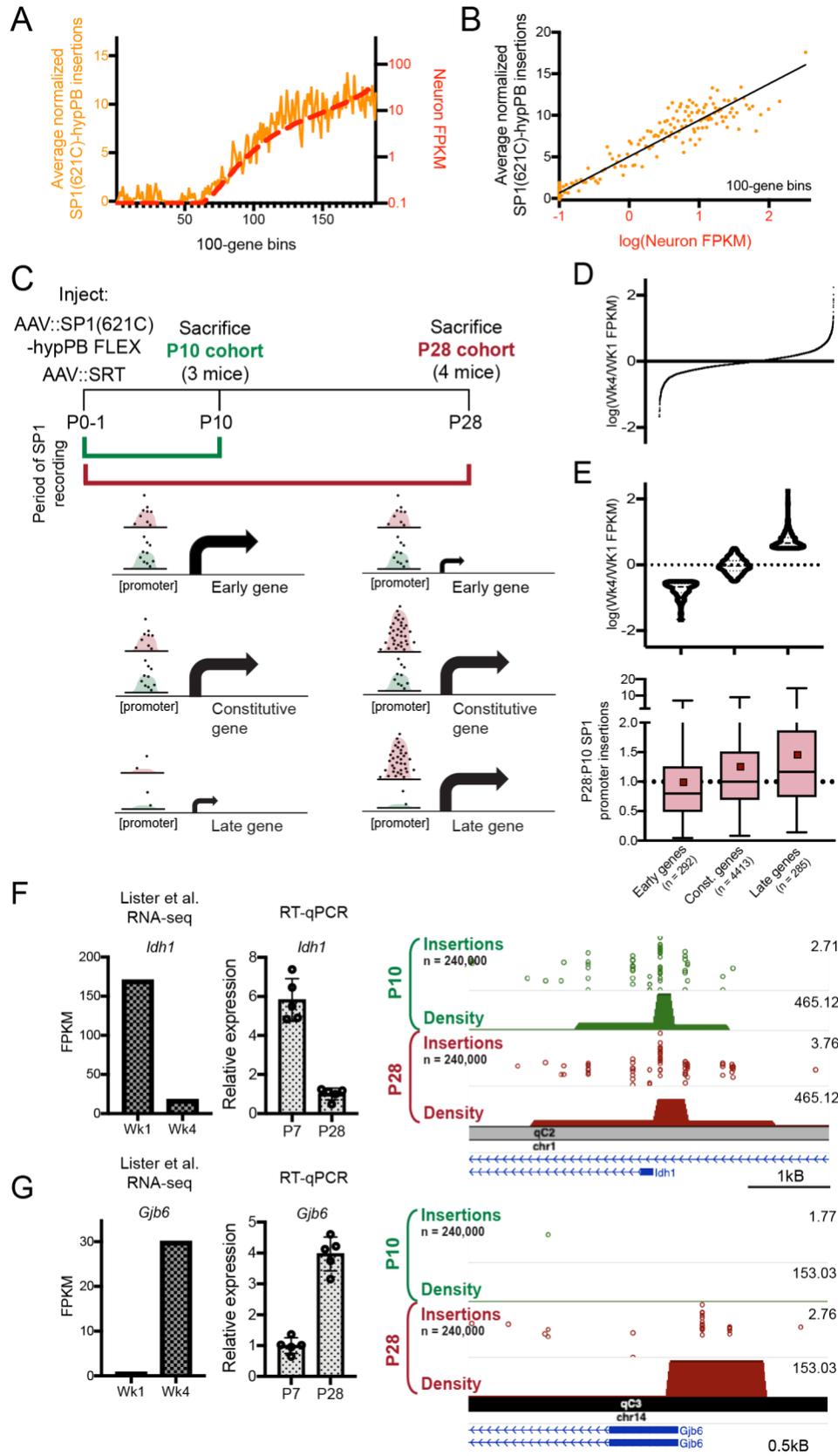


Figure 2.5. Longitudinal SP1 profiling reports integrated record of SP1 binding.

(A-B) Normalized number of SP1(621C)-hypPB directed insertions at promoter proximal regions after subtraction of unfused hypPB insertions, versus neuron-specific gene expression for all genes, binned and averaged into 100-gene bins. In (A), left y-axis represents number of promoter insertions normalized to 10⁶ total insertions in the sample and right y-axis displays neuron-specific RNA expression from Zhang et al., *Journal of Neuroscience*, 2014. (B) displays strong correlation of SP1(621C)-hypPB promoter insertions with gene expression after subtraction of unfused hypPB insertions ($R=0.96$, $p<0.0001$). (C) Experimental paradigm and predicted temporal SP1 occupancy for early, constitutive, and late expressing genes. (D) Distribution of Wk4/Wk1 expression ratios for all expressed and SP1-bound genes. (E) (top) Categorization of genes into “early” ($\log(\text{Wk4/Wk1 FPKM}) < -0.5$), “constitutive” ($-0.5 < \log(\text{Wk4/Wk1 FPKM}) < 0.5$), and “late” ($\log(\text{Wk4/Wk1 FPKM}) > 0.5$) gene sets. RNA-seq from Lister et al., *Science*, 2013. (bottom) SP1-derived promoter insertions for early, constitutive, and late gene sets, demonstrating efficient capture of transient SP1 binding events at early gene promoters and continued integration of constitutive and late gene promoters in the P28 cohort relative to the P10 cohort (One-way ANOVA [$F(2, 4987) = 16.92$], $p<0.0001$). Early genes_{mean} = 0.98, const. genes_{mean} = 1.25, late genes_{mean} = 1.45. Red square = mean, line = median. (F-G) Example of an (F) early expressed gene (*Idh1*) displaying equivalent SP1 binding in both cohorts and a (G) late expressed gene (*Gjb6*) displaying SP1 binding only in the P28 cohort. Left bar graph displays reference RNA-seq read counts from Lister et al., *Science*, 2013, while right bar graph displays our own confirmatory RT-qPCR from brains of C57bl6/J mice at P7 and P28.

FLEX splicing variant, “AAV::hypPB Frontflip,” reduces Cre-independent transposition

During our analysis of FLEX calling cards in neurons and astrocytes we observed the presence of some background transposition activity in Cre(-) animals (**Fig 2.5SD**), which could impinge on our ability to apply FLEX calling cards to analyze more rare cell populations. We hypothesized that this aberrant hypPB expression could be due to cryptic transcription of the hypPB gene on the antisense strand of the AAV::hypPB FLEX sequence, which would result in expression independent of Cre-mediated flip-excision. To overcome this, we designed a variant FLEX construct in which only the N-terminus of hypPB is present within the FLEX cassette and is flanked by the 5' end of a synthetic intron(Younis et al. 2010), the 3' end of which is located upstream of the hypPB C-terminus. Thus, aberrant transcription in either direction would produce only a non-functional, truncated hypPB protein. However, upon Cre recombination, the N-terminal hypPB segment is flipped in frame with its C-terminus, split by the newly-reconstituted artificial intron which will then be spliced out of the nascent RNA, producing a full length, functioning hypPB. We coined this FLEX variant “hypPB Frontflip” (**Fig 2.6A**).

We first tested the ability of hypPB Frontflip to reduce Cre(-) background *in vitro* by co-transfecting it into HEK293T cells along with the BrokenHeart reporter transposon. Previously, we had shown that the original hypPB FLEX construct produces no Cre-independent transposition *in vitro* at 24-hours post-transfection (**Fig 2.5SA**). However, at 96-hours post-transfection, we began to observe some background transposition in the absence of Cre, mimicking our *in vivo* results. In contrast, the hypPB Frontflip variant eliminated this Cre-independent transposition *in vitro* (**Fig 2.6B**). We then packaged hypPB Frontflip into an AAV9 vector and intracranially co-injected it with AAV::SRT into GFAP::Cre animals at P0-1. At P28, we observed a significant and dramatic reduction in Cre-independent insertions in Cre(-) animals

accompanied by a >30-fold increase in Cre(+) littermates (**Fig 2.6C**). To confirm that these insertions were indeed astrocyte-derived, as we had previously done with the original AAV::hypPB FLEX, we identified differential BRD4 binding peaks compared to the Syn1:Cre line and analyzed the cell type expression patterns of proximal genes. Indeed, we found that genes near differential binding peaks again exhibited increased expression in astrocytes, as expected (**Fig 2.6D**). In summary, AAV::hypPB Frontflip is a tightly Cre-controlled conditional expression viral system, which will allow for the implementation of AAV calling cards in more rare cell types in the future.

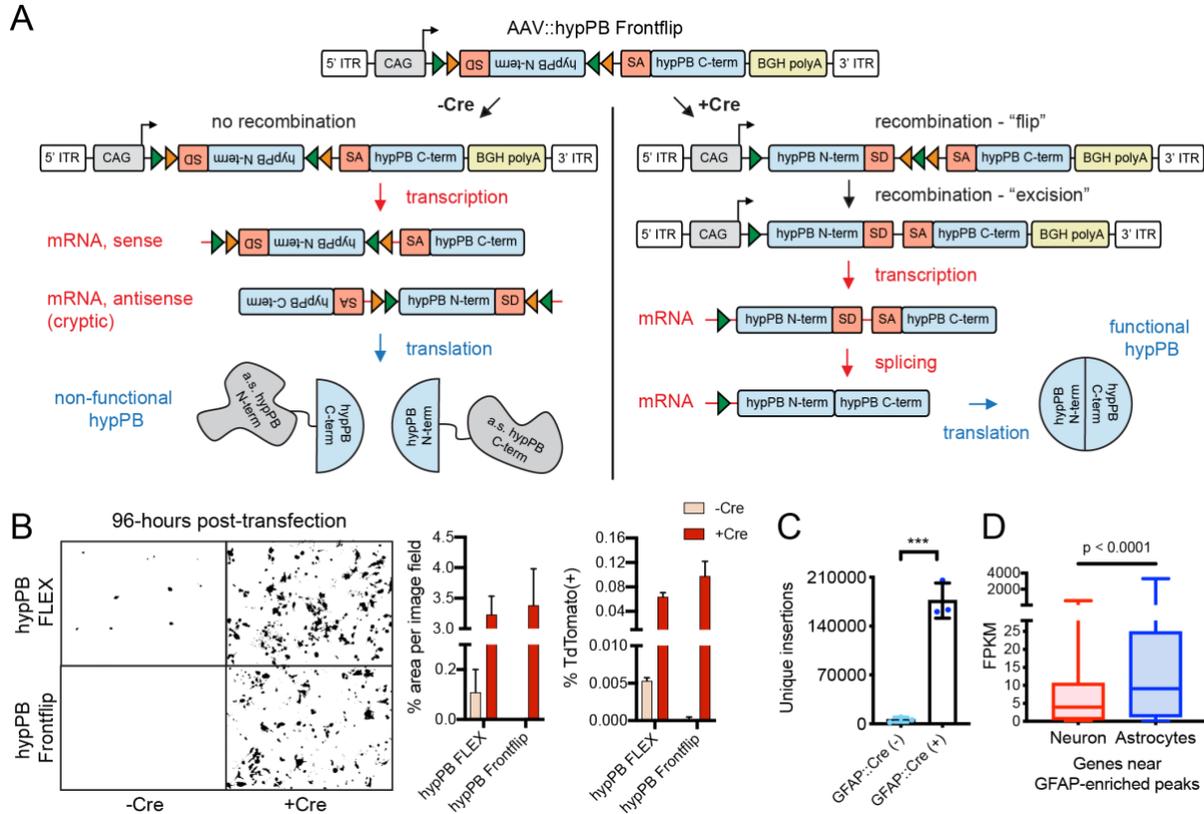


Figure 2.6. AAV::hypPB Frontflip reduces Cre-independent transposition *in vitro* and *in vivo*

(A) Schematic of AAV::hypPB Frontflip. Prior to Cre-mediated recombination, hypPB is split into its N- and C-termini, with the N-terminus in reverse orientation and inside a FLEX cassette. On the 3' end of the N-terminus is the splice donor (SD) and 5' end of an artificial intron. On the 5' end of the C-terminus is the splice acceptor (SA) and 3' end of the intron. Upon recombination, the N-terminal fragment is flipped into frame and the artificial intron is reconstituted. This sequence is then transcribed into mRNA, spliced, and translated into an uninterrupted, functional hypPB protein. (B) hypPB FLEX or hypPB Frontflip plasmids were co-transfected into HEK293T cells along with the BrokenHeart fluorescent transposition reporter. At 96-hours post-transfection, Cre(-) background is observed from hypPB FLEX, but not hypPB Frontflip. For quantification of fluorescent imaging (left graph): n=2 wells per condition, 3 images per well. For flow cytometry (right graph): n=2 wells per condition. (C) AAV::hypPB Frontflip significantly reduced transposition in Cre(-) animals (<10,000 insertions per brain), with a >30-fold increase in insertion total in Cre(+) animals (two-tailed, unpaired Student's t-test: *** p<0.001). GFAP::Cre (+) vs (-): p=0.00031, t=11.63, df=4, 95% C.I._{diff} [129,913 : 211,436]. (D) Quantifications of neuron and astrocyte specific expression of genes near GFAP::Cre [Neuron_{median} = 3.96 FPKM, Astrocyte_{median} = 9.10 FPKM, n = 615 genes] enriched insertion peaks, called via AAV::hypPB Frontflip data (p<10⁻⁷), showing significant preferential expression in astrocytes (two-tailed Mann-Whitney U test: p<0.0001).

Discussion

In this report, we have successfully developed an *in vivo*, virally mediated calling cards approach, for which we have demonstrated utility for TF and RE profiling in the live mouse brain. This technology builds on previously developed *in vitro* calling card methodologies, adapting this method for investigation of epigenetic regulation in the mammalian brain. Further, in proof-of-principle experiments, we demonstrated effectiveness of this protocol for 1) profiling TFs without an antibody, 2) cell type-specific RE profiling without cellular population purification, and 3) integrative recording of TF binding events over time. Our use of SRTs in this paradigm now also enable calling cards for single cell analyses (Moudgil et al. 2019), which expands and highlights the versatility of this toolkit in future studies.

Calling cards technologies, as gene-based systems, have several unique features that provide advantages over biochemical TF profiling methods such as ChIP-seq for certain applications (Qi, Wilkinson, et al. 2017; Wang et al. 2011, 2012). One such property of the FLEX version of the AAV calling card methodology is that there is no requirement for physical isolation of cell types or nuclei for cell type-specific analyses. This allows for the same protocol to be used for any cell type of interest, the identify of which is determined by the Cre-driver mouse line used, and avoids potential disassociation-related artifacts (van den Brink et al. 2017). Secondly, while not explored here, one could envision simple manipulations of the AAV calling cards system to allow for temporal control of the system (Zhang et al. 2014). Such adaptations could allow for innovative studies in which TF binding is recorded only during defined windows of time (Mitchell et al. 2016). In a similar vein, we have demonstrated here the ability of AAV calling cards to integrate TF binding information over time, which will allow for retrospective analysis of historical TF activity in cells. By applying this unique utility to SP1, we identified

promoter regions of genes with accumulating SP1 binding across postnatal development and captured transient SP1 binding events at early-expressed genes. Finally, AAV calling cards does not require a TF-specific antibody, allowing for TF profiling for, in theory, any packagable TF, simply by fusing it to hypPB. This being a virally-mediated system allows for simple and rapid application to animal models without the need for expensive and time-consuming breeding. Intracranial injection for a standard size litter of mice can be completed in under an hour. Finally, the non-Cre dependent versions of the system should be equally applicable in other species of interest such as rats and primates. Reagents, cloning strategies, and user-friendly analysis pipelines are available upon request, making AAV calling cards readily available for neuroscience research.

Of course, there are caveats to be considered as well. Most notably, there is potential for induced mutation, given the tendency for transposons to insert into or near critical gene regulatory regions. Indeed, transposon technologies are often used in mutagenesis screens in which transposon-mediated gene disruption can be deleterious (Rad et al. 2014). However, in such studies, the transposons are specifically engineered with splice-site gene or enhancer traps, while the SRT used in AAV calling cards only drives expression of a reporter gene and the genomic sequence immediately downstream of its insertion site. Further, the transposition rate of the *piggyBac* transposase is inherently low (<20-100 per cell (Kettlun et al. 2011; Yusa et al. 2011)), suggesting that it is highly unlikely for insertions to disrupt regulatory regions on both alleles in the same cell. Consistent with this, we observed no excess degeneration or behavioral/developmental deficits in AAV calling card-injected animals beyond that induced by needle injury, and in general, calling card technologies have not exhibited marked deleterious effects in previous reports (Wang et al. 2011, 2012).

The relatively low “per cell” transposition rate of hypPB also helps to alleviate concerns of reduced insertion efficiency at specific loci due to saturation after long periods of calling cards recording. Indeed, with millions of available TTAA sites and thousands of active REs across the genome, the chances of two insertions falling in the same site in the same cell are on the order of less than 10^{-6} to 10^{-12} . Nevertheless, it remains possible that a small subset of calling cards transposition events could perturb RE activity and nascent gene expression, which could, in theory, alter the ability for hypPB to insert a second transposon into a nearby site. Further, while we have demonstrated here the ability to record TF binding peaks across 28 days of FLEX calling card expression, it should be noted that calling cards recording may be limited in some cases by transposon availability, which could become exhausted after long periods of recording. Our final experiment with SP1 at P10 and P28 demonstrates the acquisition of new peaks at late gene promoters, despite early genes using many of the available transposons, suggesting that such exhaustion effects are limited here.

Next, while useful for profiling enhancers and super enhancers more broadly, the natural affinity of hypPB for BRD4 does necessitate that any experiment using a TF-hypPB be accompanied by a control with unfused hypPB only, such that TF binding peaks can be identified with differential peak calling, as was done for SP1 in this report. Future versions of AAV calling cards systems may be improved by fusion of TFs to other, non-BRD4-biased transposases (Yoshida et al. 2017), though the efficiency of transposon re-direction would need to be tested empirically in each case.

Finally, while we do see a clear Cre-induction of AAV::hypPB FLEX, providing a proof of principle for its use in cell type-specific profiling, we also observed some background insertion events in the absence of Cre, which could be limiting for profiling of rare cell types in

which signal is likely to be reduced. To directly address this, we designed and tested a variant FLEX virus, AAV::hypPB Frontflip, and in doing so, substantially reduced the Cre(-) background observed with the original FLEX construct both *in vitro* and *in vivo*. This major improvement will now allow for the application of FLEX calling cards to more rare cell types, where insertion number is likely to be diminished. Of note, “leaky” expression of sensitive enzymes in FLEX cassettes in AAV vectors is a common issue in neuroscience, with a number of other strategies also having been proposed to reduce Cre(-) background (e.g. start codon/kozak outside of FLEX cassette (Wall et al. 2010), mutated loxp sites (Fischer, Collins, and Callaway 2019), AAV titration (Lavin et al. 2019)), which may also be useful in reducing FLEX calling cards background. Alternatively, simply changing the viral serotype (Hammond et al. 2017) or promoter (Graybuck et al. 2019; Jüttner et al. 2019) could allow for similar analyses in cell types not explored here without the need for Cre-dependent conditional expression.

In summary, we have introduced AAV calling cards as a viable method for recording TF binding and active REs *in vivo* and demonstrated its effectiveness in profiling cell type-specific and historical TF and RE activity in the brain. Future applications of this technology to animal models of development and disease could unlock important insights into epigenetic gene regulation in a variety of neuroscience disciplines.

Acknowledgments

This work was supported by U01MH10913301 (to JDD and RDM), RF01MH117070-01 (to JDD and RDM), R21HG009750 (to RDM), and the Hope Center Viral Vectors Core at Washington University School of Medicine. AJC was supported by T32GM008151-32n and the Lucille P. Markey Special Emphasis Pathway in Human Pathobiology. AM was supported by T32GM007200, T32HG000045, and F30HG009986. JC was supported by a McDonnell Scholarship and the Lucille P. Markey Special Emphasis Pathway in Human Pathobiology. MJV was supported by F32NS105363-02. TL was supported by T32GM007067. We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for genomic analysis. The Center is partially supported by NCI Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center and by ICTS/CTSA Grant# UL1 TR000448 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. This publication is solely the responsibility of the authors and does not necessarily represent the official view of NCRR or NIH. We thank the Edison Family Center for Genome Sciences & Systems Biology, specifically Jessica Hoisington-Lopez and MariaLynn Crosby, for assistance with genomic analysis. We thank Bernard Mulvey for providing hsp68::dsRed plasmids. Finally, we thank the Kevin Noguchi, Brant Swiney, and the Intellectual and Developmental Disabilities Research Center (IDDRC) at Washington University (NICHD U54-HD087011) for neuropathological assessments.

Author contributions

Study designed by AJC, AM, JC, MJV, AY, TL, SEM, RDM, and JDD. AJC, AM, JC, JH, XC, MNW, and MH designed and generated DNA constructs. AM developed SRT analysis pipelines. AJC, JC, MJV, MS, KM, AY, TL, SEM, XC, RDM, and JDD generated and analyzed data. Manuscript written by AJC and edited by AM, JC, MJV, AY, TL, TMM, RDM, and JDD.

Declaration of interests

RDM, AM, and MNW have filed a patent application on SRT technology. No other authors have disclosures to report.

Bibliography

- van den Aamele, Jelle, Robert Krautz, and Andrea H. Brand. 2019. “TaDa! Analysing Cell Type-Specific Chromatin in Vivo with Targeted DamID.” *Current Opinion in Neurobiology* 56: 160–66. <https://doi.org/10.1016/j.conb.2019.01.021>.
- Bailey, Timothy L. et al. 2009. “MEME Suite: Tools for Motif Discovery and Searching.” *Nucleic Acids Research* 37(SUPPL. 2): 202–8.
- Beishline, Kate, and Jane Azizkhan-Clifford. 2015. “Sp1 and the ‘Hallmarks of Cancer.’” *FEBS Journal* 282(2): 224–58.
- Bonn, Stefan et al. 2012. “Cell Type-Specific Chromatin Immunoprecipitation from Multicellular Complex Samples Using BiTS-Chip.” *Nature Protocols* 7(5): 978–94. <http://dx.doi.org/10.1038/nprot.2012.049>.
- van den Brink, Susanne C et al. 2017. “Single-Cell Sequencing Reveals Dissociation-Induced Gene Expression in Tissue Subpopulations.” *Nature Methods* 14(10): 935–36.
- Cheetham, Seth W. et al. 2018. “Targeted DamID Reveals Differential Binding of Mammalian Pluripotency Factors.” *Development* (September 5): dev.170209. <http://dev.biologists.org/lookup/doi/10.1242/dev.170209>.
- Consortium, The ENCODE Project. 2012. “An Integrated Encyclopedia of DNA Elements in the Human Genome.” *Nature* 489(7414): 57–74.
- Cooney, Ashley L., Brajesh K. Singh, and Patrick L. Sinn. 2015. “Hybrid Nonviral/Viral Vector Systems for Improved PiggyBac DNA Transposon in Vivo Delivery.” *Molecular Therapy* 23(4): 667–74.
- Corradin, Olivia, and Peter C. Scacheri. 2014. “Enhancer Variants: Evaluating Functions in Common Disease.” *Genome Medicine* 6(10): 1–14.

- Creyghton, M. P. et al. 2010. “Histone H3K27ac Separates Active from Poised Enhancers and Predicts Developmental State.” *Proceedings of the National Academy of Sciences* 107(50): 21931–36. <http://www.pnas.org/cgi/doi/10.1073/pnas.1016071107>.
- Deal, Roger B., and Steven Henikoff. 2010. “A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue.” *Developmental Cell* 18(June 15): 1030–40. <http://dx.doi.org/10.1016/j.devcel.2010.05.013>.
- DeVos, Sarah L, and Timothy M Miller. 2013. “Direct Intraventricular Delivery of Drugs to the Rodent Central Nervous System.” *JoVE* (75): e50326. <https://www.jove.com/video/50326>.
- Ernst, Jason et al. 2011. “Mapping and Analysis of Chromatin State Dynamics in Nine Human Cell Types.” *Nature* 473(7345): 43–49. <http://dx.doi.org/10.1038/nature09906>.
- Fischer, Kyle B., Hannah K. Collins, and Edward M. Callaway. 2019. “Sources of Off-Target Expression from Recombinase-dependent AAV Vectors and Mitigation with Cross-over Insensitive ATG-out Vectors.” *Proceedings of the National Academy of Sciences of the United States of America* 116(52): 27001–10.
- Girdhar, Kiran et al. 2018. “Cell-Specific Histone Modification Maps in the Human Frontal Lobe Link Schizophrenia Risk to the Neuronal Epigenome.” *Nature Neuroscience* 21(8): 1126–36. <http://dx.doi.org/10.1038/s41593-018-0187-0>.
- Gogol-Doring, Andreas et al. 2016. “Genome-Wide Profiling Reveals Remarkable Parallels between Insertion Site Selection Properties of the MLV Retrovirus and the PiggyBac Transposon in Primary Human CD4+ T Cells.” *Molecular Therapy* 24(3): 592–606. <http://dx.doi.org/10.1038/mt.2016.11>.
- Grady, R. Mark, David F. Wozniak, Kevin K. Ohlemiller, and Joshua R. Sanes. 2006. “Cerebellar Synaptic Defects and Abnormal Motor Behavior in Mice Lacking α - and Beta-

- Dystrobrevin.” *Journal of Neuroscience* 26(11): 2841–51.
- Graybuck, Lucas T et al. 2019. “Prospective, Brain-Wide Labeling of Neuronal Subclasses with Enhancer-Driven AAVs.” *bioRxiv [Preprint]*: doi.org/10.1101/525014.
<https://www.biorxiv.org/content/early/2019/01/20/525014>.
- Hammond, Sean L., Ashley N. Leek, Evan H. Richman, and Ronald B. Tjalkens. 2017. “Cellular Selectivity of AAV Serotypes for Gene Delivery in Neurons and Astrocytes by Neonatal Intracerebroventricular Injection.” *PLoS ONE* 12(12): 1–22.
- Heintzman, Nathaniel D. et al. 2009. “Histone Modifications at Human Enhancers Reflect Global Cell-Type-Specific Gene Expression.” *Nature* 459(7243): 108–12.
<http://dx.doi.org/10.1038/nature07829>.
- Heinz, Sven et al. 2010. “Simple Combinations of Lineage-Determining Transcription Factors Prime Cis-Regulatory Elements Required for Macrophage and B Cell Identities.” *Molecular Cell* 38(4): 576–89. <http://dx.doi.org/10.1016/j.molcel.2010.05.004>.
- Hnisz, Denes et al. 2013. “Super-Enhancers in the Control of Cell Identity and Disease.” *Cell* 155(4): 934–47. <http://linkinghub.elsevier.com/retrieve/pii/S0092867413012270>.
- Jung, Marie et al. 2014. “Affinity Map of Bromodomain Protein 4 (BRD4) Interactions with the Histone H4 Tail and the Small Molecule Inhibitor JQ1.” *Journal of Biological Chemistry* 289(13): 9304–19.
- Jüttner, Josephine et al. 2019. “Targeting Neuronal and Glial Cell Types with Synthetic Promoter AAVs in Mice, Non-Human Primates and Humans.” *Nature Neuroscience*.
- Kadonaga, James T, Albert J Courey, Joseph Ladika, and Robert Tjian. 1988. “Distinct Regions of Sp1 Modulate DNA Binding and Transcriptional Activation.” *Science* 242(4885): 1566–70.

- Kanno, Tomohiko et al. 2004. "Selective Recognition of Acetylated Histones by Bromodomain Proteins Visualized in Living Cells." *Molecular Cell* 13(1): 33–43.
- . 2014. "BRD4 Assists Elongation of Both Coding and Enhancer RNAs by Interacting with Acetylated Histones." *Nature Structural and Molecular Biology* 21(12): 1047–57. <http://dx.doi.org/10.1038/nsmb.2912>.
- Kettlun, Claudia et al. 2011. "Manipulating Piggybac Transposon Chromosomal Integration Site Selection in Human Cells." *Molecular Therapy* 19(9): 1636–44.
- Khalaf-Nazzal, R., and F. Francis. 2013. "Hippocampal Development - Old and New Findings." *Neuroscience* 248: 225–42. <http://dx.doi.org/10.1016/j.neuroscience.2013.05.061>.
- Kwon, Deborah Y., Ying Tao Zhao, Janine M. Lamonica, and Zhaolan Zhou. 2017. "Locus-Specific Histone Deacetylation Using a Synthetic CRISPR-Cas9-Based HDAC." *Nature Communications* 8(May): 1–8. <http://dx.doi.org/10.1038/ncomms15315>.
- Laeremans, Annelies et al. 2013. "Protein Expression Dynamics During Postnatal Mouse Brain Development." *Journal of Experimental Neuroscience* 7: 61–74.
- Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. 2009. "Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome." *Genome Biology* 10(3).
- Lavin, Thomas K., Lei Jin, Nicholas E. Lea, and Ian R. Wickersham. 2019. "Monosynaptic Tracing Success Depends Critically on Helper Virus Concentrations." *bioRxiv [Preprint]* (doi: <http://dx.doi.org/10.1101/736017>).
- LeRoy, Gary, Brenden Rickards, and S. J. Flint. 2008. "The Double Bromodomain Proteins Brd2 and Brd3 Couple Histone Acetylation to Transcription." *Molecular Cell* 30(1): 51–60.
- Li, Lin, Shihua He, Jian-Min Sun, and James R Davie. 2004. "Gene Regulation by Sp1 and

- Sp3.” *Biochemistry and Cell Biology* 82(4): 460–71.
<http://www.nrcresearchpress.com/doi/abs/10.1139/o04-045>.
- Lister, Ryan et al. 2013. “Global Epigenomic Reconfiguration During Mammalian Brain Development.” *Science* 341(August 9): 1237905-1:12.
- Machanick, Philip, and Timothy L. Bailey. 2011. “MEME-ChIP: Motif Analysis of Large DNA Datasets.” *Bioinformatics* 27(12): 1696–97.
- Maurano, Matthew T. et al. 2012. “Systematic Localization of Common Disease-Associate Variation in Regulatory DNA.” *Science* 337(September): 1190–95.
- Mitchell, Amanda C. et al. 2016. “Longitudinal Assessment of Neuronal 3D Genomes in Mouse Prefrontal Cortex.” *Nature Communications* 7: 1–10.
<http://dx.doi.org/10.1038/ncomms12743>.
- Mo, Alisa et al. 2015. “Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain.” *Neuron* 86(6): 1369–84. <http://dx.doi.org/10.1016/j.neuron.2015.05.018>.
- Mochizuki, Kazuki et al. 2008. “The Bromodomain Protein Brd4 Stimulates G1 Gene Transcription and Promotes Progression to S Phase.” *Journal of Biological Chemistry* 283(14): 9040–48.
- Moudgil, Arnav et al. 2019. “Self-Reporting Transposons Enable Simultaneous Readout of Gene Expression and Transcription Factor Binding in Single Cells.” *BioRxiv [Preprint]* (<https://doi.org/10.1101/538553>).
- Noguchi, Kevin K., Brian Nemmers, and Nuri B. Farber. 2005. “Age Has a Similar Influence on the Susceptibility to NMDA Antagonist-Induced Neurodegeneration in Most Brain Regions.” *Developmental Brain Research* 158(1–2): 82–91.
- Qi, Zongtai et al. 2017a. “An Optimized, Broadly Applicable *PiggyBac* Transposon Induction

- System.” *Nucleic Acids Research*: gkw1290. <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw1290>.
- . 2017b. “An Optimized, Broadly Applicable PiggyBac Transposon Induction System.” *Nucleic Acids Research* 45(7): 1–13.
- Quinlan, Aaron R., and Ira M. Hall. 2010. “BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features.” *Bioinformatics* 26(6): 841–42.
- Rad, Roland et al. 2014. “A Conditional PiggyBac Transposition System for Genetic Screening in Mice Identifies Oncogenic Networks in Pancreatic Cancer.” *Nature Genetics* 47(1): 47–56. <http://www.nature.com/doi/10.1038/ng.3164>.
- Sakers, Kristina et al. 2017. “Astrocytes Locally Translate Transcripts in Their Peripheral Processes.” *Proceedings of the National Academy of Sciences* 114(19): E3830–38. <http://www.pnas.org/lookup/doi/10.1073/pnas.1617782114>.
- Schmid, Manfred, Thérèse Durussel, and Ulrich K. Laemmli. 2004. “ChIC and ChEC: Genomic Mapping of Chromatin Proteins.” *Molecular Cell* 16(1): 147–57.
- Semple, Bridgette D. et al. 2013. “Brain Development in Rodents and Humans: Identifying Benchmarks of Maturation and Vulnerability to Injury across Species.” *Progress in Neurobiology* 106–107: 1–16.
- Skene, Peter J., Jorja G. Henikoff, and Steven Henikoff. 2018. “Targeted in Situ Genome-Wide Profiling with High Efficiency for Low Cell Numbers.” *Nature Protocols* 13(5): 1006–19. <http://dx.doi.org/10.1038/nprot.2018.015>.
- Smith, Ryan P., Jesse D. Riordan, Charlotte R. Feddersen, and Adam J. Dupuy. 2015. “A Hybrid Adenoviral Vector System Achieves Efficient Long-Term Gene Expression in the Liver via PiggyBac Transposition.” *Human Gene Therapy* 26(6): 377–85.

- <http://online.liebertpub.com/doi/10.1089/hum.2014.123>.
- Southall, Tony D. et al. 2013. “Cell-Type-Specific Profiling of Gene Expression and Chromatin Binding without Cell Isolation: Assaying RNA Pol II Occupancy in Neural Stem Cells.” *Developmental Cell* 26(1): 101–12. <http://dx.doi.org/10.1016/j.devcel.2013.05.020>.
- Spitz, François, and Eileen E M Furlong. 2012. “Transcription Factors: From Enhancer Binding to Developmental Control.” *Nature Reviews Genetics* 13(9): 613–26. <http://dx.doi.org/10.1038/nrg3207>.
- van Steensel, Bas, and Steven Henikoff. 2000. “Identification of in Vivo DNA Targets of Chromatin Proteins Using Tethered Dam Methyltransferase.” *Nature Biotechnology* 18(April).
- Stogsdill, Jeff A. et al. 2017. “Astrocytic Neuroligins Control Astrocyte Morphogenesis and Synaptogenesis.” *Nature* 551(7679): 192–97. <http://dx.doi.org/10.1038/nature24638>.
- Stroud, Hume et al. 2017. “Early-Life Gene Expression in Neurons Modulates Lasting Epigenetic States.” *Cell* 171(5): 1151–1164.e16. <https://doi.org/10.1016/j.cell.2017.09.047>.
- Tosti, Luca et al. 2018. “Mapping Transcription Factor Occupancy Using Minimal Numbers of Cells in Vitro and in Vivo.” *Genome Research* 28(4): 592–605.
- Visel, Axel et al. 2009. “ChIP-Seq Accurately Predicts Tissue-Specific Activity of Enhancers.” *Nature* 457(7231): 854–58. <http://dx.doi.org/10.1038/nature07730>.
- Vogel, Maartje J. et al. 2006. “Human Heterochromatin Proteins Form Large Domains Containing KRAB-ZNF Genes.” *Genome Research* 16(12): 1493–1504.
- Wall, Nicholas R. et al. 2010. “Monosynaptic Circuit Tracing in Vivo through Cre-Dependent Targeting and Complementation of Modified Rabies Virus.” *Proceedings of the National Academy of Sciences of the United States of America* 107(50): 21848–53.

- Wang, Haoyi et al. 2008. “‘Calling Cards’ Method for High-Throughput Identification of Targets of Yeast DNA-Binding Proteins.” *Nature Protocols* 3(10): 1569–77.
- . 2011. “‘Calling Cards’ Enable Multiplexed Identification of the Genomic Targets of DNA-Binding Proteins.” *Genome Research* 21(5): 748–55.
- . 2012. “‘Calling Cards’ for DNA-Binding Proteins in Mammalian Cells.” *Genetics* 190(3): 941–49.
- Wells, Alan et al. 2015. “The Anatomical Distribution of Genetic Associations.” *Nucleic Acids Research* 43(22): 10804–20.
- Whyte, Warren A. et al. 2013. “Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes.” *Cell* 153(2): 307–19.
<http://linkinghub.elsevier.com/retrieve/pii/S0092867413003929>.
- Wozniak, David F. et al. 2004. “Apoptotic Neurodegeneration Induced by Ethanol in Neonatal Mice Is Associated with Profound Learning/Memory Deficits in Juveniles Followed by Progressive Functional Recovery in Adults.” *Neurobiology of Disease* 17(3): 403–14.
- Xu, X. et al. 2014. “Cell Type-Specific Expression Analysis to Identify Putative Cellular Mechanisms for Neurogenetic Disorders.” *Journal of Neuroscience* 34(4): 1420–31.
- Yoshida, Junko et al. 2017. “Chromatin States Shape Insertion Profiles of the PiggyBac, Tol2 and Sleeping Beauty Transposons and Murine Leukemia Virus.” *Scientific Reports* 7(May 2016): 1–18. <http://dx.doi.org/10.1038/srep43613>.
- Younis, I. et al. 2010. “Rapid-Response Splicing Reporter Screens Identify Differential Regulators of Constitutive and Alternative Splicing.” *Molecular and Cellular Biology* 30(7): 1718–28.
- Yusa, Kosuke et al. 2011. “A Hyperactive PiggyBac Transposase for Mammalian Applications.”

Proceedings of the National Academy of Sciences of the United States of America 108(4):
1531–36.

Zentner, Gabriel E et al. 2015. “ChEC-Seq Kinetics Discriminates Transcription Factor Binding Sites by DNA Sequence and Shape in Vivo.” *Nature Communications* 6(8733).
<http://dx.doi.org/10.1038/ncomms9733>.

Zhang, Ye et al. 2014. “An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex.” *Journal of Neuroscience* 34(36):
11929 –11947.

Zhang, Yong et al. 2008. “Model-Based Analysis of ChIP-Seq (MACS).” *Genome Biology*
9(R137).

Zhou, Pingzhu et al. 2017. “Mapping Cell Type-Specific Transcriptional Enhancers Using High Affinity, Lineage-Specific Ep300 BioChIP-Seq.” *eLife* 6: 1–29.

Supplemental figures

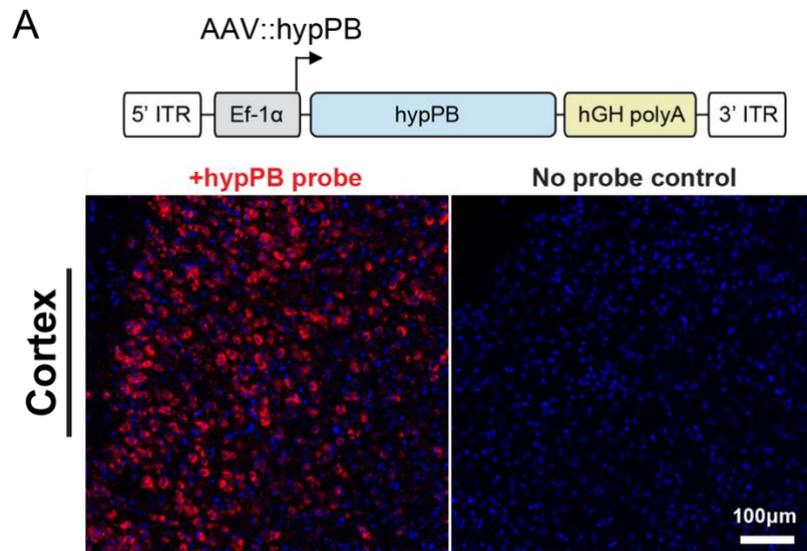
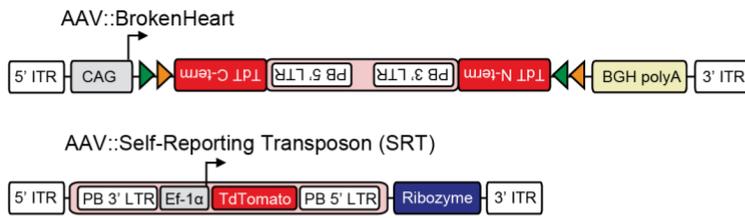


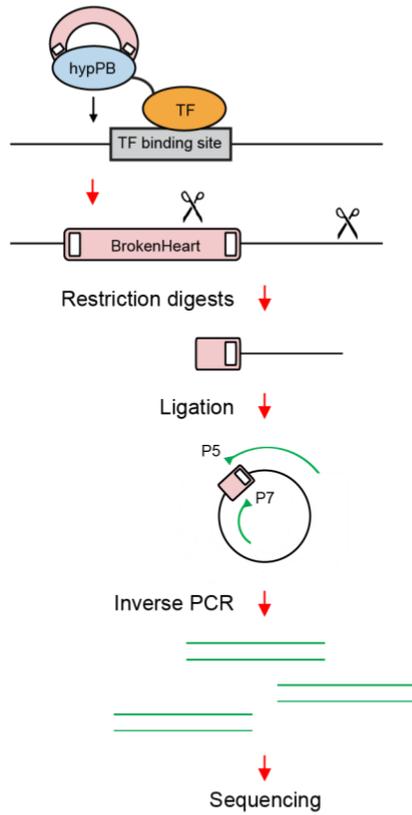
Figure 2.1S. AAV9 intracranial injection induces hypPB expression in the cortex.

(A) *In situ* hybridization of hypPB RNA and no probe control displaying widespread cortical expression of hypPB after delivery of AAV::hypPB.

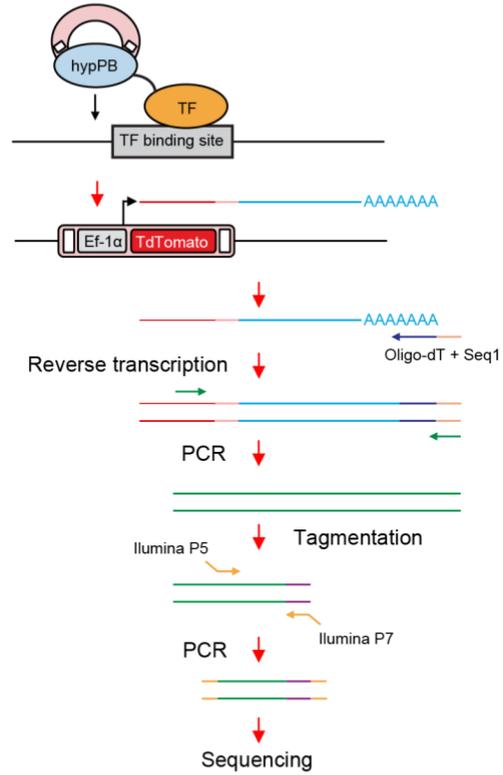
A



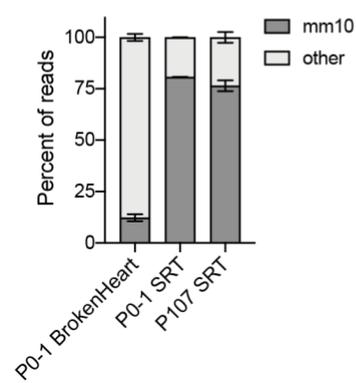
B



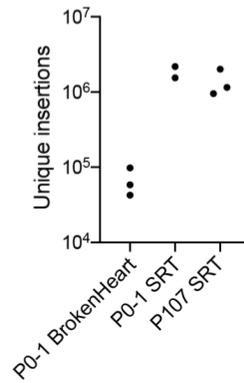
C



D



E



F

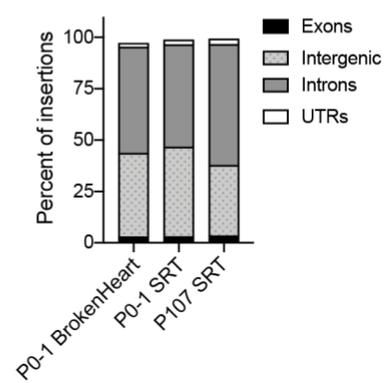


Figure 2.2S. Comparison of neonatal and adult AAV delivery of traditional DNA calling card and SRT calling card systems.

(A) Viral donor transposon constructs for DNA (AAV::BrokenHeart) and RNA (Self-Reporting Transposon; SRT) calling cards. (B) DNA calling cards library preparations were carried out as previously described in Wang et al., *Genetics*, 2012. TF-hyPB fusions insert BrokenHeart transposon DNA near TF binding sites. Genomic DNA is then harvested and digested with restriction enzymes that cut near the end of the transposon and in downstream genomic sequence. These fragments are subsequently self-ligated and circularized. From these, transposons and their flanking genomic sequences are amplified with inverse PCR, using primers that contain Illumina sequencing primers and adapters. Final products are sequenced and aligned to the mouse genome to map transposons to genomic locations. (C) Schematic of RNA calling cards library preparation protocol. SRTs contain a ubiquitous promoter that drives the transcription of a TdTomato reporter gene with no 3' poly-A termination signal. When the SRT gene is transcribed episomally in the context of the AAV genome (i.e. prior to transposition), a downstream self-cleaving ribozyme is incorporated into the transcript, destabilizing it and ultimately causing its degradation. However, once the transposon is inserted into the genome, the SRT is no longer affiliated with the ribozyme, and the SRT transcript becomes stabilized and extends to include flanking genomic sequence downstream of the TdTomato gene. RNA is then reverse transcribed and PCR amplified, and Illumina adapters are added for sequencing. (D) Read mapping rates for BrokenHeart DNA calling cards and SRT RNA calling cards libraries after P0-1 or adult cortical delivery, showing increased efficiency of recovery for genome-mapping reads in SRT relative to BrokenHeart. (n=3 mice for P0-1 BrokenHeart [12.3% mapping to mm10]; n=2 mice for P0-1 SRT [80.7%]; n=3 mice for adult SRT [76.5%]) (E) Number of unique insertions observed from BrokenHeart DNA calling cards and SRT RNA calling cards libraries (n=3 mice for P0-1 BrokenHeart [198,981 insertions were recovered at mean read coverage of 8.98 reads/insertion]; n=2 mice for P0-1 SRT [3,732,694 insertions at 3.84 reads/insertion]; n=3 mice for adult SRT [4,114,106 insertions at 5.92 reads/insertion]), displaying greatly increased recovery with SRT. (F) Mapping locations of transposon insertions in various genomic regions from BrokenHeart and SRT calling card libraries.

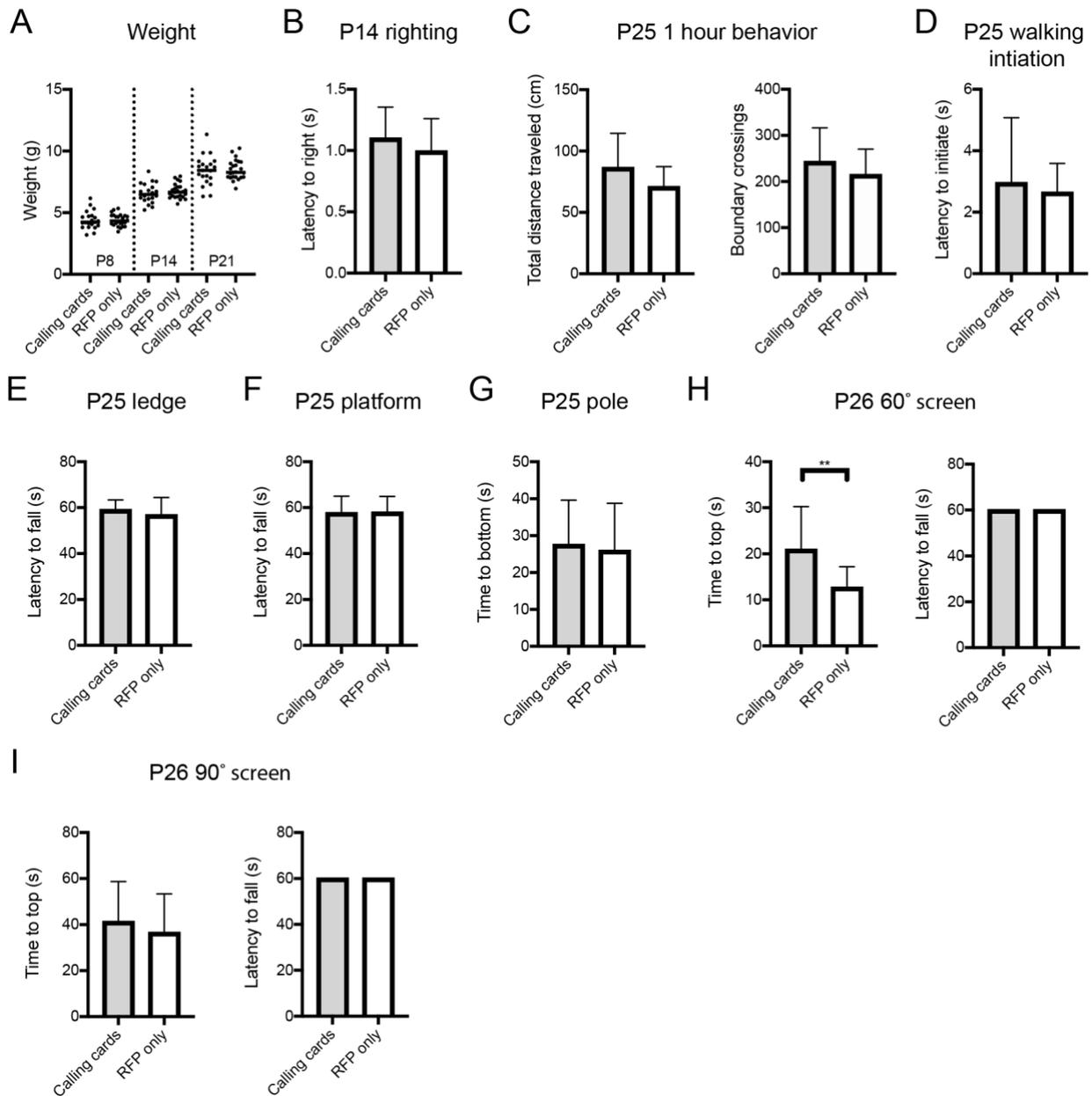


Figure 2.3S. AAV SRT calling cards system does not induce excess degeneration or behavioral/developmental deficits.

(A-I) Behavioral and developmental assessments of mice injected at P0-1 with SRT calling cards (n=21) or control, RFP only (n=24) viruses revealed few or no developmental, sensorimotor, or anxiety-related deficits in calling card animals relative to control. All group comparisons were done with two-tailed, unpaired Student's t-test, with Bonferroni corrected $\alpha=0.05$ as a significance threshold (including all tests in Fig 1). ** $p < 0.01$ (Bonferroni corrected).

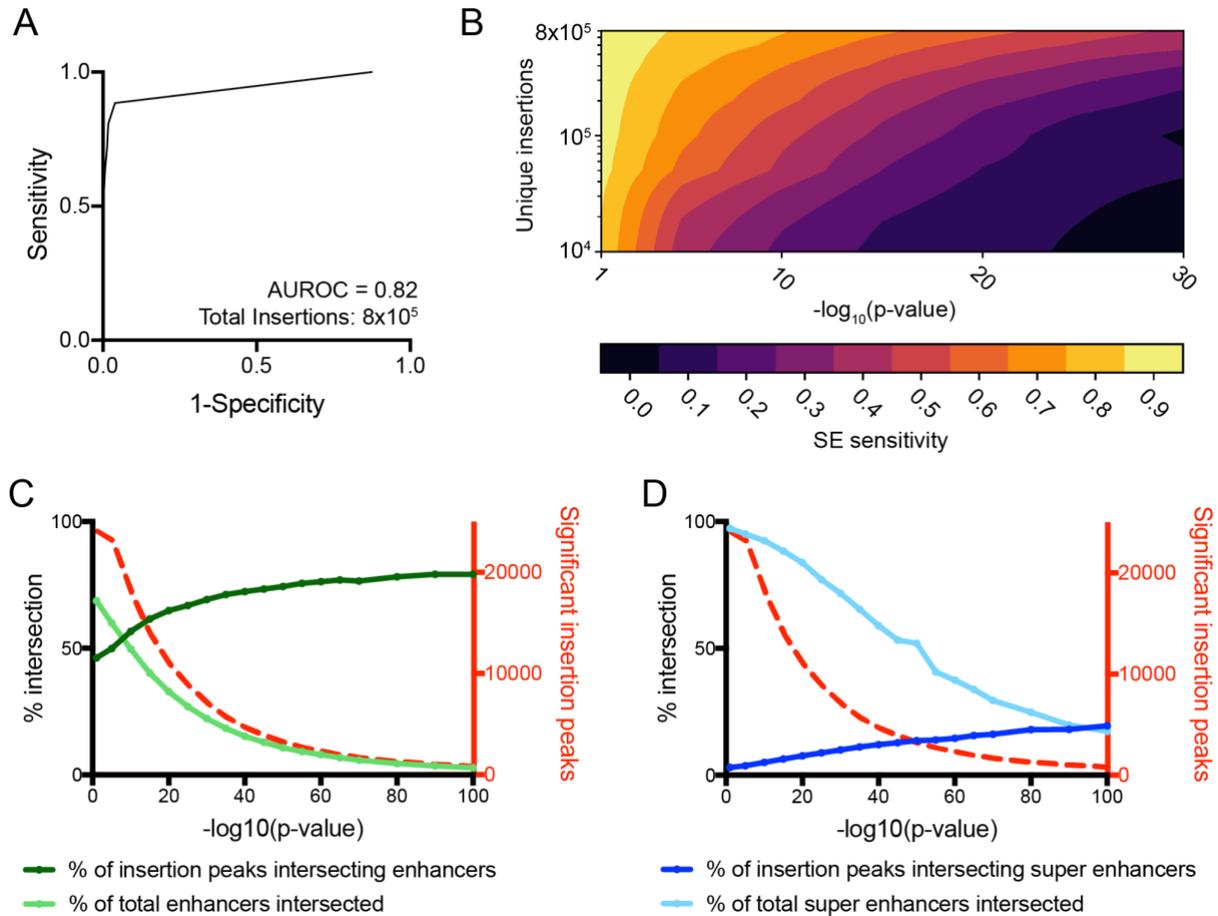


Figure 2.4S. *In vitro* and *in vivo* sensitivity of unfused hypPB calling card libraries for active REs.

(A-B) Sensitivity and specificity of super enhancer (SE) identification for unfused hypPB calling cards libraries in N2a cells. (A) Receiver-operator characteristic (ROC) curve for identification of active super enhancers using unfused hypPB peaks called from 800,000 unique insertions. Area under ROC curve: 0.82. (B) Super enhancer sensitivity at various significance thresholds and insertion totals, demonstrating high super enhancer sensitivity at even very low (10^4) insertion totals. (C-D) Intersections between significantly-enriched insertion peaks derived from *in vivo*, cortical unfused hypPB calling cards libraries and H3K27ac-marked (C) enhancers or (D) super enhancers at a range of insertion peak significance thresholds. Red line represents total number of significant peaks at each p-value threshold. Quantifications in Fig. 2F-I represent intersections at $p=10^{-30}$ significance threshold.

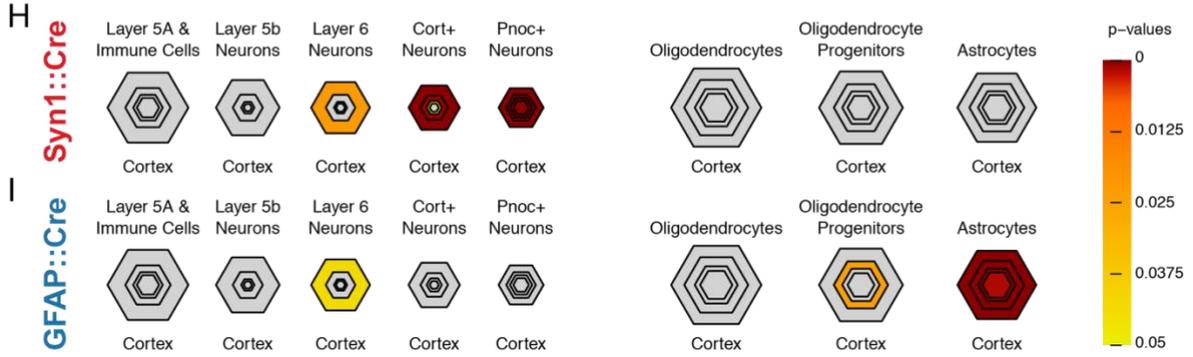
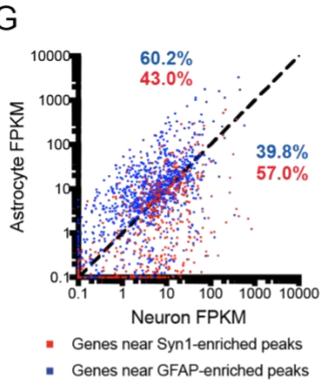
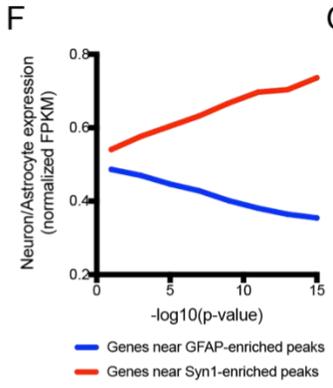
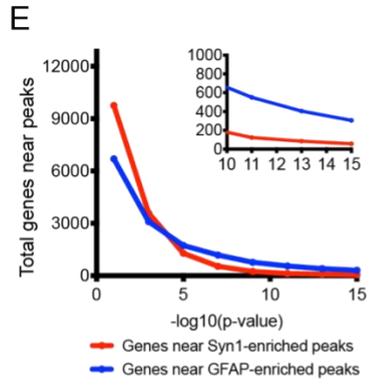
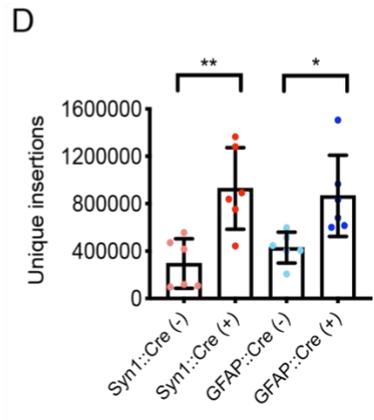
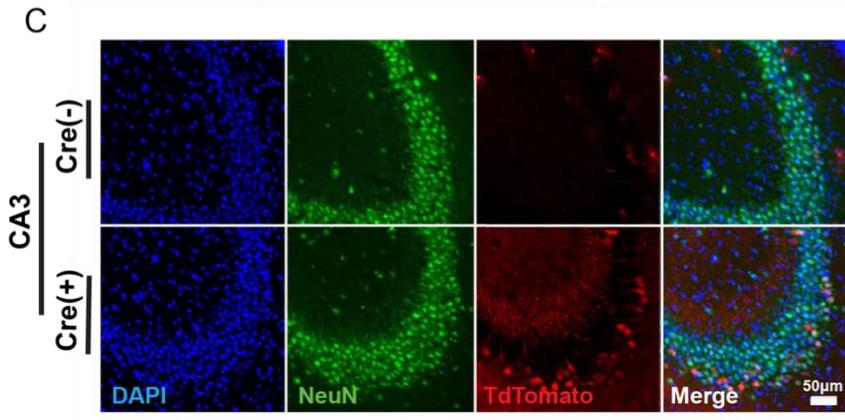
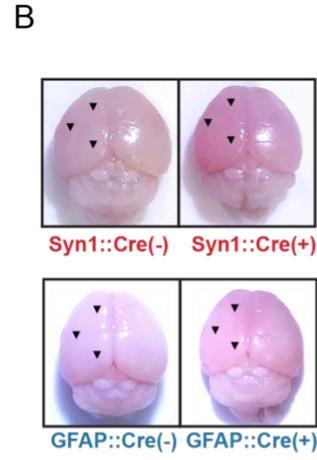
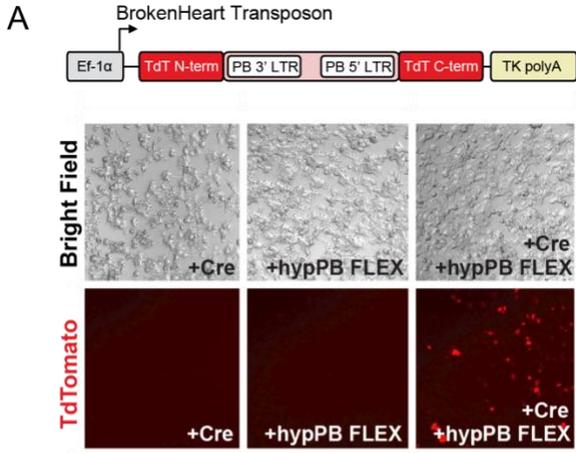


Figure 2.5S. Supplemental to: FLEX calling cards system generates cell type-specific RE profiles.

(A) Transfection of BrokenHeart transposons, hypPB FLEX plasmid, and Cre recombinase into HEK293T cells. n=3 wells per condition, 5 image fields per well, representative images shown. No TdTomato reporter from reconstituted BrokenHeart transposons observed in the absence of Cre or hypPB FLEX at 24-hours post-transfection. (B) Representative images of Syn1::Cre and GFAP::Cre positive and negative littermate brains, displaying increased SRT-derived TdTomato reporter signal in Syn1::Cre positive animals. (C) Representative images displaying preferential expression in Neu(+) neurons in CA3 hippocampal regions of Syn1::Cre(+) mice, but not negative littermates. TdTomato images taken at equal exposure times for direct comparison. (D) Quantifications of unique insertions in Syn1::Cre or GFAP::Cre positive and negative littermate mice (two-tailed, unpaired Student's t-test: * p<0.05, ** p<0.01). Syn1::Cre (+) vs (-): p=0.0032, t=3.86, df=10, 95% C.I._{diff} [268,014 : 1,000,132]; GFAP::Cre (+) vs (-): p=0.015, t=2.92, df=10, 95% C.I._{diff} [103,883 : 771,416] (E) Number of genes near differentially enriched insertion peaks in Syn1::Cre and GFAP::Cre animals at a range of significance thresholds. (F) Normalized neuron-to-astrocyte expression ratio [Neuron FKPM/(Neuron FKPM + Astrocyte FKPM)] for genes near Syn1::Cre or GFAP::Cre enriched peaks at a range of significance thresholds for defining differentially enriched insertion peaks. As significance threshold becomes more stringent, expression of nearby genes becomes more cell type-specific. RNA-seq expression from Zhang et al., *Journal of Neuroscience*, 2014. (G) Astrocyte versus neuron expression of all genes near Syn1::Cre or GFAP::Cre enriched insertion peaks at a stringent peak-calling significance threshold (p=10⁻⁷). Percent of genes on either side of the y=x midline shown. (H-I) Graphical representation of cortical cell type enrichment based on gene sets near either (H) Syn1::Cre (p<10⁻¹¹; top 123 genes) or (I) GFAP::Cre (p<10⁻²¹; top 131 genes) enriched insertion peaks. Legend displays Benjamini-Hochberg corrected Fisher's Exact Test p-value for overlap of reference cell type-specific gene sets and Syn1::Cre or GFAP::Cre candidate gene sets. Stringencies for enrichment for each pre-defined reference set are represented by size of each hexagon, with the outer ring being the least stringent set and inner ring being the most stringent set.

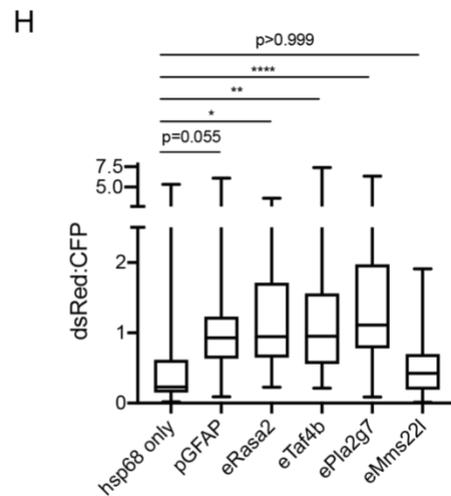
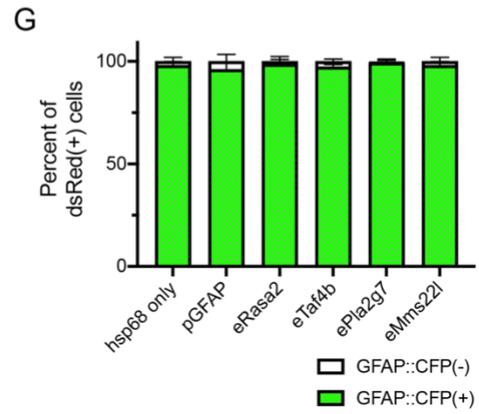
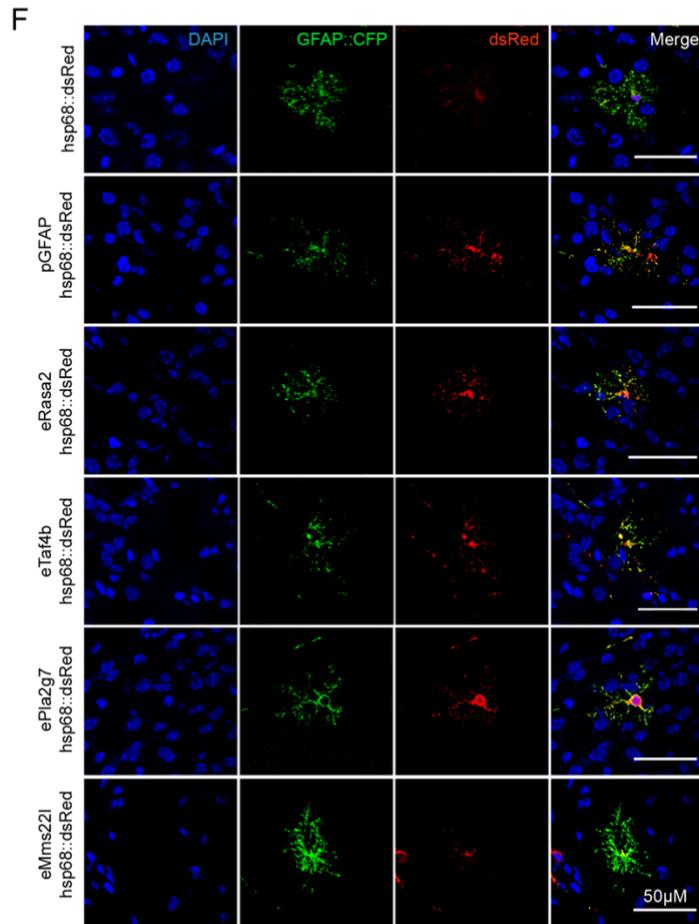
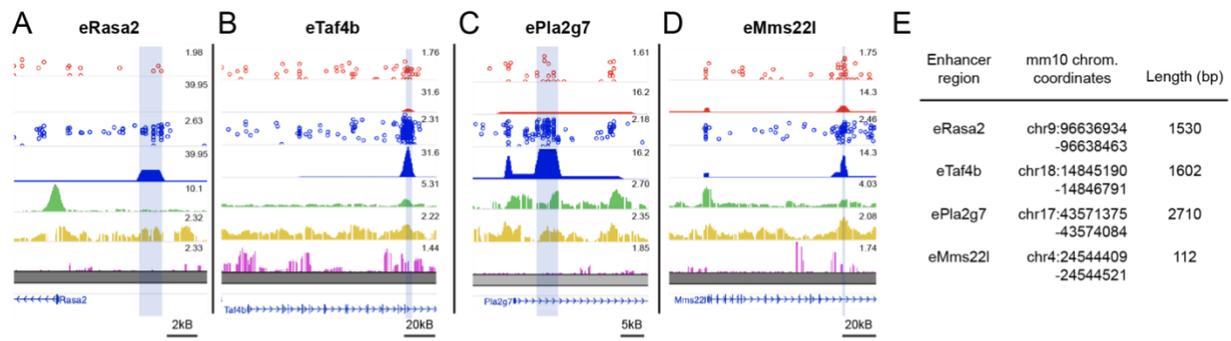


Figure 2.6S. Functional validation of enhancer activity at P7 for GFAP::Cre enriched unfused hypPB peaks.

(A-D) Candidate GFAP::Cre enriched insertion peaks that were chosen for functional enhancer validation based on enhancer size, significant GFAP::Cre enrichment, and astrocyte-specific RNA expression of their nearest genes. Each candidate RE (highlighted in blue) was separately cloned into a plasmid upstream of the hsp68 minimal promoter and a dsRed reporter gene for *in vivo* testing. (E) Chromosomal coordinates and lengths of candidate REs. (F) Candidate RE reporter constructs were co-delivered along with a GFAP::CFP plasmid to ventricle-proximal radial glia, including astrocytes, via PALE (Stogsdill et al., 2017). Expression of dsRed was enhanced by both the canonical GFAP promoter (pGFAP; positive control) and three of the four candidate REs (all but eMms22l). (G) Percentage of dsRed(+) cells in cortex co-labeled with GFAP::CFP, demonstrating that >96% of labeled cells are glia at this timepoint in all conditions. n=150 dsRed(+) cells from 3 brains per condition. (H) Quantification of dsRed expression enhancement in GFAP::CFP(+) astrocytes by pGFAP and candidate RE constructs. n=34-42 GFAP::CFP(+) cells from 3 brains per condition (one-way ANOVA with Dunnett's multiple comparisons test; *p<0.05; **p<0.01; ****p<0.0001). pGFAP_{mean diff.} = 0.66, 95% C.I._{diff} [0.01 : 1.33]. eRasa2_{mean diff.} = 0.67, 95% C.I._{diff} [0.03 : 1.32]. eTaf4b_{mean diff.} = 0.84, 95% C.I._{diff} [0.21 : 1.48]. ePla2g7_{mean diff.} = 1.22, 95% C.I._{diff} [0.58 : 1.86]. eMms22l_{mean diff.} = 0.002, 95% C.I._{diff} [-0.65 : 0.65].

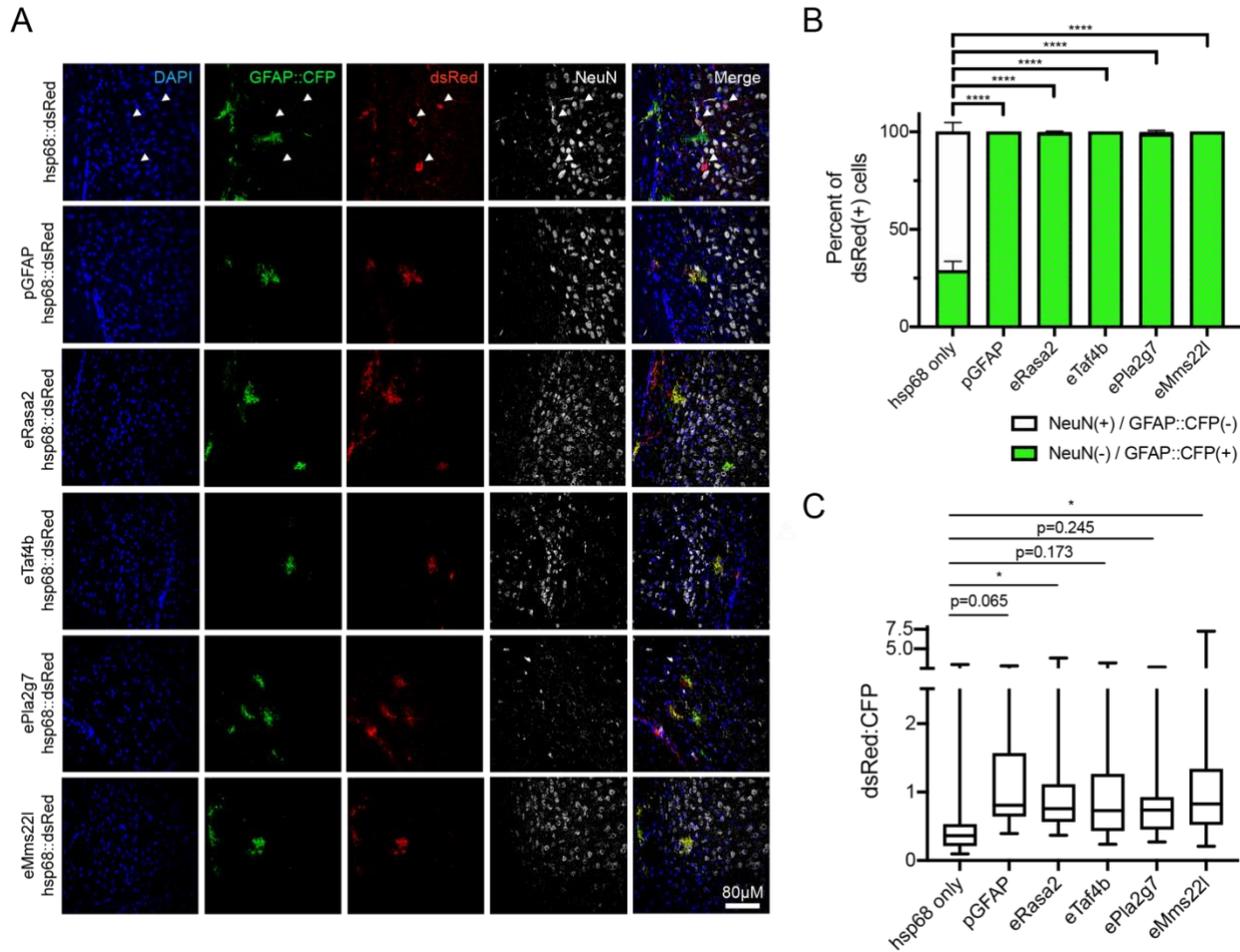


Figure 2.7S. Candidate astrocyte enhancers identified with FLEX calling cards direct cell type-specific expression in astrocytes by P21.

(A-B) Candidate astrocyte enhancers derived from FLEX calling cards enrichment (see Fig 6S) were electroporated into P0-1 mouse pups via PALE (Stogsdill et al., 2017) and animals were sacrificed at P21 for IF analysis. In animals receiving a plasmid containing dsRed driven by a minimal promoter only (hsp68::dsRed), dsRed expression in the cortex was evident in a population of NeuN(+) neurons (white arrows) that was not observed at P7 (see Fig 6SF-G). In contrast, in animals receiving either the pGFAP-driven positive control plasmid or plasmids containing the candidate enhancers, dsRed expression was significantly limited to GFAP::CFP(+) cells, indicating that the enhancer candidates facilitate cell type-specificity of gene expression. n=287-769 dsRed(+) cells from two brains per condition for RE candidates and n=67 dsRed(+) cells from one brain for pGFAP (one-way ANOVA with Dunnett's multiple comparisons test; ****p<0.0001). pGFAP_{mean diff.} = 74.47%, 95% C.I._{diff} [69.66% : 79.28%]. eRasa2_{mean diff.} = 73.57%, 95% C.I._{diff} [68.76% : 78.38%]. eTaf4b_{mean diff.} = 74.47%, 95% C.I._{diff} [69.66% : 79.28%]. ePla2g7_{mean diff.} = 72.37%, 95% C.I._{diff} [67.56% : 77.18%]. eMms221_{mean diff.} = 74.47%, 95% C.I._{diff} [69.66% : 79.28%]. (C) Quantification of dsRed expression enhancement in GFAP::CFP(+) astrocytes by pGFAP and candidate RE constructs. n=28-32 GFAP::CFP(+) cells from 2 brains per condition for RE candidates and n=17 GFAP::CFP(+) cells from one brain for pGFAP (one-way ANOVA with Dunnett's multiple comparisons test; *p<0.05). pGFAP_{mean diff.} = 0.59, 95% C.I._{diff} [-0.03 : 1.21]. eRasa2_{mean diff.} = 0.58, 95% C.I._{diff} [0.04 : 1.13]. eTaf4b_{mean diff.} = 0.42, 95% C.I._{diff} [-0.11 : 0.95]. ePla2g7_{mean diff.} = 0.40, 95% C.I._{diff} [-0.15 : 0.95]. eMms221_{mean diff.} = 0.62, 95% C.I._{diff} [0.10 : 1.15].

A

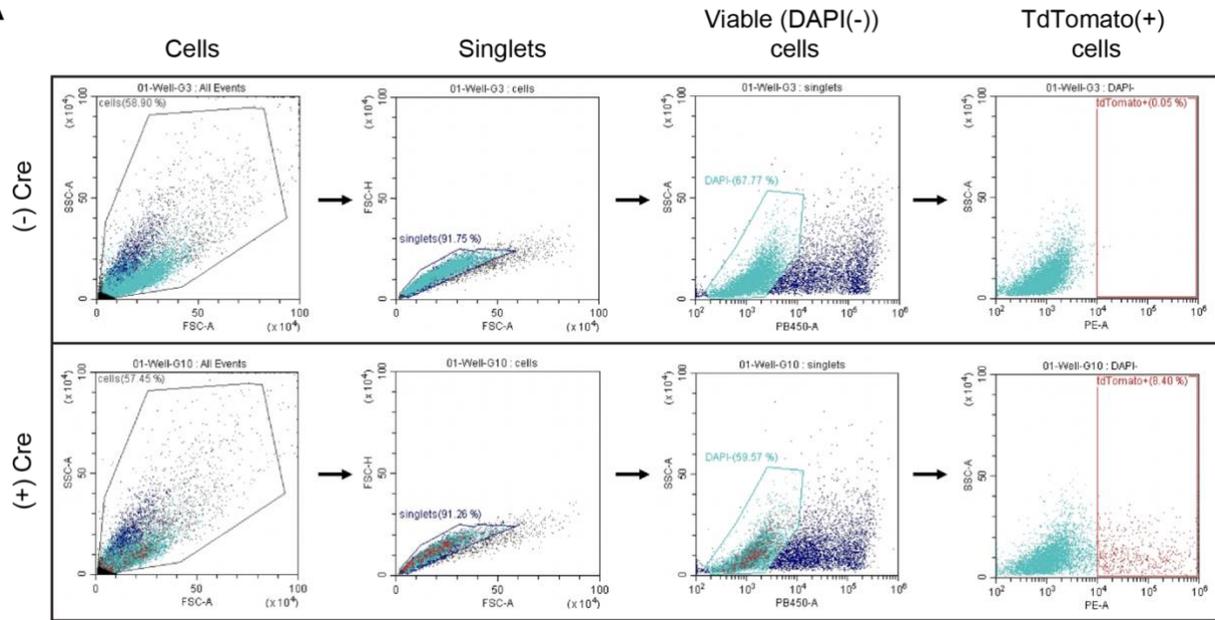


Figure 2.8S. Gating procedure for flow cytometry in cells with BrokenHeart-derived TdTomato fluorescence.

(A) Cells were dissociated into suspension and spiked with 0.1 μ g/ml DAPI as a viability dye, allowing discrimination of intact from plasma membrane-compromised cells. Gates were drawn based on Cre(-) conditions. The analyzed TdTomato(+) populations were selected from the parent cells/singlets/DAPI(-) subsets.

CHAPTER 3

**Retrospective BRD4 occupancy profiling reveals predictive vulnerability
and susceptibility loci in an acute seizure model**

Preface

This chapter contains contents from a manuscript in preparation:

Retrospective BRD4 occupancy profiling reveals predictive vulnerability and susceptibility loci in an acute seizure model

Alexander J. Cammack¹, Mark Shabsovich¹, Joseph D. Dougherty^{3,4}, Robi D. Mitra^{2,3}, and Timothy M. Miller¹

¹Washington University School of Medicine, Department of Neurology, St. Louis, MO

²Washington University School of Medicine, Edison Family Center for Genome Sciences and Systems Biology, St. Louis, MO

³Washington University School of Medicine, Department of Genetics, St. Louis, MO

⁴Washington University School of Medicine, Department of Psychiatry, St. Louis, MO

Abstract

Epilepsy syndromes are characterized by the recurrence of acute seizures. Previous evidence suggests that epigenetic mechanisms may be involved in epileptogenesis and/or progression to severe seizure. One such mechanism is enhancer activation, which is mediated by the chromatin reader, BRD4. BRD4 mediates activity-dependent expression of immediate-early genes in neurons, and BRD4 inhibition has been shown to be protective from seizure. Here, we profile BRD4 genome occupancy in a mouse model of acute seizure, prior to the induction of the seizure itself, and in doing so, identify enhancers for which BRD4 binding predicts eventual seizure severity outcomes. Many of these enhancers were proximal to genes with known roles in seizure and epilepsy, and gene ontological analysis revealed enrichment of synaptic transmission-related categories. These results suggest that BRD4 binding at specific loci may predispose animals to severe seizures and support a role of BRD4 in mediating seizure susceptibility pathways. Manipulation of BRD4 and/or its binding partners may be an effective therapeutic strategy for seizure in the future.

Introduction

Epilepsy is the fourth most common neurological disorder in the United States, with an incidence of 68 per 100,000 persons (Fiest et al. 2017) and a lifetime risk of 1 in 26 (England et al. 2012). The term epilepsy refers to an amalgam of neurological conditions, each affecting various at-risk populations, however these disorders are all characterized by the unifying phenotype of unprovoked, recurrent seizures (Myers and Mefford 2015). Perhaps the most detrimental cases of seizures are *status epilepticus* (SE), in which a single seizure persists for a sustained period of time (>30 min) without recovery of consciousness. The occurrence of SE can cause severe neuronal damage and result in impairment of vital functions which can be fatal (Betjemann and Lowenstein 2015). While a number of anti-epileptic drugs (AEDs) have been identified and are now commonly used for the treatment of seizures (Löscher 2011; Rogawski 2006), there remains at least one third of the epileptic individuals for which their seizures are unresponsive to treatment and continue uncontrolled (Devinsky et al. 2018).

Seizures are the result of imbalances in inhibitory and excitatory neural conductances, which lead to massive, simultaneous propagations of neuronal activity (Staley 2015). These imbalances are typically brief, however in instances of SE, they continue unchecked, resulting in a sustained, severe seizure. Excessive excitation during the early stages of seizure and/or ineffective recruitment of inhibitory, GABAergic signaling may both be involved in progression from mild seizure to severe seizure in SE, though the underlying mechanisms are not yet fully understood (Betjemann and Lowenstein 2015). In some cases, conductance imbalances are caused by mutations in proteins directly involved in membrane potential and excitability, such as ion channels, and are thus termed “channelopathies” (Myers and Mefford 2015). However, these mutations do not account for the majority of epilepsy incidences (Helbig et al. 2008; Helbig and

Lowenstein 2013). Thus, it is critical to understand the molecular mechanisms underlying both onset and termination of seizure for the development of effective therapeutics.

Recent evidence suggests that transcription factor (TF)- and enhancer-mediated epigenetic regulation may be involved in determining susceptibility to seizure. One such factor is a chromatin reader called BRD4, a member of the bromo and extraterminal (BET) domain family, which binds to acetylated histone tails in enhancer regions (Jung et al. 2014; Kanno et al. 2004, 2014) and facilitates initiation of transcription at associated gene promoters (LeRoy, Rickards, and Flint 2008; Mochizuki et al. 2008). BRD4 is highly expressed in neurons, where it regulates the expression of activity-dependent genes, such as immediate-early genes, and other genes involved in synaptic transmission and membrane excitability (Korb et al. 2015). Critically, pharmacological inhibition of BRD4 in mice prior to chemical induction of acute seizure (via the GABA antagonist, pentylentetrazole (PTZ)) is sufficient to reduce seizure severity (Korb et al. 2015), suggesting that BRD4 may regulate mechanisms underlying progression from mild to severe and prolonged seizure. However, the downstream targets of BRD4 involved in this process are not yet understood.

Here we identify BRD4-bound genomic loci that correlate with seizure severity in the PTZ-induced acute seizure model in order to find factors that may underlie seizure vulnerability and susceptibility. To do so, we utilize a newly developed epigenomic profiling technique, AAV calling cards (Cammack et al. 2019), which allows for the recording of BRD4 genome occupancy in the mouse cortex *prior to* seizure induction; thus these profiles represent the “pre-seizure” state which is unaltered by the seizure itself. In this study, we compare pre-seizure BRD4 occupancy between mice incurring mild versus severe seizures and, in doing so, identify a number of regulatory elements (REs) predictive of seizure severity outcomes *in vivo*.

Methods

Animals

All animal procedures were approved by the Washington University in St. Louis Institutional Animal Care and Use Committee (IACUC) in accordance with National Institutes of Health (NIH) guidelines. All mice used in this study were of the Synapsin 1 (Syn1)::Cre (RRID:IMSR_JAX:003966) transgenic mouse strain on a C57BL/6J background. Mice had unlimited access to food and water and were housed in a facility with a 12-hour light/dark cycle. For euthanasia and tissue collection, mice were anesthetized with isoflurane and perfused with 15ml of cold saline (PBS), after which brains were dissected and flash frozen in liquid nitrogen. For P0-1 injections, pregnant females were monitored daily for litter delivery and pups were injected with adeno-associated virus (AAV) within 24 hours of being found in the cage. As all mice in this study were euthanized by P28, mice were not weaned prior to sacrifice.

PTZ-induced seizure induction and evaluation

PTZ was dissolved in sterile PBS at 5mg/mL prior to injection. For induction of acute seizure, mice were injected intraperitoneally (i.p.) with PTZ at 65mg/mL and evaluated for 15 minutes post-injection while recording with a video camera. Immediately following the 15-minute evaluation period, mice were euthanized as described above, and brains were dissected and collected in liquid nitrogen.

Seizures were scored from the video recording based on previously published scales (DeVos et al. 2013; Löscher and Nolting 1993; Racine 1972) with scores as follows: 0: normal; 1: immobility; 2: spasm, tremble, or twitch; 3: tail extension; 4: forelimb clonus; 5: generalized clonic activity; 6: jumping or running seizures; 7: full tonic extension; and 8: death. For the

purposes of this study, scores of 4 or less were considered “mild,” while 5 and above were “severe.” After scoring, animals were selected for BRD4 calling cards analysis.

Virus generation and injections

A conditional expression, or “FLEX,” calling cards system was chosen for BRD4 profiling (Cammack et al. 2019). A Cre-dependent hypPB construct and a (self-reporting transposon (SRT) construct were separately packaged into AAV serotype 9 (AAV9) vectors by the Hope Center Viral Vectors Core at Washington University School of Medicine. For all experiments, AAV9::hypPB FLEX and AAV9::SRT viruses were mixed equivolume at maximum titer and co-injected into 3 cortical sites per hemisphere of P0-1 pups (1 μ l viral mix per site), as previously described (Cammack et al. 2019).

Viral titers (viral genomes per milliliter) prior to mixing were as follows:

AAV9::hypPB FLEX = 8.0×10^{12} - 1.0×10^{13} vg/mL

AAV9::SRT = 1.6×10^{13} - 2.2×10^{13} vg/mL

Calling cards library preparation and sequencing

Self-reporting transposon (SRT) calling card libraries were prepared from cortical RNA samples as previously described (Cammack et al. 2019; Moudgil et al. 2019). The cortex of each animal was randomly separated into 10 pieces from which RNA samples were individually prepped. Each RNA sample was subsequently reverse transcribed and prepared for sequencing, as previously described (Cammack et al. 2019; Moudgil et al. 2019), resulting in 10 barcodes per animals which were sequenced on Illumina HiSeq 2500, NextSeq 500, or MiniSeq platforms (Illumina, San Diego, CA, USA). Raw reads were filtered for features of true insertion events

and mapped to the mouse genome (mm10) with Novoalign 3 (Novocraft Technologies). Unique insertions were defined as existing either in different TTAA's in the same barcode or the same TTAA but in different barcodes, and all insertions were all considered equal, regardless of read depth.

Identification of differentially BRD4-bound loci in mild versus severe animals

Significantly enriched insertion peaks between mild and severe animals were identified as previously described (Cammack et al. 2019; Moudgil et al. 2019). Prior to peak calling, each animal's library was individually downsampled to 440,000 unique insertions. Downsampled libraries were then pooled into "severe" and "mild" groups (n=6 animals per group), to create final pooled profiles of 2,640,000 insertions per group. After this, the genome was segmented into blocks of constant insertion density and a p-value was assigned to each block based on the differential enrichment between the two pooled profiles based on a count-based statistical comparison. In these analyses, a significance threshold of $p < 10^{-7}$ was used, and all blocks surpassing this threshold were considered peaks significantly enriched in mild over severe or severe over mild profiles. All significant peaks past this cutoff reach statistical significance after Bonferroni correction based on the total number of blocks in each final pooled library. To further delineate predictive seizure outcome factors, BRD4-directed insertion depth at these loci was then correlated with severity score via linear regression.

Gene-enhancer pairing and gene ontology

The nearest gene (or genes for direct intersections) to each differentially enriched BRD4 calling cards peak was defined and filtered out if greater than 10,000 bases away from the peak,

as these are low confidence gene-peak pairings. Gene ontology (GO) enrichment analysis was performed on resultant gene sets with the PANTHER v14 classification system (Mi et al. 2017).

Statistical analyses

Statistical tests were done with GraphPad Prism v8.1.2 assuming normal distributions.

Data and code availability

Raw and processed data will be made available through GEO following publication. Calling card analysis software were previously published (Moudgil et al. 2019) and are available upon request.

Results

Recording pre-seizure BRD4 occupancy profiles in a chemically induced acute seizure model

We sought to identify BRD4-bound genomic loci predictive of seizure severity in the PTZ-induced acute seizure model. We delivered BRD4-directed, conditionally expressed (“FLEX”), AAV calling cards viral vectors to the cortices of a large cohort of P0-1 mouse pups of the *Syn1::Cre* genotype, which enriches BRD4 signal in neurons while maintaining some background in astrocytes (Cammack et al. 2019). At P28, after recording pre-seizure BRD4 occupancy profiles in each animal, mice were administered acute seizures through i.p. injection of the GABA antagonist, PTZ, at 65mg/kg (**Fig 3.1A**). Mice were euthanized immediately following seizure, and cortical tissues were collected for calling cards analyses. To determine the degree of variability in seizure severity following PTZ injection, seizures were scored on a standard severity scale (DeVos et al. 2013; Löscher and Nolting 1993; Racine 1972), ranging from 1 (e.g. immobility) to 8 (e.g. death). Of the 27 total mice in the study, 14 scored with a “mild” seizure (0-4) and 13 with a “severe” (5-8), which were determined by the absence or presence, respectively, of generalized clonic seizure. Next, we asked whether this variability could be explained by sex, genotype, or weight. We found no correlation between severity score and any of these factors (**Fig 3.1B-D**).

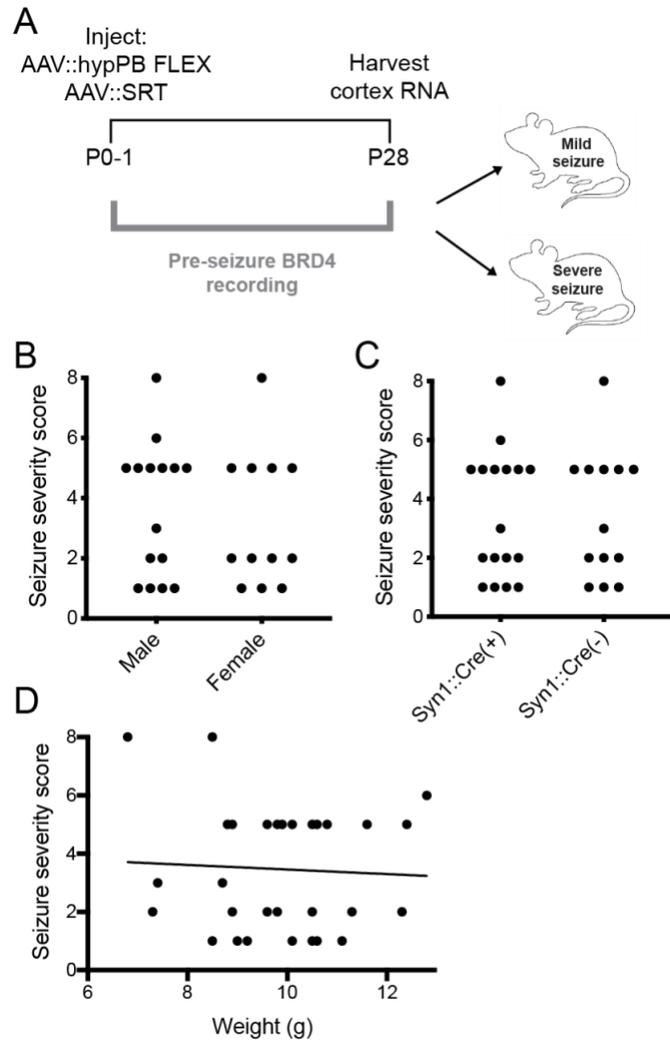


Figure 3.1: Seizure severity does not correlate with mouse sex, genotype, or weight.

(A) Experimental paradigm for longitudinal, pre-seizure BRD4 recording. (B-D) Variability in seizure severity is not explained by (B) sex, (C) *Syn1::Cre* genotype, or (D) animal weight.

Identification of loci differentially BRD4-bound in mild versus severe seizures

To test whether BRD4 occupancy at specific loci might associate with seizure severity, we compared BRD4 calling card profiles between mice incurring mild versus severe seizures. We prepared and sequenced calling card libraries from a sex and littermate-matched subset of animals with mild or severe seizures (n=6 per category) and called differential BRD4 peaks between the combined mild and severe profiles. We observed 341 REs enriched for BRD4 binding in severe seizure animals and 226 enriched in mild at $p < 10^{-7}$. A number of these loci exhibited a strong positive or negative correlation with peak seizure severity; thus, BRD4 occupancy at these peaks is predictive of seizure outcomes. Intriguingly, some of these peaks were near to genes with known roles in membrane excitability and/or seizure susceptibility. For example, one differential binding peak was within an intron of the *CNTNAP2* gene, of which mutations cause cortical dysplasia/focal epilepsy syndrome (Strauss et al. 2006) and knockout mice display profound hyperactivity and epileptic seizures (Peñagarikano et al. 2011); accordingly, we observed that BRD4 binding at this locus is negatively correlated with seizure severity (**Fig 3.2A-B**). Two other peaks, which displayed positive correlations of BRD4 occupancy and seizure severity, were proximal to *SYN2* (**Fig 3.2D-E**) and *NBEA* genes which also harbor epilepsy- and neurodevelopmental disease-associated mutations (Cavalleri et al. 2007; Corradi et al. 2014; Mulhern et al. 2018) and have displayed epileptic phenotypes in cellular and animal models (Etholm and Heggelund 2009; Feliciano, Andrade, and Bykhovskaia 2013).

To extend these observations genome-wide, we performed GO for gene sets proximal to BRD4 calling card peaks enriched in severe or mild animals. Interestingly, we found that GO terms for both gene sets were highly enriched for synaptic and neurodevelopmental genes (**Fig**

3.2C,F), suggesting that BRD4 regulates pathways and processes critical to proper neuronal development and function.

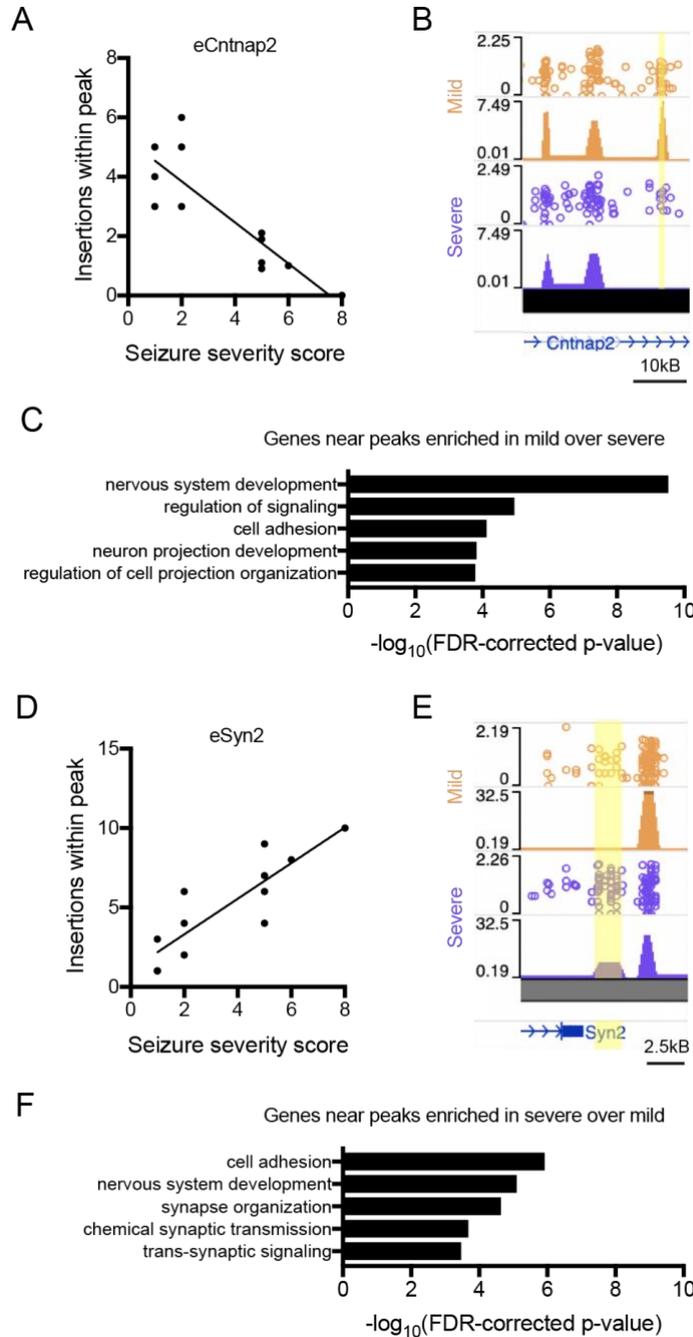


Figure 3.2: Retrospective analysis of longitudinally-recorded BRD4 occupancy identifies vulnerability and resiliency loci that are predictive of seizure outcome.

(A,B,D,E) Example genes and (C,F) gene ontology terms of genes proximal to differentially enriched BRD4 binding sites in mild over severe animals (A-C) or severe over mild animals (D-F). $n = 6$ mice per group. Peaks defined at $p < 10^{-7}$; eCntnap2: $R=0.85$, $p=0.0004$; eSyn2: $R=0.86$, $p=0.0003$.

Discussion

Epileptic seizures are caused by spontaneous neuronal overactivity. This overactivity induces numerous changes in gene and protein expression, many of which occur immediately following seizure onset. Indeed, within minutes of incurrence of seizure, a number of genes, including immediate-early genes such as *c-Fos*, have been shown to be dramatically altered at the mRNA and protein levels (Peng and Houser 2005), a process which is facilitated by BRD4 (Korb et al. 2015). Thus, we hypothesized that BRD4 binding at specific genomic loci would be predictive of and may mechanistically underlie downstream seizure severity phenotypes.

However, activity-dependent changes induced by seizure may mask important, inherent susceptibility factors present at seizure onset. To overcome these limitations, we longitudinally recorded the prodromal genome occupancy of BRD4 with AAV calling cards in an acute seizure model *prior to seizure*. Utilizing this novel approach, we identified a number of loci which are predictive of seizure outcomes. Of these differentially BRD4-bound regulatory elements (REs), we observed that many were located near to or within synaptic and excitability genes, perhaps explaining why BRD4 inhibition protects from severe seizure. Moreover, for some of these genes there is preceding evidence supporting their involvement in seizure susceptibility (Cavalleri et al. 2007; Corradi et al. 2014; Etholm and Heggelund 2009; Feliciano et al. 2013; Mulhern et al. 2018; Peñagarikano et al. 2011; Strauss et al. 2006). Critically, by recording BRD4 binding before seizure induction, these profiles are largely unaltered by the seizure itself and thus represent the pre-seizure state. This retrospective paradigm is particularly important here, given the rapidity with which epigenetic and gene expression profiles can change in the cellular context upon insult or stimulation (Stavreva et al. 2015).

However, there are number of limitations to this study as well. Firstly, it remains unclear whether these associations are correlative or causative in nature. Future studies in knockout or overexpression models may help to delineate whether individual loci directly regulate seizure susceptibility. Moreover, while pairing REs to specific genes via proximity is widely done (Ernst et al. 2011) and is generally a valid proxy for RE-mediated gene identification (Creyghton et al. 2010; Heintzman et al. 2009; Visel et al. 2009), this method is inherently imperfect (Maurano et al. 2012); thus, it is likely that some differentially BRD4-bound loci regulate genes other than those analyzed here. Finally, while the four-week BRD4 recording period dwarfs the 15-minute seizure observance period, it remains possible that BRD4 occupancy profiles may be altered in some cases during the insult itself; however, given the length of time of BRD4 recording and the previously-described permanent recording ability of AAV calling cards (Cammack et al. 2019), we consider this unlikely to be a significant contributor to the profiles described here.

In summary, these results support that pre-seizure BRD4 genome occupancy is predictive of seizure outcomes. This work delineates specific REs, genes, and pathways through which BRD4 acts to functionally alter seizure susceptibility, providing potential mechanisms to explain the previously demonstrated role for BRD4 in neuronal excitability (Korb et al. 2015). However, perhaps most importantly, this study introduces a novel experimental paradigm to solve a common biological conundrum: how to study pre-insult states while also observing post-insult phenotypic outcomes. We feel similar retrospective analysis paradigms may be equally applicable in the future in other neural contexts in which historical epigenetic information would be useful, such as models of neurodevelopment or neurodegeneration.

Acknowledgements

This study was funded by the Muscular Dystrophy Association, NINDS R01NS078398 (to TMM), U01MH10913301 (to JDD and RDM), RF01MH117070-01 (to JDD and RDM), R21HG009750 (to RDM) and the Hope Center Viral Vectors Core at Washington University School of Medicine. AJC was supported by 5T32GM008151-32 and the Lucille P. Markey Special Emphasis Pathway in Human Pathobiology. We thank the Edison Family Center for Genome Sciences & Systems Biology, specifically Jessica Hoisington-Lopez and MariaLynn Crosby, for assistance with genomic analysis.

Author contributions

Study designed by AJC, JDD, RDM, and TMM. Data generation and analysis by AJC and MS. Manuscript written by AJC.

Bibliography

- Betjemann, John P., and Daniel H. Lowenstein. 2015. "Status Epilepticus in Adults." *The Lancet Neurology* 14(6): 615–24. [http://dx.doi.org/10.1016/S1474-4422\(15\)00042-3](http://dx.doi.org/10.1016/S1474-4422(15)00042-3).
- Cammack, Alexander J. et al. 2019. "A Viral Toolkit for Recording Transcription Factor-DNA Interactions in Live Mouse Tissues." *BioRxiv [Preprint]* (<https://doi.org/10.1101/538504>).
- Cavalleri, Gianpiero L. et al. 2007. "Multicentre Search for Genetic Susceptibility Loci in Sporadic Epilepsy Syndrome and Seizure Types: A Case-Control Study." *Lancet Neurology* 6(11): 970–80.
- Corradi, Anna et al. 2014. "SYN2 Is an Autism Predisposing Gene: Loss-Of-function Mutations Alter Synaptic Vesicle Cycling and Axon Outgrowth." *Human Molecular Genetics* 23(1): 90–103.
- Creyghton, M. P. et al. 2010. "Histone H3K27ac Separates Active from Poised Enhancers and Predicts Developmental State." *Proceedings of the National Academy of Sciences* 107(50): 21931–36. <http://www.pnas.org/cgi/doi/10.1073/pnas.1016071107>.
- Devinsky, Orrin et al. 2018. "Epilepsy." *Nature Reviews Disease Primers* 3(18024).
- DeVos, Sarah L. et al. 2013. "Antisense Reduction of Tau in Adult Mice Protects against Seizures." *Journal of Neuroscience* 33(31): 12887–97.
- England, Mary Jane, Catharyn T. Liverman, Andrea M. Schultz, and Larisa M. Strawbridge. 2012. "Epilepsy Across the Spectrum: Promoting Health and Understanding." *Washington DC: National Academies Press*. <https://www.ncbi.nlm.nih.gov/books/NBK100596/>.
- Ernst, Jason et al. 2011. "Mapping and Analysis of Chromatin State Dynamics in Nine Human Cell Types." *Nature* 473(7345): 43–49. <http://dx.doi.org/10.1038/nature09906>.
- Etholm, Lars, and Paul Heggelund. 2009. "Seizure Elements and Seizure Element Transitions

- during Tonic-Clonic Seizure Activity in the Synapsin I/II Double Knockout Mouse: A Neuroethological Description.” *Epilepsy and Behavior* 14(4): 582–90.
<http://dx.doi.org/10.1016/j.yebeh.2009.02.021>.
- Feliciano, Pedro, Rodrigo Andrade, and Maria Bykhovskaia. 2013. “Synapsin Ii and Rab3A Cooperate in the Regulation of Epileptic and Synaptic Activity in the CA1 Region of the Hippocampus.” *Journal of Neuroscience* 33(46): 18319–30.
- Fiest, Kirsten M. et al. 2017. “Author Response: Prevalence and Incidence of Epilepsy: A Systematic Review and Meta-Analysis of International Studies.” *Neurology* 89(6): 641–42.
- Heintzman, Nathaniel D. et al. 2009. “Histone Modifications at Human Enhancers Reflect Global Cell-Type-Specific Gene Expression.” *Nature* 459(7243): 108–12.
<http://dx.doi.org/10.1038/nature07829>.
- Helbig, Ingo, and Daniel H. Lowenstein. 2013. “Genetics of the Epilepsies: Where Are We and Where Are We Going?” *Current Opinion in Neurology* 26(2): 179–85.
- Helbig, Ingo, Ingrid E. Scheffer, John C. Mulley, and Samuel F. Berkovic. 2008. “Navigating the Channels and beyond: Unravelling the Genetics of the Epilepsies.” *The Lancet Neurology* 7(3): 231–45.
- Jung, Marie et al. 2014. “Affinity Map of Bromodomain Protein 4 (BRD4) Interactions with the Histone H4 Tail and the Small Molecule Inhibitor JQ1.” *Journal of Biological Chemistry* 289(13): 9304–19.
- Kanno, Tomohiko et al. 2004. “Selective Recognition of Acetylated Histones by Bromodomain Proteins Visualized in Living Cells.” *Molecular Cell* 13(1): 33–43.
- . 2014. “BRD4 Assists Elongation of Both Coding and Enhancer RNAs by Interacting with Acetylated Histones.” *Nature Structural and Molecular Biology* 21(12): 1047–57.

<http://dx.doi.org/10.1038/nsmb.2912>.

- Korb, Erica et al. 2015. “BET Protein Brd4 Activates Transcription in Neurons and BET Inhibitor Jq1 Blocks Memory in Mice.” *Nature Neuroscience* 18(10): 1464–73.
- LeRoy, Gary, Brenden Rickards, and S. J. Flint. 2008. “The Double Bromodomain Proteins Brd2 and Brd3 Couple Histone Acetylation to Transcription.” *Molecular Cell* 30(1): 51–60.
- Löscher, Wolfgang. 2011. “Critical Review of Current Animal Models of Seizures and Epilepsy Used in the Discovery and Development of New Antiepileptic Drugs.” *Seizure* 20(5): 359–68. <http://dx.doi.org/10.1016/j.seizure.2011.01.003>.
- Löscher, Wolfgang, and Björn Nolting. 1993. “The Role of Technical, Biological, and Pharmacological Factors in the Laboratory Evaluation of Anticonvulsant Drugs. IV. Protective Indicies.” *Epilepsy Research* 9: 1–10.
- Maurano, Matthew T. et al. 2012. “Systematic Localization of Common Disease-Associate Variation in Regulatory DNA.” *Science* 337(September): 1190–95.
- Mi, Huaiyu et al. 2017. “PANTHER Version 11: Expanded Annotation Data from Gene Ontology and Reactome Pathways, and Data Analysis Tool Enhancements.” *Nucleic Acids Research* 45(D1): D183–89.
- Mochizuki, Kazuki et al. 2008. “The Bromodomain Protein Brd4 Stimulates G1 Gene Transcription and Promotes Progression to S Phase.” *Journal of Biological Chemistry* 283(14): 9040–48.
- Moudgil, Arnav et al. 2019. “Self-Reporting Transposons Enable Simultaneous Readout of Gene Expression and Transcription Factor Binding in Single Cells.” *BioRxiv [Preprint]* (<https://doi.org/10.1101/538553>).
- Mulhern, Maureen S. et al. 2018. “NBEA: Developmental Disease Gene with Early Generalized

- Epilepsy Phenotypes.” *Annals of Neurology* 84(5): 788–95.
- Myers, Candace T, and Heather C Mefford. 2015. “Advancing Epilepsy Genetics in the Genomic Era.” *Genome medicine* 7(1): 91.
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4549122&tool=pmcentrez&rendertype=abstract>.
- Peñagarikano, Olga et al. 2011. “Absence of CNTNAP2 Leads to Epilepsy, Neuronal Migration Abnormalities, and Core Autism-Related Deficits.” *Cell* 147(1): 235–46.
- Peng, Zechun, and Carolyn R. Houser. 2005. “Temporal Patterns of Fos Expression in the Dentate Gyrus after Spontaneous Seizures in a Mouse Model of Temporal Lobe Epilepsy.” *Journal of Neuroscience* 25(31): 7210–20.
- Racine, Ronald J. 1972. “Modification of Seizure Activity by Electrical Stimulation: II. Motor Seizure.” *Electroencephalography and Clinical Neurophysiology* 32: 281–94. https://ac.els-cdn.com/0013469472901770/1-s2.0-0013469472901770-main.pdf?_tid=eb2b4ab8-4f74-4878-85a2-e7753e50eb04&acdnat=1525166976_62ef7d695a5275b96dff44ff3876c892.
- Rogawski, Michael A. 2006. “Diverse Mechanisms of Antiepileptic Drugs in the Development Pipeline.” *Epilepsy Research* 69(3): 273–94.
- Staley, Kevin. 2015. “Molecular Mechanisms of Epilepsy.” *Nature Neuroscience* 18(3): 367–72.
<http://www.nature.com/doifinder/10.1038/nn.3947>.
- Stavreva, Diana A. et al. 2015. “Dynamics of Chromatin Accessibility and Long-Range Interactions in Response to Glucocorticoid Pulsing.” *Genome Research* 25(6): 845–57.
- Strauss, Kevin A. et al. 2006. “Recessive Symptomatic Focal Epilepsy and Mutant Contactin-Associated Protein-like 2.” *New England Journal of Medicine* 354(13): 1370–77.
- Visel, Axel et al. 2009. “ChIP-Seq Accurately Predicts Tissue-Specific Activity of Enhancers.”

Nature 457(7231): 854–58. <http://dx.doi.org/10.1038/nature07730>.

CHAPTER 4

Brain and spinal cord enhancers are misregulated in *C9orf72* ALS mouse models and human tissues and harbor ALS risk-associated non-coding genetic variation

Preface

This chapter contains contents from a manuscript in preparation:

Brain and spinal cord enhancers are misregulated in *C9orf72* ALS mouse models and human tissues and harbor ALS risk-associated non-coding genetic variation

Alexander J. Cammack¹, Nathan Pomper¹, Kathleen M. Schoch¹, Mark Shabsovich¹, Arnav Moudgil^{2,3,4}, Joseph D. Dougherty^{3,5}, Robi D. Mitra^{2,3}, and Timothy M. Miller¹

¹Washington University School of Medicine, Department of Neurology, St. Louis, MO

²Washington University School of Medicine, Edison Family Center for Genome Sciences and Systems Biology, St. Louis, MO

³Washington University School of Medicine, Department of Genetics, St. Louis, MO

⁴Washington University School of Medicine, Medical Scientist Training Program, St. Louis, MO

⁵Washington University School of Medicine, Department of Psychiatry, St. Louis, MO

Abstract

Large repeat expansion mutations in *chromosome 9 open reading frame 72*, or *C9orf72*, are the largest genetic cause of amyotrophic lateral sclerosis (C9ALS). While the mechanisms leading to C9ALS are unclear, recent evidence suggests that dysregulation of transcriptional and translational epigenetic pathways may be involved. Firstly, microRNAs (miRNAs), which are short regulatory RNAs that bind to and repress mRNAs, have been demonstrated to be dysregulated in both human and mouse models of ALS and be involved in ALS-related pathogenic mechanisms. Secondly, aberrant, *C9orf72* repeat-derived peptides, called dipeptide repeat proteins (DPRs) have been shown to disrupt chromatin dynamics, which could affect activity of epigenetic transcriptional regulatory elements, such as enhancers and super enhancers. Further, enhancers harbor a large portion of genetic variation associated with neurodegenerative disease, suggesting that their misregulation leads to increased disease risk. However, whether miRNAs or enhancers are disrupted in C9ALS remains undetermined. Here, we profiled miRNA expression level and enhancer activity in both C9ALS mouse models and postmortem CNS tissues and found broad patterns of dysregulation in both systems. These results implicate C9ALS-associated pathologies, such as DPRs, in directly driving alterations in epigenetic control.

Introduction

A hexanucleotide (G₄C₂) repeat expansion in *chromosome 9 open reading frame 72*, or *C9orf72*, causes 38% of familial and 6% of non-familial or singleton amyotrophic lateral sclerosis (sALS), representing the largest genetically identified subgroup of ALS to date (DeJesus-Hernandez et al. 2011; Majounie et al. 2012; Renton et al. 2011). Additionally, the same expansion can cause frontotemporal lobar degeneration (FTLD) in isolation or comorbidly with ALS (van Blitterswijk et al. 2013; Majounie et al. 2012; van der Zee et al. 2013). While this locus typically has under 30 repeats, ALS individuals with *C9orf72* expansion mutations (C9ALS) have hundreds to thousands (van Blitterswijk et al. 2013; Dols-Icardo et al. 2014; Gijssels et al. 2016; Nordin et al. 2015). In addition to the common ALS pathology of TDP-43 aggregation (Mackenzie et al. 2013), presence of this large expansion leads to the accumulation of two C9ALS-specific pathologies: nuclear RNA foci consisting of repeat-derived RNA, and dipeptide repeat proteins (DPRs), which are translated through repeat associated, non-ATG (RAN) translation (Ash et al. 2013; Gendron et al. 2013; Mori et al. 2013; Zu et al. 2013). For C9ALS G₄C₂ repeats, RAN translation yields five distinct species of DPRs: poly(GP), poly(GA), poly(PR), poly(GR), and poly(PA), each shown to aggregate in CNS tissues of C9ALS/FTLD individuals (Ash et al. 2013; Gendron et al. 2013; Mackenzie et al. 2015; Mori et al. 2013; Zu et al. 2013).

Mounting evidence in cellular and animal models suggests DPRs are toxic and may be directly involved in disease pathogenesis (Gendron and Petrucelli 2017). While the proposed mechanisms through which these aberrant peptides induce cellular dysfunction are diverse (Freibaum and Taylor 2017), recent evidence suggests that epigenetic alteration may be a key contributor. In one study, it was discovered that poly(PR) co-localizes to regions of

heterochromatin in both a transgenic poly(PR)-expressing mouse model and in C9ALS/FTLD postmortem cortical tissue and induces alterations in histone H3 posttranslational modifications (Zhang et al. 2019). Poly(PR), along with the other arginine-containing DPR, poly(GR), has been demonstrated to undergo LLPS, a dynamic process in which proteins demix from a homogenous solution to form membraneless condensates (Boeynaems et al. 2017). In overexpression models, these aberrant DPRs localize to biological condensates in the cell, such as nucleoli, and can interfere with the dynamicity of these structures, promoting liquid-to-solid transitions (Lee et al. 2016). Interestingly, one possible explanation for the observations of poly(PR) at heterochromatin is co-phase separation with heterochromatin protein 1 (HP1), a direct interactor with methylated histone tails which participates in gene silencing and has itself been shown to undergo LLPS in biological contexts (Strom et al. 2017). Indeed, in *in vitro* droplet experiments, it was found that poly(PR) induces aberrant phase transitions of HP1 to a solid, aggregated state, suggesting that this DPR alters chromatin dynamics through LLPS (Zhang et al. 2019).

Extending this one step further, it has been recently recognized that a number of other chromatin-associated proteins, such as transcription factors (TFs) and co-factors (Boija et al. 2018; Sabari et al. 2018), and even chromatin itself (Gibson et al. 2019), undergo natural phase transitions which may underlie epigenetic regulation (Hnisz et al. 2017). It follows then that presence of *C9orf72* expansion-related pathologies may further influence the genome-wide regulation of these factors as well. One such protein is the chromatin reader, BRD4, which binds to enhancer-associated acetylations on histone tails (Jung et al. 2014; Kanno et al. 2004, 2014) and is involved in the initiation of gene transcription (LeRoy, Rickards, and Flint 2008; Mochizuki et al. 2008). BRD4 is highly associated with super enhancers (Hnisz et al. 2013) and is critical for the transcriptional response to neuronal activation in the brain (Korb et al. 2015).

Additionally, multiple studies, including those from our group, suggest that epigenetic regulation at the translational level may also be altered in disease and contribute to ALS pathogenesis (Hoye et al. 2017, 2018; Koval et al. 2013; Reichenstein et al. 2019; Varcianna et al. 2019). microRNAs (miRNAs), which are short, regulatory RNAs that bind to and repress mRNAs, display altered expression in ALS mouse models (Hoye et al. 2017; Koval et al. 2013), patient-derived induced pluripotent stem cell (iPSC) models (Varcianna et al. 2019), and patient tissues (Hoye et al. 2017; Reichenstein et al. 2019). These changes modulate neuronal and glial function and have been implicated in ALS-related pathogenic mechanisms (Amin et al. 2015; Hoye et al. 2017, 2018; Koval et al. 2013; Reichenstein et al. 2019; Varcianna et al. 2019). However, whether similar alterations exist in C9ALS is unclear.

Here we profile multiple key epigenetic regulatory mechanisms in C9ALS mouse models and human postmortem tissues to investigate potential contributions to disease pathogenesis (**Table 4.1**). First, we profile miRNA levels in C9ALS transgenic mice compared to non-expanded and non-transgenic controls. Then, using an AAV calling cards method (Cammack, Moudgil, et al. 2019), we test whether *C9orf72* expansions lead to altered BRD4 binding and enhancer activity in a C9ALS mouse model. Finally, we profile genome-wide enhancer activity with H3K27ac ChIP-seq in human postmortem C9ALS, sALS, and non-ALS control lumbar spinal cord samples and test whether these human spinal cord enhancers harbor known ALS genetic risk variants.

Table 4.1

Sample type	Assay	Target	Sample size
<i>C9orf72</i> mouse line (O'Rourke et al., 2016)	Microarray	miRNA	n = 3 C9 _{exp} n = 3 C9 _{non-exp} n = 3 wt
<i>C9orf72</i> mouse line (Liu et al., 2016)	AAV calling cards	BRD4	n = 6 C9(+) n = 6 C9(-)
<i>C9orf72</i> mouse line (Liu et al., 2016)	RNA-seq	mRNA	n = 3 C9(+) n = 3 C9(-)
Human postmortem lumbar spinal cord	ChIP-seq	H3K27ac	n = 4 C9ALS n = 4 sALS n = 4 Con

Methods

Animals and genotyping

All animal practices and procedures were approved by the Washington University in St. Louis Institutional Animal Care and Use Committee (IACUC) in accordance with National Institutes of Health (NIH) guidelines. Transgenic mouse strains used in this study include Tg(C9orf72_3) line 112 (The Jackson Laboratory, 023099), hence referred to as “C9_{exp}” and littermate controls as “wt”; Tg(C9orf72_2) line 8 (The Jackson Laboratory, 023088), hence referred to as “C9_{non-exp}”; and C9-500 (The Jackson Laboratory, 029099), hence referred to as “C9(+)” and littermate controls as “C9(-).” The C9_{exp} and C9_{non-exp} lines were bred to the C57BL/6J background, while the C9(+) line was bred to the FVB/N6 background. All C9orf72 transgenic lines were genotyped with repeat-primed PCR (rpPCR) as previously described (Cammack, Atassi, et al. 2019). At indicated endpoints, mice were anesthetized with isoflurane and perfused with 15ml of cold saline (PBS). Afterwards, brains and spinal cords were dissected and flash frozen in liquid nitrogen.

miRNA extraction and microarray/RT-qPCR expression analyses

miRNAs were extracted from cervical spinal cord samples from C9_{exp}, C9_{non-exp}, and wt animals with miRNeasy kits (Qiagen), following manufacturer’s instructions. miRNAs were then reverse transcribed (Megaplex™ Primer Pools, Rodent Pools Set v3.0, ThermoFisher) and run (without preamplification) in low-density rodent miRNA A&B cards sets 3.0 (Life Technologies) on a QuantStudio 12K Flex Real-Time PCR System for 40 cycles. Three animals (2F/1M) per genotype were used for microarray assessments.

For microarray analyses, miRNA with an average raw CT count of <34 across all 9 samples were discarded, leaving 263 for downstream analysis. Each CT was normalized to the geomean of miR-24, miR-191, and miR-30c and then analyzed with the $2^{-\Delta\Delta CT}$ method. These three miRNAs were previously demonstrated to be consistently expressed across CNS cell types in spinal cord (Hoye et al. 2017).

Virus generation and stereotactic injections

A constitutively expressed AAV calling cards system was chosen for BRD4 profiling (Cammack, Moudgil, et al. 2019). An unfused hypPB construct and a self-reporting transposon (SRT) construct were separately packaged into AAV serotype 9 (AAV9) vectors by the Hope Center Viral Vectors Core at Washington University School of Medicine. For all experiments, AAV9::hypPB and AAV9::SRT viruses were mixed equivolume at maximum titer and unilaterally co-injected into 2 cortical sites per hemisphere with coordinates relative to bregma of 1.25mm rostral; 1.5mm lateral; 0.55mm depth; and 1.06mm caudal; 1.5mm lateral; 0.55mm depth. 2 μ l of viral mix was delivered at a rate of 0.2 μ l/minute.

Viral titers (viral genomes per milliliter) prior to mixing were as follows:

AAV9::hypPB = 7.1×10^{12} - 1.0×10^{13} vg/mL

AAV9::SRT = 2.2×10^{13} vg/mL

SRT library preparation and sequencing

Prior to RNA extraction, cortex samples were randomly separated into 10 pieces. RNA samples were then extracted from cortex samples using RNeasy kits (Qiagen) following the manufacturer's protocol. SRT calling card libraries were prepared from RNA samples as

previously described (Cammack, Moudgil, et al. 2019; Moudgil et al. 2019) and sequenced on the Illumina NextSeq 500 platform. After filtering and aligning reads to the mouse genome (mm10), unique insertions were defined as existing either in different TTAA's in the same RNA sample or the same TTAA but in different samples. All insertions were considered equally independent of read depth.

Significant BRD4 peak calling

BRD4 peaks significantly enriched in C9(+) over C9(-) animals (and visa versa) were identified as previously described (Cammack, Moudgil, et al. 2019; Moudgil et al. 2019) based on the differential insertion enrichment between the two profiles at genomic loci. Prior to peak calling analyses, each mouse library was downsampled to exactly 740,000 insertions. Then, mice were grouped by genotype (n=6 mice per genotype) and peaks were called based on the pooled datasets (4,440,000 insertions per genotype). Here, the significance threshold was set at $p < 10^{-7}$ at which significant peaks pass Bonferroni correction based on the number of genome blocks in each dataset (130,000-135,000 per dataset). To eliminate peaks near background levels, we filtered out any peaks with < 30 total insertions across all 12 animals assayed, with the remaining peaks being utilized in subsequent analyses. Finally, to account for within-genotype variability, unpaired Student's t-tests were run on individual differential peaks to compare C9(+) to C9(-) enrichment. As the peaks were initially defined with a Bonferroni-corrected p-value, these resultant t-test p-values were used without further correction.

Gene-enhancer pairing and gene ontology

Peak-proximal genes near each differentially enriched BRD4 calling cards peak were defined and used for “cell component” gene ontology (GO) enrichment analysis with the PANTHER v14 classification system (Mi et al. 2017). Prior to GO, genes were filtered out if greater than 10kB away from the a differentially enriched BRD4 peak.

RNA sequencing (RNA-seq) and analysis

RNA-seq libraries were prepared, sequenced, and analyzed by the Genome Technology Access Center (GTAC) at Washington University in St. Louis (St. Louis, MO, USA). The below methods were provided by GTAC.

Samples were prepared according to library kit manufacturer’s protocol, indexed, pooled, and sequenced on an Illumina HiSeq. Basecalls and demultiplexing were performed with Illumina’s bcl2fastq software and a custom python demultiplexing program with a maximum of one mismatch in the indexing read. RNA-seq reads were then aligned to the Ensembl release 76 primary assembly with STAR version 2.5.1a. Gene counts were derived from the number of uniquely aligned unambiguous reads by Subread:featureCount version 1.4.6-p5. Sequencing performance was assessed for the total number of aligned reads, total number of uniquely aligned reads, and features detected. The ribosomal fraction, known junction saturation, and read distribution over known gene models were quantified with RSeQC version 2.6.2.

All gene counts were then imported into the R/Bioconductor package EdgeR and TMM normalization size factors were calculated to adjust for samples for differences in library size. Ribosomal genes and genes not expressed in the smallest group size minus one sample greater than one count-per-million were excluded from further analysis. The TMM size factors and the

matrix of counts were then imported into the R/Bioconductor package Limma. Weighted likelihoods based on the observed mean-variance relationship of every gene and sample were then calculated for all samples with the `voomWithQualityWeights`. The performance of all genes was assessed with plots of the residual standard deviation of every gene to their average log-count with a robustly fitted trend line of the residuals.

Chromatin Immunoprecipitation sequencing (ChIP-seq) and analysis

H3K27ac ChIP-seq libraries were prepared from fresh frozen autopsy lumbar spinal cord samples were sent to ActiveMotif epigenetics services team (Carlsbad, CA, USA) for end-to-end H3K27ac ChIP-seq services. Samples included patients with either C9ALS, sALS, or no ALS (Con). Demographics and clinical information about these patients are provided in **Table 4.2**. The below methods were modified from ActiveMotif's description of their propriety analysis steps.

75-nt single-end sequence reads were generated by Illumina sequencing (using the NextSeq500 platform) and mapped to the genome using the BWA algorithm with default settings. Only reads that passed Illumina's purity filter, aligned with no more than 2 mismatches, and mapped uniquely to the genome (hg38) were used in subsequent analyses. In addition, duplicate reads ("PCR duplicates") were removed. To identify the density of reads along the genome, the genome was divided into 32-nt bins and the number of reads in each bin was determined. Final libraries were downsampled to the level of the sample in the group with the fewest usable number of reads. Peaks were called using either MACS with p-value cutoffs of 10^{-7} for narrow peaks and 10^{-1} for broad peaks; or SICER with a p-value cutoff of FDR 10^{-10} and gap

parameter of 600 bp. Peak filtering was performed by removing false ChIP-Seq peaks as defined within the ENCODE blacklist.

To compare peak metrics between samples, overlapping peaks were grouped into “merged peaks”, which are defined by the start coordinate of the most upstream peak and the end coordinate of the most downstream peak. Read density metrics at each merged peak were then normalized within sample to the mean of read counts across all peaks; thus, the average read density across all merged peaks was equivalent between samples. An unpaired Student’s test was then used to identify enriched merged peaks in C9ALS or sALS over Con samples.

The identification of super enhancers was done with a proprietary algorithm that gives a very similar result as the ROSE software. In a first step, MACS or SICER peaks generated by the standard ChIP-Seq analysis were merged if their inner distance was equal or less than 12,500 bp. Then, the merged peak regions with the strongest signals (top 5%) were identified as super enhancers.

For subsequent analyses, all enhancers with low read counts (less than 15 in all samples) were removed.

Analysis of ALS-associated risk variants in spinal cord enhancers

Nine non-coding single nucleotide polymorphisms (SNPs) were found to associated with or suggestive of increased risk of ALS in a recent genome wide association study (GWAS) (Nicolas et al. 2018). These SNP coordinates were converted from hg19 to hg38 with UCSC liftOver and compared to merged region coordinates from H3K27ac ChIP-seq postmortem lumbar spinal cord libraries (merged across all 12 samples) with BEDtools intersect.

To compare ALS-associated SNP intersections in other tissue enhancers, BED files from a publicly available dataset of H3K27ac ChIP-seq-defined enhancers from 86 different human autopsy tissues and cell lines were utilized (Hnisz et al. 2013) and enhancer coordinates were compared to SNP coordinates with BEDtools intersect. Prior to intersection analysis, enhancer coordinates were also converted to hg38 with UCSC liftOver.

Statistical analyses

Statistical tests were done with GraphPad Prism v8.1.2 and Microsoft Excel v16.16.18.

Data and code availability

Raw and processed data will be made available through GEO following publication. Calling card analysis software were previously published (Moudgil et al. 2019) and are available upon request.

Table 4.2

Autopsy ID	Genotype	Sex	Time to autopsy	Age at onset	Age at death	Co-morbidities
AU-022	C9ALS	F	18 hours	67	69	FTD
AU-060	C9ALS	F	20 hours	63	68	FTD
AU-073	C9ALS	M	33 hours	62	64	
AU-038	C9ALS	M	Unknown	75	79	
AU-064	sALS	F	20 hours	76	77	
AU-065	sALS	F	26 hours	75	76	
AU-077	sALS	M	64 hours	Unknown	76	
AU-075	sALS	M	15 hours	68	69	
AU-020	Con	F	15 hours	N/A	73	Myotonic dystrophy type II (<i>ZFN9</i> mutation)
AU-032	Con	F	20 hours	N/A	83	Renal cell disease, pneumonia
AU-078	Con	M	24 hours	N/A	47	Autonomic neuropathy, alcoholism
AU-068	Con	M	23 hours	N/A	73	Parkinsonism

Results

C9orf72 expansions induce differential miRNA expression in the mouse spinal cord

To investigate potential misregulation of miRNA expression due to *C9orf72* mutations, we profiled miRNA expression levels via microarray in cervical spinal cord samples from 14-week old transgenic mice harboring multiple copies (ranging ~100-1000 repeats) of an expanded *C9orf72* mutation (“C9_{exp}”), as well as two controls: the negative littermates of the C9_{exp} line (“wt”) and a separate transgenic line carrying a normal repeat size allele (15 repeats; “C9_{non-exp}”) (**Fig 4.1A**). In total, we examined expression levels of 381 miRNA species, of which 263 were found to be reliably expressing above background in the mouse spinal cord. Normalized miRNA expression levels were then compared between C9_{exp} animals and controls to identify differentially expressing species.

Likely due to the low sample size in these arrays (n=3 per genotype), no miRNAs were found to be significantly altered after Bonferroni multiple comparison correction. However, we did find a significant correlation between the fold change in expression of all miRNAs when comparing C9_{exp} to either the C9_{non-exp} or the wt control, suggesting that these global changes are likely not artifactual (**Fig 4.1B-D**). Interestingly, these expression alterations generally appeared to have a greater magnitude change when comparing C9_{exp} to wt than to C9_{non-exp}, suggesting that presence of even a small repeat may induce mild changes in miRNA expression (**Fig 4.1B-D**).

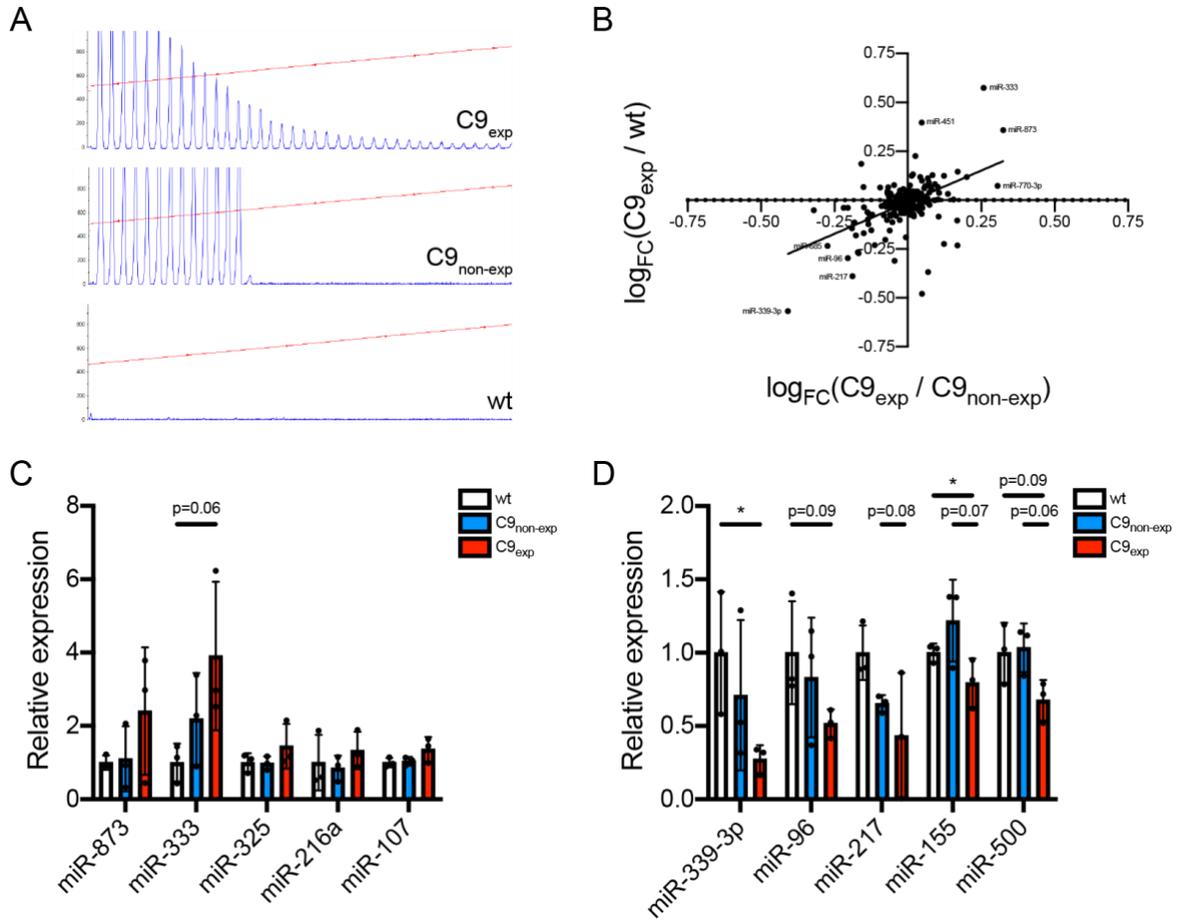


Figure 4.1: Differentially expressed miRNAs in cervical spinal cord of *C9orf72* transgenic mice. (A) rpPCR results from $C9_{exp}$, $C9_{non-exp}$, and wt animals, demonstrating their repeat expansion sizes. (B) Fold-changes in miRNA expression in $C9_{exp}$ are correlated when compared to either the $C9_{non-exp}$ or the wt controls ($R=0.51$, $p<0.0001$). (C-D) Microarray expression levels of individual miRNAs either (C) upregulated or (D) downregulated when compared to either the $C9_{non-exp}$ or the wt control. P-values generated with unpaired Student's t-tests, without multiple comparison corrections.

BRD4 profiling in the cortex of C9orf72 transgenic mice reveals altered enhancer usage

We hypothesized that *C9orf72* mutations would induce changes in enhancer activity at specific loci, especially those bound by the chromatin reader, BRD4. To test this, we assayed genome-wide occupancy of BRD4 with AAV calling cards across a month-long window from 7.5 to 8.5 months of age in the cortices of C9(+) and C9(-) transgenic animals. Briefly, this technique assesses BRD4 occupancy by tagging binding events with a marker transposon, allowing for quantification of BRD4 binding via transposon insertion density. Additionally, we performed RNA sequencing (RNA-seq) on a subset of C9(+) and C9(-) mice to assess effect of enhancer misregulation on gene expression.

By comparing insertion density between C9(+) and C9(-) littermates, we identified 61 peaks significantly enriched for BRD4 binding in C9(+) animals and 83 significantly de-enriched (**Fig 4.2A,E**). At the mRNA level, while no genes were found to be significantly altered after FDR multiple comparisons correction, we were nonetheless able to identify a number of genes with trends toward altered expression in the same direction as a nearby significantly enriched or de-enriched BRD4 peak, suggesting that misregulation of these peaks may have downstream consequences for transcription (**Fig 4.2B-D, F-H**). Finally, to investigate potential downstream pathways that could be affected by BRD4 misregulation, we performed gene ontology on gene sets proximal to differentially bound BRD4 peaks and found that the most enriched categories were involved in the synapse and synaptic membrane (**Fig 4.2I**). This is of particular interest, given the known susceptibility of C9ALS iPSC-derived neurons to excitotoxicity (Donnelly et al. 2013).

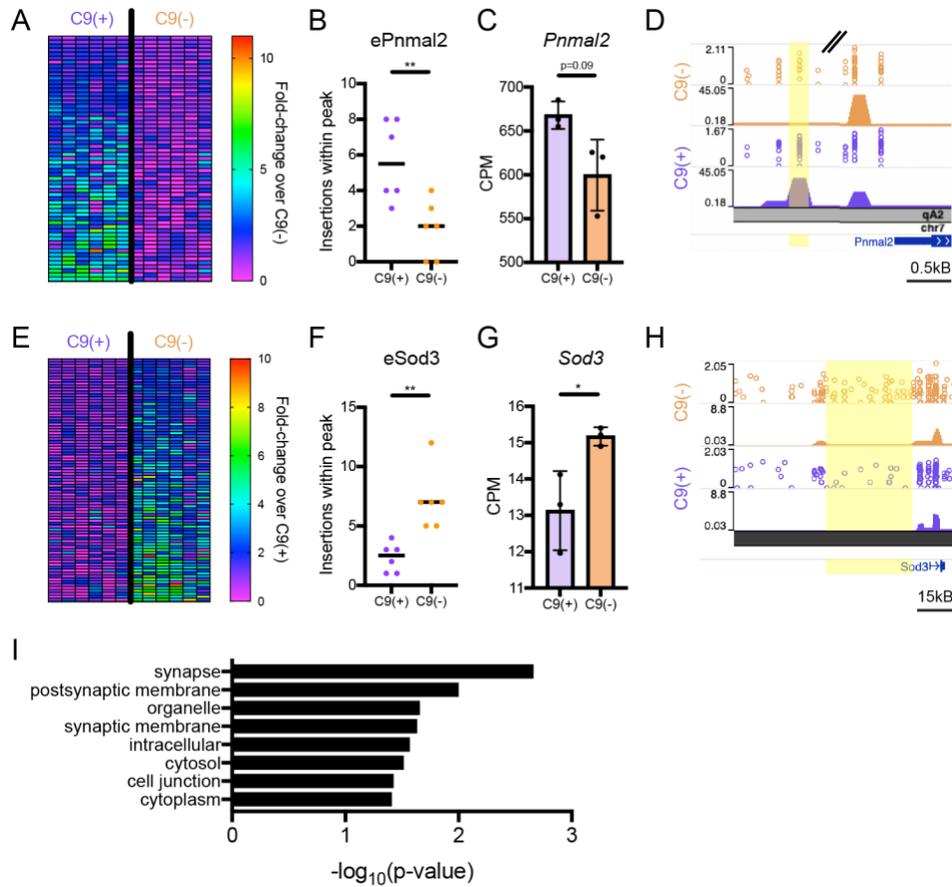


Figure 4.2: BRD4-bound enhancer profiling in *C9orf72* transgenic mice.

(A) Heatmaps of peaks (A) upregulated or (E) downregulated in *C9orf72* transgenic mice compared to non-transgenic controls (Unpaired Student's t-test, unadjusted $p < 0.05$). Example (B-D) upregulated and (F-H) downregulated REs with (B, F) quantifications of BRD4 calling card insertions within peak, (C, G) RNA-seq counts per million (CPM) of proximal genes, and (D, H) genome browser screenshots. Unpaired Student's t-test, unadjusted $*p < 0.05$, $**p < 0.01$. (I) Gene ontology (cellular component) terms of genes proximal to differentially enriched BRD4 binding sites (n = 6 mice per group).

Differential enhancer activity in human spinal cord is correlated between C9ALS and sALS

Next, we sought to investigate potential misregulation of enhancer activity in human ALS. To do so, we profiled genome-wide H3K27ac with ChIP-seq in postmortem lumbar spinal cord samples from C9ALS, sALS, and non-ALS (Con) patients. We were able to identify a number of enhancers with a pattern of up or downregulation in C9ALS compared to non-ALS (**Fig 4.3A-B**). Interestingly, the fold-change between C9ALS and Con was highly correlated with that between sALS and Con, suggesting that these changes may be representative of ALS more broadly (**Fig 4.3C-D**). Of note, the magnitude of change was on average greater in C9ALS than sALS. Finally, we separated out super enhancers, and found that these elements display a similar pattern to all enhancers as a whole (**Fig 4.3E-H**). In sum, we found global misregulation of enhancer activity in ALS spinal cords, supporting that aberrant epigenetic regulation extends to the human disease condition.

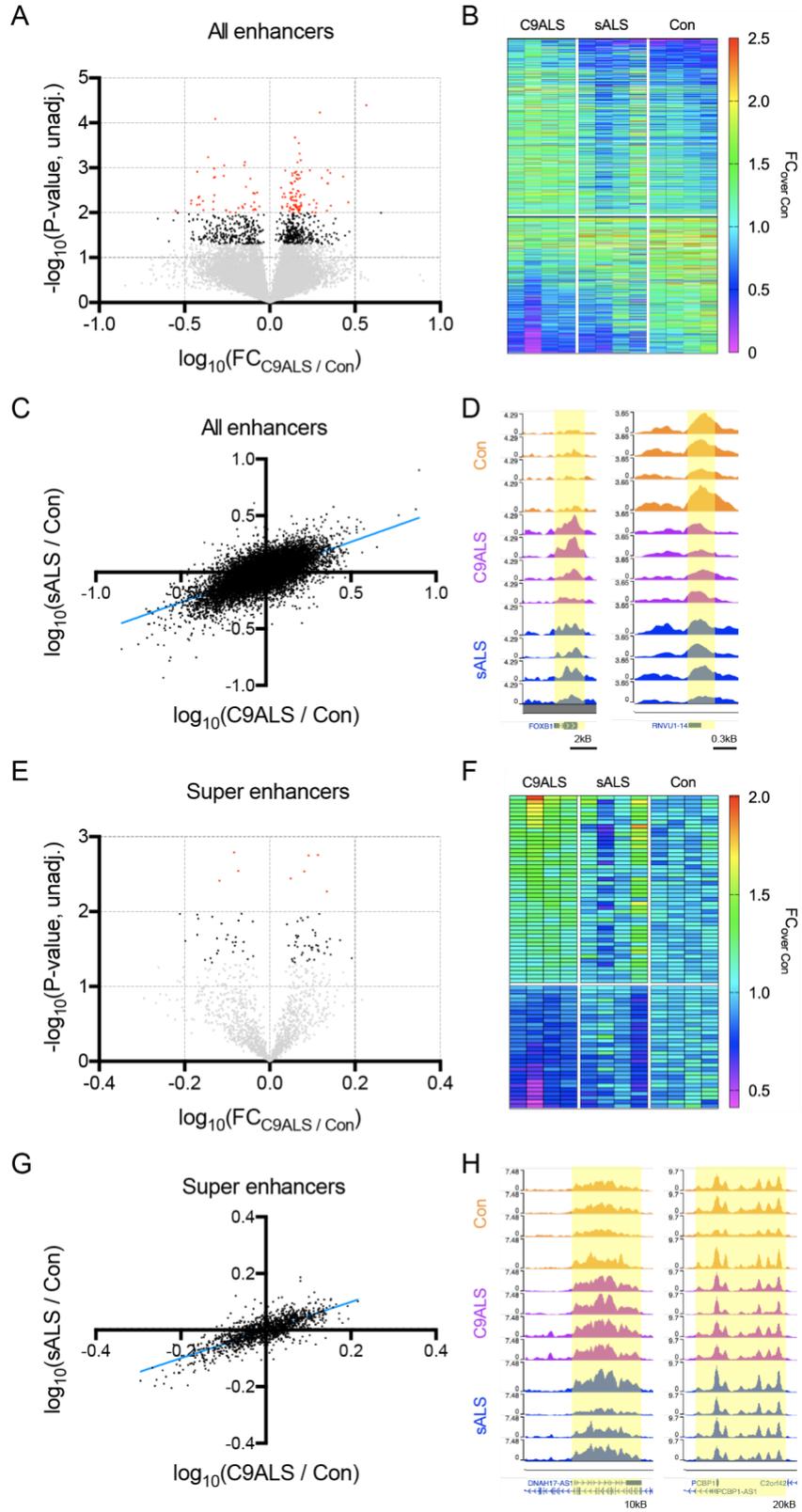


Figure 4.3: H3K27ac ChIP-seq in postmortem C9ALS, sALS, and Con lumbar spinal cord.

H3K27ac ChIP-seq enhancer activity of (A-D) all enhancers (each dot = separate enhancer; black dots: unadjusted $p < 0.05$; red dots: unadjusted $p < 0.01$) or (E-H) super enhancers profiled in postmortem lumbar spinal cords of C9ALS, sALS, and Con patients. (A, E) Volcano plots of enhancer activity at (A) all enhancers or (E) super enhancers. Black and red dots are enhancers which reached unadjusted $p < 0.05$ and $p < 0.01$, respectively (unpaired Student's t-test). (B, F) Heatmaps displaying the fold-change (FC) of normalized H3K27ac density relative to the average of the 4 Con samples across (B) all enhancers and (F) super enhancers which reached unadjusted $p < 0.05$ (unpaired Student's t-test). Heatmaps are sorted by average FC in C9ALS over Con. Rows above line break were enriched in C9ALS while below are de-enriched. (C, G) FC in C9ALS relative to Con is highly correlated with sALS over Con for (C) all enhancers (linear regression, $R = 0.68$, $p < 0.0001$) as well as (G) the super enhancer subset (linear regression, $R = 0.78$, $p < 0.0001$). (D, H) Example (D) typical enhancers and (H) super enhancers displaying (left) enrichment or (right) de-enrichment in C9ALS relative to Con. Y-axis represents normalized read depth with smoothing filter applied.

Spinal cord enhancers harbor ALS-associated non-coding genetic risk variants

Lastly, we sought to test the hypothesis that spinal cord enhancers may harbor ALS risk-associated genetic variation. To do this, we asked whether SNPs associated with increased ALS risk via GWAS could be found in enhancer regions profiled above with H3K27ac CHIP-seq. The most recent ALS GWAS found that 9 non-coding SNPs achieved or almost achieved genome wide significance, indicating that these variant alleles confer increased risk of ALS (Nicolas et al. 2018). Of note, 6/9 of these SNPs were proximal to genes which are known to cause ALS when mutated, such as *TBK1*, *C9orf72*, and *Kif5a*, and thus may be simply “marker” SNPs of disease-associated coding mutations. However, the remaining 3 SNPs do not appear near to any ALS-associated gene and remain unclear as to why they cause increased disease risk.

In our analysis, we discovered that 2 out of 3 (rs10463311 and rs75087725) of these non-marker SNPs fall within spinal cord enhancers (**Fig 4.4A-B**). The enhancer harboring rs75087725 was within the super enhancer subset. We then asked whether these enhancers were specific to the spinal cord or shared across other human tissues. To do so, we intersected these two SNPs with enhancer coordinates from a previously published dataset of H3K27ac CHIP-seq in 86 different human tissues and cell lines (Hnisz et al. 2013). Indeed, we found that these SNPs fell within active enhancers in a subset of human tissues in addition to spinal cord (**Fig 4.4C**). The super enhancer harboring rs75087725 was shared primarily with brain tissues, suggesting that this super enhancer is neural-specific. In contrast, the typical enhancer harboring rs10463311 was shared with 19 other tissues/cell lines. Interestingly, the majority (14/19) of these shared tissues were either muscle or muscle-containing tissues, but surprisingly not brain. Thus, this enhancer appears to be specific to spinal cord and muscle. Of the 86 tissues in the dataset, only

two other tissues (hippocampus and left ventricle) harbored more than 1/9 ALS risk SNPs and none harbored more than 1/3 of the non-marker ALS risk SNPs as was seen in spinal cord.

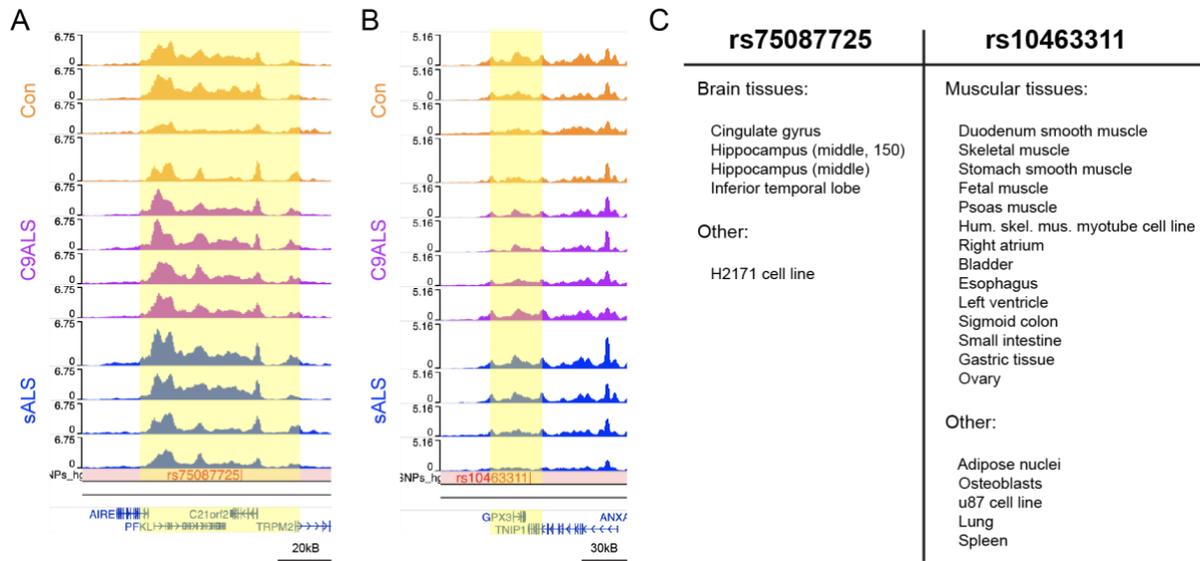


Figure 4.4: ALS risk-associated genetic variants fall in active spinal cord enhancers.

(A-B) Of the 9 non-coding SNPs found to significantly or nearly significantly increase risk of ALS in a recent GWAS (Nicolas et al. 2018), 2 fell within spinal cord enhancers. (A) rs75087725 was found to reside in a super enhancer, while (B) rs10463311 was within a typical enhancer. (C) The active enhancers harboring rs75087725 and rs10463311 were also found in other human tissues and cell lines. The tissues and cell lines with enhancers harboring rs75087725 were largely neural, while those harboring rs10463311 were mostly either muscle or muscle-containing tissues.

Discussion

In this study, we profiled miRNA expression and enhancer activity in the brains and spinal cords of C9ALS transgenic mice and postmortem C9ALS and sALS tissues and, in doing so, found widespread epigenetic misregulation at both the transcriptional and translational levels in both conditions. These results suggest that C9ALS-associated pathologies, such as DPRs, drive alterations in epigenetic control. Further, as the transgenic models used here do not exhibit overt neurodegeneration (and were studied at early timepoints), our findings support that these changes may occur prior to disease onset and are likely driven by pathological species rather than post-onset insults, such as neuroinflammation. Finally, we observed misregulated enhancer activity in sALS as well, suggesting that these alterations may extend beyond just *C9orf72*-mediated neurodegeneration.

Our finding that spinal cord enhancers harbor ALS risk-associated SNPs suggests that misregulation of enhancer activity may contribute to increased risk of neurodegeneration. Further, that enhancers of the spinal cord harbored 2/3 of the non-coding ALS SNPs not near known ALS-associated genes (Nicolas et al. 2018), while no other human tissue (out of 86 in our reference dataset) harbored more than one, suggests that misregulated enhancer activity may even contribute to the motor neuron-specific dysfunction associated with ALS. In support of this, previous studies have shown that tissue-specific enhancers often harbor disease-associated genetic variants of diseases uniquely associated with that tissue; for example, brain-specific super enhancers harbor the majority of Alzheimer's disease non-coding genetic risk, while, in contrast, super enhancers specifically active in immune cell populations harbor the risk for the autoimmune disorder, multiple sclerosis (Hnisz et al. 2013). It follows then that spinal cord enhancers would harbor risk variation for a spinal cord disorder, such as ALS. Interestingly, the

few tissues with which the ALS-associated SNP-harboring enhancers were shared were also either neural (as was the case for rs75087725) or muscular (as in rs10463311). While these observations are yet to be mechanistically understood, they add support to the idea that tissue-specific enhancer misregulation may delineate the cell or tissue type likely to experience dysfunction.

However, there are number of aspects to this study which require further analysis. It is important to note that the results described here are largely observational and correlative in nature and require functional validation in future studies to elucidate their mechanistic involvement in disease pathogenesis. Indeed, while profiling such as this is useful for uncovering general trends and pathways, individual enhancers or miRNAs (which represent potential therapeutic and mechanistic targets) cannot be confidently linked to pathogenesis without further study, for example in knockout or overexpression models. Secondly, it would be interesting and useful to directly compare RE activity between mouse and human to understand whether specific differential loci are conserved between species. Here we were able to demonstrate that general trends of epigenetic dysregulation observed in mice are recapitulated in human tissues. However, specific loci were not compared, due to difficulty of properly matching sequences between species in the less conserved, non-coding space. Our findings could be enhanced through implementation of orthogonal methodologies, such as chromatin capture, which would allow for functional pairing of REs to genes and may increase confidence that a given RE is properly matched between mouse and human. Lastly, future studies will be necessary to understand the mechanisms underlying the misregulated epigenetic control observed here. One possible explanation for the observed changes in chromatin dynamics is altered LLPS of TFs, co-factors, and/or chromatin itself induced by aberrant *C9orf72*-associated pathologies, such as DPRs.

While such mechanisms have been demonstrated for some cellular proteins (Lee et al. 2016), including the chromatin-associated enzyme, HP1 (Zhang et al. 2019), this has yet to be explored for other TFs and co-factors, such as BRD4, which was directly assayed and found to be misregulated in this study. Similarly, LLPS has been implicated as a critical mechanistic component of miRNA biogenesis (Jiang et al. 2017; Sheu-Gruttadauria and MacRae 2018). In either case, it is plausible that the presence of aberrant phase separating proteins such as DPRs could interfere with natural biological LLPS and thus underlie the functional epigenetic alterations observed here. Future biochemical and cell biological studies will help to elucidate whether these or other mechanisms contribute the epigenetic dysregulation we observed in C9ALS.

In summary, we have provided here a useful profile of epigenetic regulation in both human and mouse models of C9ALS, providing a reference resource for future studies of epigenetic control in motor neuron neurodegeneration. In doing so, we have uncovered a novel pattern of epigenetic misregulation, both transcriptionally and translationally, associated with presence of the *C9orf72* mutation. Future studies will help to understand the functional consequences of the changes observed here.

Acknowledgements

This study was funded by the Muscular Dystrophy Association, NINDS R01 NS078398, and the Hope Center Viral Vectors Core at Washington University School of Medicine. AJC was supported by 5T32GM008151-32 and the Lucille P. Markey Special Emphasis Pathway in Human Pathobiology. We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for genomic analysis. The Center is partially supported by NCI Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center and by ICTS/CTSA Grant# UL1 TR000448 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. This publication is solely the responsibility of the authors and does not necessarily represent the official view of NCRR or NIH. We thank the Edison Family Center for Genome Sciences & Systems Biology, specifically Jessica Hoisington-Lopez and MariaLynn Crosby, for assistance with genomic analysis.

Author contributions

Study designed by AJC, NP, KMS, and TMM. Data generation and analysis by AJC, NP, KMS, MS, and AM. Manuscript written by AJC.

Bibliography

- Amin, Neal D et al. 2015. “Loss of Motoneuron-Specific MicroRNA-218 Causes Systemic Neuromuscular Failure.” *Science* 350(6267): 1525–29.
- Ash, Peter E a et al. 2013. “Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to C9FTD/ALS.” *Neuron* 77(4): 639–46.
<http://dx.doi.org/10.1016/j.neuron.2013.02.004>.
- van Blitterswijk, Marka et al. 2013. “Association between Repeat Sizes and Clinical and Pathological Characteristics in Carriers of C9ORF72 Repeat Expansions (Xpansize-72): A Cross-Sectional Cohort Study.” *The Lancet Neurology* 12(10): 978–88.
[http://dx.doi.org/10.1016/S1474-4422\(13\)70210-2](http://dx.doi.org/10.1016/S1474-4422(13)70210-2).
- Boeynaems, Steven et al. 2017. “Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics.” *Molecular Cell* 65(6): 1044-1055.e5.
<http://linkinghub.elsevier.com/retrieve/pii/S1097276517301284>.
- Boija, Ann et al. 2018. “Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains.” *Cell*: 1–14.
<https://linkinghub.elsevier.com/retrieve/pii/S0092867418313989>.
- Cammack, Alexander J., Arnav Moudgil, et al. 2019. “A Viral Toolkit for Recording Transcription Factor-DNA Interactions in Live Mouse Tissues.” *BioRxiv [Preprint]* (<https://doi.org/10.1101/538504>).
- Cammack, Alexander J., Nazem Atassi, et al. 2019. “Prospective Natural History Study of C9orf72 ALS Clinical Characteristics and Biomarkers.” *Neurology* 93(17): e1605–17.
- DeJesus-Hernandez, Mariely et al. 2011. “Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS.” *Neuron*

- 72(2): 245–56. <http://dx.doi.org/10.1016/j.neuron.2011.09.011>.
- Dols-Icardo, Oriol et al. 2014. “Characterization of the Repeat Expansion Size in C9orf72 in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia.” *Human Molecular Genetics* 23(3): 749–54.
- Donnelly, Christopher J. et al. 2013. “RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention.” *Neuron* 80(2): 415–28.
<http://dx.doi.org/10.1016/j.neuron.2013.10.015>.
- Freibaum, Brian D., and J. Paul Taylor. 2017. “The Role of Dipeptide Repeats in C9ORF72-Related ALS-FTD.” *Frontiers in Molecular Neuroscience* 10(February): 1–9.
<http://journal.frontiersin.org/article/10.3389/fnmol.2017.00035/full>.
- Gendron, Tania F. et al. 2013. “Antisense Transcripts of the Expanded C9ORF72 Hexanucleotide Repeat Form Nuclear RNA Foci and Undergo Repeat-Associated Non-ATG Translation in C9FTD/ALS.” *Acta Neuropathologica* 126(6): 829–44.
- Gendron, Tania F, and Leonard Petrucelli. 2017. “Disease Mechanisms of C9ORF72 Repeat Expansions.” *Cold Spring Harbor Laboratory Press*: 1–22.
- Gibson, Bryan A et al. 2019. “Organization of Chromatin by Intrinsic and Regulated Phase Separation.” *Cell* 179(2): 470-484.e21. <http://www.ncbi.nlm.nih.gov/pubmed/31543265>.
- Gijssels, I. et al. 2016. “The C9orf72 Repeat Size Correlates with Onset Age of Disease, DNA Methylation and Transcriptional Downregulation of the Promoter.” *Molecular Psychiatry* 21(8): 1112–24.
- Hnisz, Denes et al. 2013. “Super-Enhancers in the Control of Cell Identity and Disease.” *Cell* 155(4): 934–47. <http://linkinghub.elsevier.com/retrieve/pii/S0092867413012270>.
- . 2017. “A Phase Separation Model for Transcriptional Control.” *Cell* 169: 13–23.

<http://dx.doi.org/10.1016/j.cell.2017.02.007>.

Hoye, Mariah L. et al. 2017. "MicroRNA Profiling Reveals Marker of Motor Neuron Disease in ALS Models." *The Journal of Neuroscience* 37(22): 3582–16.

<http://www.jneurosci.org/lookup/doi/10.1523/JNEUROSCI.3582-16.2017>.

———. 2018. "Motor Neuron-Derived MicroRNAs Cause Astrocyte Dysfunction in Amyotrophic Lateral Sclerosis." *Brain* 141(9): 2561–75.

Jiang, Li et al. 2017. "NEAT1 Scaffolds RNA-Binding Proteins and the Microprocessor to Globally Enhance Pri-MiRNA Processing." *Nature Structural and Molecular Biology* 24(10): 816–24.

Jung, Marie et al. 2014. "Affinity Map of Bromodomain Protein 4 (BRD4) Interactions with the Histone H4 Tail and the Small Molecule Inhibitor JQ1." *Journal of Biological Chemistry* 289(13): 9304–19.

Kanno, Tomohiko et al. 2004. "Selective Recognition of Acetylated Histones by Bromodomain Proteins Visualized in Living Cells." *Molecular Cell* 13(1): 33–43.

———. 2014. "BRD4 Assists Elongation of Both Coding and Enhancer RNAs by Interacting with Acetylated Histones." *Nature Structural and Molecular Biology* 21(12): 1047–57.
<http://dx.doi.org/10.1038/nsmb.2912>.

Korb, Erica et al. 2015. "BET Protein Brd4 Activates Transcription in Neurons and BET Inhibitor Jq1 Blocks Memory in Mice." *Nature Neuroscience* 18(10): 1464–73.

Koval, Erica D. et al. 2013. "Method for Widespread MicroRNA-155 Inhibition Prolongs Survival in ALS-Model Mice." *Human Molecular Genetics* 22(20): 4127–35.

Lee, Kyung-Ha et al. 2016. "C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles." *Cell* 167(3): 774-788.e17.

<http://linkinghub.elsevier.com/retrieve/pii/S0092867416313836>.

LeRoy, Gary, Brenden Rickards, and S. J. Flint. 2008. "The Double Bromodomain Proteins Brd2 and Brd3 Couple Histone Acetylation to Transcription." *Molecular Cell* 30(1): 51–60.

Mackenzie, Ian R. et al. 2013. "Dipeptide Repeat Protein Pathology in C9ORF72 Mutation Cases: Clinico-Pathological Correlations." *Acta Neuropathologica* 126(6): 859–79.

———. 2015. "Quantitative Analysis and Clinico-Pathological Correlations of Different Dipeptide Repeat Protein Pathologies in C9ORF72 Mutation Carriers." *Acta Neuropathologica*. <http://link.springer.com/10.1007/s00401-015-1476-2>.

Majounie, Elisa et al. 2012. "Frequency of the C9orf72 Hexanucleotide Repeat Expansion in Patients with Amyotrophic Lateral Sclerosis and Frontotemporal Dementia: A Cross-Sectional Study." *Lancet Neurology* 11: 323–30.

[http://www.thelancet.com/pdfs/journals/lanneur/PIIS1474-4422\(12\)70043-1.pdf](http://www.thelancet.com/pdfs/journals/lanneur/PIIS1474-4422(12)70043-1.pdf) (May 3, 2015).

Mi, Huaiyu et al. 2017. "PANTHER Version 11: Expanded Annotation Data from Gene Ontology and Reactome Pathways, and Data Analysis Tool Enhancements." *Nucleic Acids Research* 45(D1): D183–89.

Mochizuki, Kazuki et al. 2008. "The Bromodomain Protein Brd4 Stimulates G1 Gene Transcription and Promotes Progression to S Phase." *Journal of Biological Chemistry* 283(14): 9040–48.

Mori, Kohji et al. 2013. "The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS." *Science* 339(March): 1335–39.

<http://www.sciencemag.org/content/suppl/2013/02/06/science.1232927.DC1.full> (May 7, 2015).

- Moudgil, Arnav et al. 2019. “Self-Reporting Transposons Enable Simultaneous Readout of Gene Expression and Transcription Factor Binding in Single Cells.” *BioRxiv [Preprint]* (<https://doi.org/10.1101/538553>).
- Nicolas, Aude et al. 2018. “Genome-Wide Analyses Identify KIF5A as a Novel ALS Gene.” *Neuron* 97(6): 1268-1283.e6. [http://www.cell.com/neuron/fulltext/S0896-6273\(18\)30148-X](http://www.cell.com/neuron/fulltext/S0896-6273(18)30148-X).
- Nordin, A. et al. 2015. “Extensive Size Variability of the GGGGCC Expansion in C9orf72 in Both Neuronal and Non-Neuronal Tissues in 18 Patients with ALS or FTD.” *Human Molecular Genetics* 24(February): 1–10. <http://www.hmg.oxfordjournals.org/cgi/doi/10.1093/hmg/ddv064>.
- Reichenstein, Irit et al. 2019. “Human Genetics and Neuropathology Suggest a Link between MiR-218 and Amyotrophic Lateral Sclerosis Pathophysiology.” *Science Translational Medicine* 11(eaav5264; 18 December): 1–12.
- Renton, Alan E. et al. 2011. “A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD.” *Neuron* 72(2): 257–68.
- Sabari, Benjamin R. et al. 2018. “Coactivator Condensation at Super-Enhancers Links Phase Separation and Gene Control.” *Science* 361(6400): eaar3958. <http://science.sciencemag.org/>.
- Sheu-Gruttadauria, Jessica, and Ian J. MacRae. 2018. “Phase Transitions in the Assembly and Function of Human MiRISC.” *Cell* 173(4): 946-957.e16. <https://doi.org/10.1016/j.cell.2018.02.051>.
- Strom, Amy R. et al. 2017. “Phase Separation Drives Heterochromatin Domain Formation.” *Nature* 547(7662): 241–45. <http://dx.doi.org/10.1038/nature22989>.

- Varcianna, André et al. 2019. “Micro-RNAs Secreted through Astrocyte-Derived Extracellular Vesicles Cause Neuronal Network Degeneration in C9orf72 ALS.” *EBioMedicine* 40: 626–35.
- van der Zee, Julie et al. 2013. “A Pan-European Study of the C9orf72 Repeat Associated with FTLD: Geographic Prevalence, Genomic Instability, and Intermediate Repeats.” *Human Mutation* 34(2): 363–73.
- Zhang, Yong-Jie et al. 2019. “Heterochromatin Anomalies and Double-Stranded RNA Accumulation Underlie C9orf72 Poly(PR) Toxicity.” *Science* 363(6428): eaav2606.
<http://science.sciencemag.org/content/363/6428/eaav2606/tab-pdf>.
- Zu, Tao et al. 2013. “RAN Proteins and RNA Foci from Antisense Transcripts in C9ORF72 ALS and Frontotemporal Dementia.” *Proceedings of the National Academy of Sciences of the United States of America* 110(51): E4968-77.
<http://www.pnas.org/content/110/51/E4968.full> (May 3, 2015).

CHAPTER 5

Prospective natural history study of *C9orf72* ALS clinical characteristics and biomarkers

Preface

This chapter contains contents from a previously published manuscript:

Prospective natural history study of *C9orf72* ALS clinical characteristics and biomarkers

Alexander J. Cammack, B.S.^{1,*}, Nazem Atassi, M.D.^{2,*}, Theodore Hyman, B.S.¹, Leonard H. van den Berg, M.D., Ph.D.³, Matthew Harms, M.D.⁴, Robert H. Baloh, M.D., Ph.D.⁵, Robert H. Brown, M.D., Ph.D.⁶, Michael A. van Es, M.D., Ph.D.³, Jan H. Veldink, M.D., Ph.D.³, Balint S. de Vries, M.D.³, Jeffrey D. Rothstein, M.D., Ph.D.⁷, Caroline Drain, M.H.S.¹, Jennifer Jockel-Balsarotti, B.S.¹, Amber Malcolm, M.S.N., R.N.¹, Sonia Boodram, B.S.¹, Amber Salter, Ph.D., M.P.H.¹, Nicholas Wightman, B.A.⁶, Hong Yu, M.S.², Alexander V. Sherman, M.Sc.², Thomas J. Esparza, B.S.¹, Diane McKenna-Yasek, B.S.⁶, Margaret A. Owegi, D.O.⁶, Catherine Douthwright, Ph.D.⁶, Alzheimer's Disease Neuroimaging Initiative (ADNI)[#], Alexander McCampbell, Ph.D.⁸, Toby Ferguson, M.D., Ph.D.⁸, Carlos Cruchaga, Ph.D.¹, Merit Cudkovicz, M.D., M.Sc.², Timothy M. Miller, M.D., Ph.D.¹

¹Washington University School of Medicine, Department of Neurology, St. Louis, MO

²Massachusetts General Hospital, Neurological Clinical Research Institute, Department of Neurology, Boston, MA

³Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, University Utrecht, Utrecht, The Netherlands

⁴Columbia University, Department of Neurology, New York, NY

⁵Cedars-Sinai Medical Center, Department of Neurology, Los Angeles, CA

⁶University of Massachusetts, Department of Neurology, Worcester, MA

⁷Johns Hopkins University, Department of Neurology, Baltimore, MD

⁸Biogen Inc., Boston, MA

Abstract

Objective:

To define the natural history of the *C9orf72* amyotrophic lateral sclerosis (C9ALS) patient population, develop disease biomarkers, and characterize patient pathologies.

Methods:

We prospectively collected clinical and demographic data from 116 symptomatic C9ALS and 12 non-ALS full expansion carriers across seven institutions in the United States and The Netherlands. In addition, we collected blood samples for DNA repeat size assessment, cerebrospinal fluid (CSF) samples for biomarker identification, and autopsy samples for dipeptide repeat protein (DPR) size determination. Finally, we collected retrospective clinical data via chart review from 208 C9ALS and 450 singleton ALS (SALS) individuals.

Results:

The mean age at onset in the symptomatic prospective cohort was 57.9 ± 8.3 years, and median duration of survival after onset was 36.9 months. The monthly change was -1.8 ± 1.7 for ALSFRS-R and $-1.4\% \pm 3.24\%$ of predicted for SVC. In blood DNA, we found that G₄C₂ repeat size correlates positively with age. In CSF, we observed that concentrations of poly(GP) negatively correlate with DNA expansion size but do not correlate with measures of disease progression. Finally, we found that size of poly(GP) dipeptides in the brain can reach large sizes similar to that of their DNA repeat derivatives.

Conclusions:

We present a thorough investigation of C9ALS natural history, providing the basis for C9ALS clinical trial design. We found that clinical features of this genetic subset are less variant than in SALS. Additionally, we identified important correlations of C9ALS patient pathologies with clinical and demographic data.

Introduction

The *C9orf72* repeat expansion mutation is the most common genetic cause of both familial ALS (fALS) and sporadic, or singleton, ALS (sALS) (DeJesus-Hernandez et al. 2011; Majounie et al. 2012; Renton et al. 2011). This mutation likely induces a gain of function toxicity, either through one or a combination of two repeat-associated pathologies, each arising following transcription of the repeat in both the sense and antisense directions: repeat-derived RNA foci (Donnelly et al. 2013; Lee et al. 2013) and dipeptide repeat proteins (DPRs) formed through repeat associated, non-ATG (RAN) translation (Ash et al. 2013; Gendron et al. 2013; Mori et al. 2013; Zu et al. 2013). While the proposed molecular mechanisms by which RNA foci and DPRs may disrupt cellular biology are numerous and diverse, it is widely accepted that, at least in cellular and animal overexpression models, these pathologies can cause a remarkable amount of neuronal toxicity (Freibaum and Taylor 2017).

However, despite the large amount of work demonstrating *C9orf72* pathology-related toxicities and mechanisms in model systems, there remains a lack of understanding of their characteristics *in vivo*. Because DPRs can be translated without a traditional start codon, it is unknown how large DPRs become in human ALS. Importantly, recent work in the C9ALS field has demonstrated a direct relationship between DPR size and toxicity. One study showed that survival of primary cortical neurons decreases as poly(GR) repeat length increases (Wen et al. 2014), while another demonstrated a similar survival decrease in *Drosophila* overexpressing poly(GR) and poly(PR) with increasing size (Mizielinska et al. 2014). Further, using much larger DPRs (>1000 repeats) in HeLa cells, one group found that electrophysiological deficits worsen with increasing DPR size (Callister et al. 2016). However, largely due to difficulties in creating and manipulating large repeats, the majority of studies focus on repeats much shorter than the

thousands that are present in the human DNA expansion mutation. Thus, understanding DPR size and its relationship to that of its DNA counterpart in the human context is critical for understanding human disease.

In addition to investigating these pathological characteristics, it is important to facilitate the transition of C9ALS-directed therapeutics from the bench to the clinic. Drugs targeted at reducing or eliminating RNA foci and/or DPRs are in development, with multiple rapidly approaching human trials (Donnelly et al. 2013; Jiang et al. 2016; Lagier-Tourenne et al. 2013; Nguyen et al. 2019; Sareen et al. 2013; Zhou et al. 2017, 2019), the design of which will be determined by the natural history of the C9ALS patient population. Thus, a detailed description of the C9ALS population is necessary for the facilitation of C9ALS clinical trials.

Here we present a prospective natural history study of C9ALS individuals in the United States and The Netherlands. We documented a wide array of demographic and clinical data in a cohort of 116 symptomatic and 12 non-ALS full expansion carriers. Additionally, we collected blood and cerebrospinal fluid (CSF) samples from these individuals for analysis of DNA repeat size and C9ALS biomarkers, respectively. As the expansion in *C9orf72* is somatically unstable (Nordin et al. 2015), we examined potential relationships between repeat size and clinical and demographic data. Previous studies suggest that CSF poly(GP) levels, while not strongly correlated with disease characteristics, are steady over time and responsive to C9ALS therapeutics, providing promise for use as a pharmacodynamics biomarker (Gendron et al. 2017; Lehmer et al. 2017; Su et al. 2014). Thus, we examined poly(GP) levels in CSF and correlated these measurements with DNA repeat expansion size and clinical characteristics. Finally, we investigated the size of poly(GP) dipeptides in postmortem CNS tissues.

Methods

Participant identification and enrollment

Participants for the prospective natural history study were enrolled at seven institutions: Washington University (WU), Massachusetts General Hospital (MGH), University Medical Center Utrecht (Utrecht), University of Massachusetts (UM), Columbia University Medical Center (CUMC), Cedars-Sinai Medical Center (CSMC), and Johns Hopkins University (JHU). Data management was handled by the Neurologic Clinical Research Institute (NCRI) at MGH. The C9ALS subgroup (n = 116) required a minimum diagnosis of possible ALS based on the most recent revision of the El-Escorial Criteria, while a separate “non-ALS” subgroup (n = 12) included individuals without ALS-related motor symptoms. Of the 12 non-ALS participants, all received cognitive assessment with the ALS Cognitive Behavioral Screen (ALS-CBS™). *C9orf72* expansions of greater than 50 repeats were confirmed for all participants via Clinical Laboratory Improvement Amendments (CLIA)-approved testing provided by Prevention Genetics (Marshfield, WI, USA).

A separate cohort of 208 C9ALS and 450 SALS individuals from WU and Utrecht was assessed retrospectively for *C9orf72* natural history, dating back to 2006. All individuals were first delineated by the presence or absence of a *C9orf72* expansion mutation, confirmed via in-house genetic testing. Those without an expansion were then screened for the presence or absence of a family history of ALS, and those with a family history were further excluded to reduce potential presence of other ALS-causing mutations (Turner et al. 2017). The remaining individuals were categorized as SALS.

Standard protocol approvals, registrations, and patient consents

This study was approved by each recruiting center's Institutional Review Board and written informed consent was provided by all participants.

Data and patient sample collection

Clinical data for the prospective study were collected in person or via telephone interviews with individuals and/or caregivers, with the exception of slow vital capacity (SVC), which was assessed via handheld electronic spirometer. Clinical assessments included the ALS Functional Rating Scale-Revised (ALSFRS-R), ALS-CBS, and ALS Caregiver Behavioral Questionnaire (ALSCBQ). Severe cognitive impairment consistent with a diagnosis of FTLD was defined as an initial ALS-CBS score of 10 or less out of a total score of 20, and moderate cognitive impairment as 11-15. Family history of ALS, dementia, or both was assessed with criteria defined in Byrne et al. (Byrne et al. 2011) Additional data included family and personal medical history, medication use, ALS-related history (onset site, onset date, timeline of symptom progression), physical and neurological examination, vital signs, and demographics. Biological specimens from prospective study participants were also collected, including whole blood, serum, CSF, and urine. Participants were enrolled over a 44-month period and data were collected until study closing. Additionally, participants were followed for survival outcomes after study closing, up until the date of this report. Study site personnel were trained on good clinical practices and study outcome measures. Data were collected by the sites and recorded on source documents, then subsequently captured in the NeuroBANK™ patient-centric platform (<http://neurobank.org/>) and monitored remotely for consistency and completeness.

Longitudinal data were acquired through follow up visits or phone calls. In total, we were able to acquire at least two longitudinal measures from 88 participants for ALSFRS-R, 53 for SVC, 42 for ALS-CBS, and 29 for ALSCBQ. The average time (and range) between measures was 2.6 months (0.6-14.5) for ALSFRS-R, 7.6 (0.7-26.6) for SVC, 7.8 (0.7-26.6) for ALS-CBS, and 3.9 (0.6-18.1) for ALSCBQ. 21 participants in the prospective C9ALS subgroup reached mortality within 6 months of study enrollment and 12 more between 6 and 12 months. Additionally, 21 C9ALS participants enrolled less than 6 months before study closing. Of the 42 subjects enrolled in the study for 6 months or less, 16 were still able to provide longitudinal data for ALSFRS-R, 12 for SVC, 10 for ALS-CBS, and 7 for ALSCBQ. Other reasons for lack of follow up are mixed and include inability to perform tasks due to disease progression, time constraints during clinic visits, and inability to travel to the study site.

In the non-ALS subgroup (n = 12 individuals), longitudinal ALS-CBS measures were available for 7 individuals, with an average (and range) of 9.2 months (5.3-18.2) between measures. No non-ALS individuals reached mortality during the study. 5 non-ALS participants were enrolled within 9 months of study closure, the average time between measures in this group, 4 of which did not provide longitudinal data. One other individual conducted their follow up remotely via phone and thus was unable to provide a longitudinal ALS-CBS assessment.

Clinical data for the retrospective cohort were collected through chart review and recorded to a central database.

Blood-derived DNA samples from unexpanded individuals were from the Knight Alzheimer's Disease Research Center (ADRC) at WU and the ADNI database (www.loni.ucla.edu/ADNI; for up-to-date information see www.adni-info.org). As there is no association between repeat size and dementia of the Alzheimer's type (DAT) status (Harms et al.

2013), all individuals were grouped together for analysis. This study was approved by each recruiting center's Institutional Review Board and was carried out in accordance with the approved protocol.

Autopsy samples used in poly(GP) size assessments were from a separate cohort of C9ALS individuals from WU. All samples were collected between 6-32 hours of death, flash frozen in liquid nitrogen, and stored at -80°C until use.

DNA repeat size measurement

Presence or absence of an expanded repeat was assessed with repeat primed PCR (rpPCR) as previously described (Harms et al. 2013) using published primer concentrations and sequences (DeJesus-Hernandez et al. 2011). PCR products were analyzed with an ABI® 3130xl Genetic Analyzer (Applied Biosystems) and GeneMapper® software. rpPCR was also used to quantify repeat sizes in unexpanded individuals.

G₄C₂ repeat size assessment of full expansions was performed with Southern blot. DNA was isolated from whole blood samples by a core facility at WU and sent to Dr. Robert H. Brown's lab for analysis. DNA was digested with AluI and DdeI, followed by gel electrophoresis and probing with a (GGGGCC)₅-DIG probe. Bands were visualized using an anti-DIG antibody and a chemiluminescent protocol. For quantification, densitometry plots were generated for each lane using GelEval software (FrogDance Software; ver. 1.37). Density peaks for each ladder band were used to create a standard curve, from which sample density peak sizes were interpolated (GraphPad Prism; ver. 7.0). Due to the upper limit of the ladder, this method could not accurately distinguish expansion sizes greater than 3855 repeats. Sixteen samples were measured twice, with a median difference of 318 repeats between measurements. All Southern

blot size assessments were blinded to eliminate experimenter bias. Prior to unblinding, samples with technical abnormalities (e.g. gel band too faint, presence of non-specific artifact, etc.) were excluded. In total, samples from 100 C9ALS and non-ALS individuals were analyzed, with 11 being excluded, leaving 89 for further analysis.

Haplotype analysis

Unexpanded individuals were genotyped for a 24 single nucleotide polymorphism (SNP) “at-risk” haplotype associated with expanded *C9orf72* repeats (Mok et al. 2012) using an Illumina 610 array. Stringent quality control criteria was applied to remove low-quality SNPs (Wijsman et al. 2011). The SNPs reported previously as part of the risk haplotype for *C9orf72* (Mok et al. 2012) were extracted and haplotype analysis was performed using PLINK. Haplotype carrier status was defined as individuals with at least 21/24 matching SNPs.

Poly(GP) measurement and size assessment

Relative concentrations of poly(GP) were measured via immunoassay. A monoclonal mouse anti-poly(GP) antibody (Biogen, Boston, MA, USA) was incubated on goat anti-mouse plates (Meso Scale Diagnostics, Rockville, MD, USA) for capture. Consecutive incubations with a polyclonal rabbit anti-poly(GP) antibody (Biogen, Boston, MA, USA) followed by a sulfo-tagged goat anti-rabbit antibody (Meso Scale Diagnostics, Rockville, MD, USA) were used for detection. “Relative concentration” designations refer to the amount of poly(GP) signal present in an equivalent amount of DPR-containing tissue or cell lysate. Thus, axes display ng/ml of total standard lysate rather than absolute concentration of poly(GP).

For assessment of poly(GP) in CSF, undiluted samples were measured in triplicate (45µl/well). Any sample displaying less than twice the average signal in blank wells was considered a zero value and all non-zero samples achieved a coefficient of variation of less than 15.

For poly(GP) size assessment, samples were first separated with size exclusion chromatography (SEC) and then quantified with the poly(GP) immunoassay. Autopsy samples were homogenized in TEN buffer (10mM Tris pH 8.0, 1mM EDTA, 100mM NaCl) with 2% sodium dodecyl sulfate (SDS) and fresh protease inhibitors (Sigma Aldrich, St. Louis, MO, USA), sonicated briefly, and centrifuged at 100,000g for 30 minutes at 4°C to remove insoluble protein species. Resultant supernatants were normalized by volume, loaded into a Superdex 200 10/300 GL SEC column (GE Healthcare, Chicago, IL, USA), and separated into 1 mL fractions at a flow rate of 0.75 mL/minute in TEN buffer + 0.5% SDS. SEC fractions were measured with the poly(GP) immunoassay in triplicate.

Statistical analyses

Statistical analyses were performed in GraphPad Prism ver. 7.0. Central tendency markers represent mean, while error bars represent standard deviation. All correlations were analyzed with linear regression. All reported p-values were corrected for multiple comparisons across the entire study (False discovery rate; FDR).

Survival analyses for the prospective C9ALS cohort included 88 individuals with definitive survival endpoints, while the remaining 28 participants that were either alive or lost to follow up were right-censored on last known date of contact. Censored datapoints are marked

with tick marks in **Fig 5.1A**. Survival comparisons of retrospective C9ALS and SALS were done with Long-rank (Mantel-Cox) test.

Data availability

All data from this study are stored in the NeuroBANK™ (<http://neurobank.org/>) data repository at the MGH NCRI and are linked with biospecimen repositories. We will share deidentified datasets with researchers who want to advance understanding of neurological disease. A limited amount of deidentified biofluid samples (DNA, serum, PBMCs, urine and CSF) collected from this study are stored at the Northeast ALS (NEALS) Biorepository (<https://www.neals.org/for-als-researchers/neals-sample-repository/>) and are publicly available to researchers around the world. All requests for data and biofluid samples will go through the NEALS Sample Repository through an application request system for qualified researchers. All biofluid samples will be available until they are depleted. Data will be made public one-year post-publication.

Results

C9ALS natural history: Prospective disease onset and survival

Data were collected prospectively from 116 symptomatic C9ALS and 12 non-ALS full expansion carriers. Of the symptomatic individuals, 53 were male and 63 female. Average age at disease onset was 57.9 ± 8.3 years (**Fig 5.1A; Table 5.1**). At the time of this report, 88 individuals had reached a survival endpoint (either mortality or initiation of invasive ventilation) and 28 were either still alive or lost to follow up. Median survival in this population was 36.9 months after disease onset (**Fig 5.1B; Table 5.1**). Site of disease onset and patient-reported family histories (following criteria defined by Byrne et al. (Byrne et al. 2011)) are described in **Fig 5.1C-D** and **Table 5.1**.

C9ALS natural history: Retrospective disease onset and survival

In addition to the prospective study, we collected retrospective survival and age at onset information from a separate cohort of 208 C9ALS and 450 SALS individuals. Mean age at onset from the retrospective cohort was similar to the prospective cohort: 59.0 ± 9.3 years for C9ALS and 59.3 ± 12.1 years for SALS (**Fig 5.1E**). Median survival was 29.9 and 30 months for C9ALS and SALS, respectively; however, notably, C9ALS had a significantly smaller fraction of slow-progressing individuals (**Fig 5.1F**).

C9ALS natural history: Prospective ALSFRS-R, SVC, ALS-CBS, and ALSCBQ

Longitudinal data for C9ALS individuals were available for ALSFRS-R (n = 88), SVC (n = 53), ALS-CBS (n = 42), and ALSCBQ (n = 29). Additionally, baseline ALS-CBS data were collected from 101 C9ALS participants. We observed an average monthly decline in ALSFRS-R

Table 5.1

Clinical values	Prospective cohort
Sex (% male)	45.7%
Average age at onset	57.9 ± 8.3 years
Median survival	36.9 months
Onset location	
<i>Limb</i>	53.8% (63/116)
<i>Bulbar</i>	38.5% (45/116)
<i>Respiratory</i>	4.3% (5/116)
<i>Cognitive</i>	1.7% (2/116)
<i>Multiple sites</i>	0.9% (1/116)
<i>Insufficient data</i>	0.9% (1/116)
Family history (<i>criteria from Byrne et al.</i>)	
<i>“definite” ALS only</i>	10.1% (13/128)
<i>“probable” ALS only</i>	19.5% (25/128)
<i>“definite” dementia only</i>	4.7% (6/128)
<i>“probable” dementia only</i>	23.4% (30/128)
<i>“definite” ALS and dementia</i>	17.2% (22/128)
<i>“probable” ALS and dementia</i>	10.9% (14/128)
<i>No known family history</i>	14.1% (18/128)
Cognitive impairment (ALS-CBS)	
<i>Cognitively normal (score = 16-20)</i>	47.5% (48/101)
<i>Moderate impairment (score = 11-15)</i>	32.7% (33/101)
<i>Severe impairment (score < 11)</i>	19.8% (20/101)

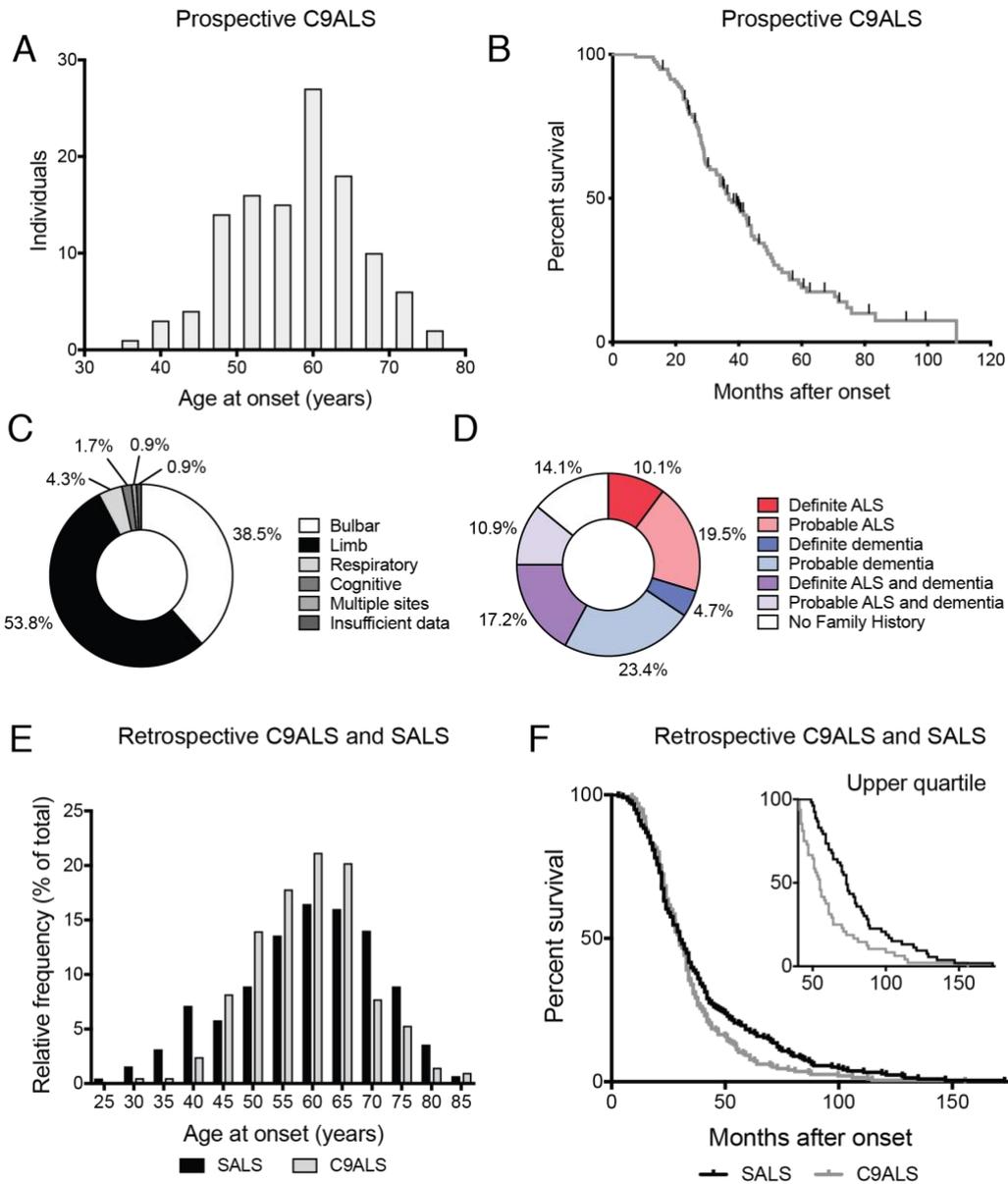


Figure 5.1. C9ALS natural history: descriptive characteristics.

(A) Distribution of ages at onset for the prospective C9ALS cohort (n = 116 individuals; mean age at onset = 57.9 ± 8.3 years). (B) Mortality in the prospective C9ALS cohort. (n = 88 individuals reaching survival endpoint; n = 28 individuals alive or lost to follow up (right-censored on last known date of contact; marked with tick marks); median survival = 36.9 months). (C) Distribution of onset locations in the prospective C9ALS cohort. (D) Family history of ALS, dementia, or both in the prospective cohort (n = 128 individuals). (E) Distribution of ages at onset for retrospective C9ALS (n = 208 individuals; mean age at onset = 59.0 ± 9.3 years) and SALS (n = 450 individuals; mean age at onset = 59.3 ± 12.1 years) cohorts. (F) Mortality in retrospective C9ALS (n = 195 individuals; median survival = 29.9 months) and SALS cohorts (n = 214 individuals; median survival = 30 months). Upper quartile survival displays significantly decreased disease duration in slow-progressing C9ALS individuals (n = 48 individuals; median survival = 55 months) as compared to SALS individuals (n = 53 individuals; median survival = 73 months) (Long-rank (Mantel-Cox) test: $p < 0.001$).

of -1.8 ± 1.7 and SVC of $-1.4\% \pm 3.24\%$ (**Fig 5.2A-B**). At first visit, 47.5% (48/101) individuals with available data presented as cognitively normal, 32.7% (33/101) as moderately impaired, and 19.8% (20/101) as severely impaired (**Table 5.1**). We observed little change in ALSCBS or ALSCBQ total scores (**Fig 5.2C-D**) or individual subscores (**Fig 5.2E-L**) over time in this cohort. Additionally, we obtained baseline ALS-CBS data for all 12 participants in the non-ALS cohort. Of these, 6/12 presented as cognitively normal, 3/12 as moderately impaired, and 3/12 as severely impaired. During the course of this study, no change in ALS-CBS impairment status was observed in the 7 non-ALS individuals for whom longitudinal measures were available.

Relationship of expansion size and age in prospective cohort

We probed G₄C₂ repeat size in blood DNA from 89 individuals in the prospective C9ALS and non-ALS cohorts. The average repeat size was 2789 ± 757 repeats, noting that 10 individuals' repeat sizes were within 318 repeats of the upper limit of detection. We observed a significant positive correlation of age at disease onset and repeat size (**Fig 5.3A**). As expected, we also found a positive correlation between repeat size and age at sample collection (**Fig 5.3B-C**). We observed no significant correlations between repeat size and survival after onset or ALSFRS-R rate of decline (**Fig 5.3D-E**).

Relationship of expansion size and age in unexpanded individuals

We hypothesized that repeat size and age correlations we observed in the prospective cohort may represent dynamic G₄C₂ expansion over time. We extended our analyses to blood-derived DNA samples from a large (n = 674) cohort of individuals with unexpanded repeat sizes (<30 repeats) from WU and ADNI. Within this population, 36% (244/674) of individuals were

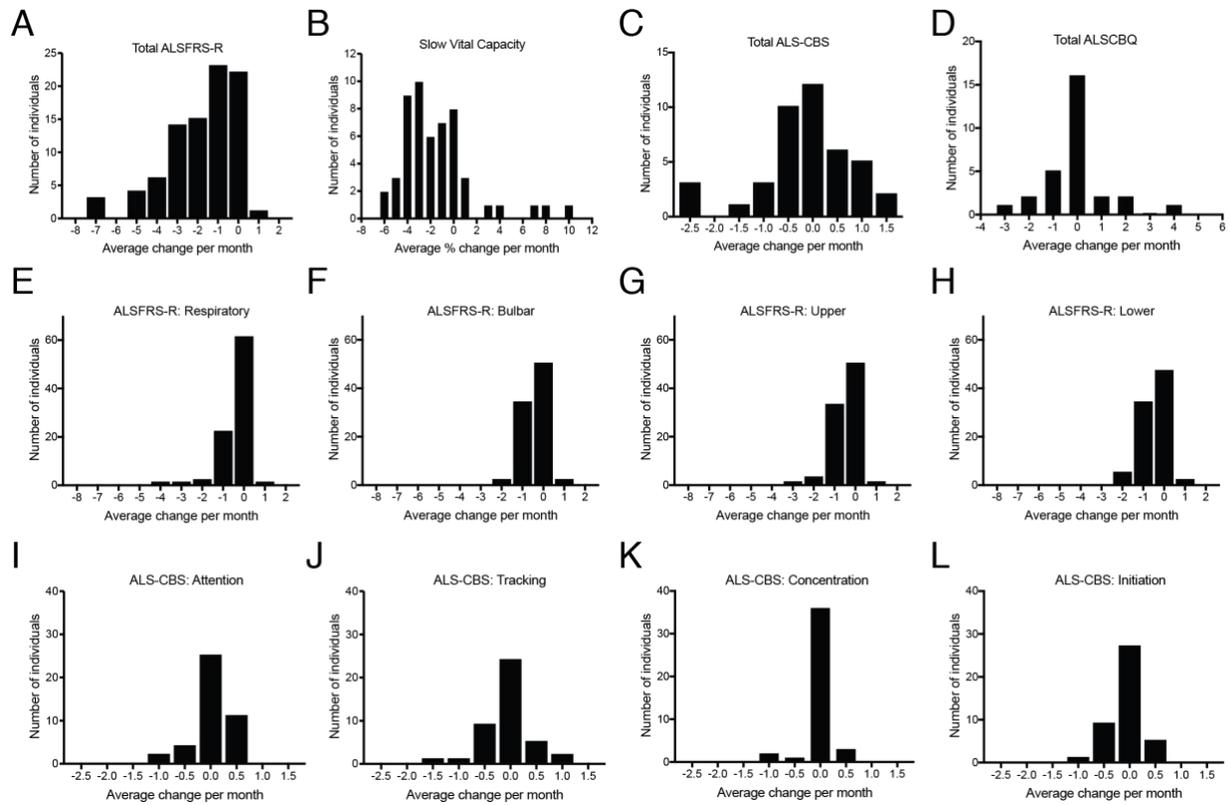


Figure 5.2. C9ALS natural history: measures of disease progression.

(A) ALS Functional Rating Scale (ALSFRRS-R), (B) Slow vital capacity (SVC), (C) ALS Cognitive Behavioral Screen (ALS-CBS), and (D) ALS Caregiver Behavioral Questionnaire (ALSCBQ) rates of decline for C9ALS individuals with at least 2 longitudinal datapoints. (ALSFRRS-R: n = 88; SVC: n = 53; ALS-CBS: n = 42; ALSCBQ: n = 29). (E-H) ALSFRS-R (n = 88) and (I-L) ALS-CBS (n = 42) subscore rates of change.

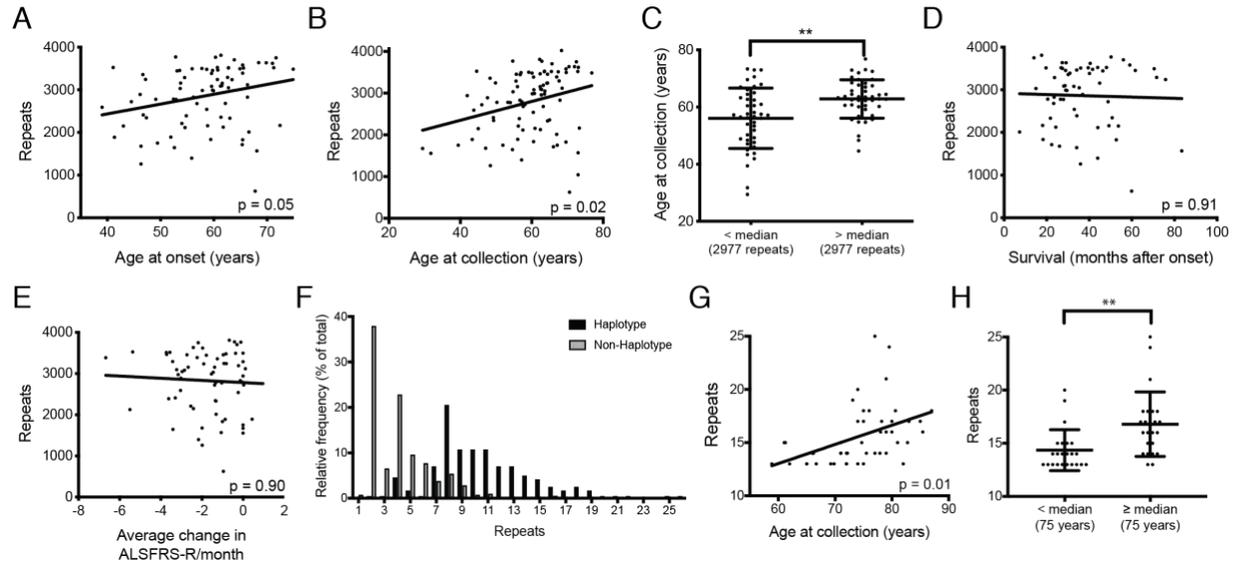


Figure 5.3. Relationship of G₄C₂ repeat size and age.

G₄C₂ repeat size in blood, as measured by Southern blot, has a significant positive correlation with (A) age at disease onset ($n = 79$; $R = 0.254$; $p = 0.05$) and (B) age at sample collection ($n = 89$; $R = 0.281$; $p = 0.02$). (C) When individuals are separated by median repeat size, age at collection is significantly higher in the top half than the bottom half ($n = 89$; Mann-Whitney U test: $p < 0.01^{**}$). Repeat size does not correlate with (D) survival ($n = 61$; $R = 0.031$; $p = 0.91$) after onset or (E) ALSFRS-R rates of decline ($n = 66$; $R = 0.053$; $p = 0.90$). (F) In an unexpanded (<30 repeats) population, haplotype carrying individuals have a significantly larger average repeat size in blood than non-carriers ($n = 244$ haplotype carriers, $n = 430$ non-carriers; Mann-Whitney U test: $p < 0.0001$). (G) In non-ALS individuals carrying the *C9orf72*-associated risk haplotype with large, yet unexpanded repeat sizes (upper quintile; 13-25 repeats) repeat size is significantly correlated with age at sample collection ($n = 47$; $R = .434$; $p < 0.01$). (H) Repeat size is significantly higher in the top half of large, unexpanded repeat carriers than the bottom half ($n = 47$; Mann-Whitney U test: $p < 0.01^{**}$). All p-values are corrected for multiple comparisons (FDR).

carriers of the *C9orf72* risk haplotype. Mirroring previously described populations (Smith et al. 2012; van der Zee et al. 2013), haplotype carriers in this cohort have a significantly larger repeat size than non-carriers, suggesting that this genetic background may be permissive to expansion (**Fig 5.3F**).

Inherited repeat size is a known risk factor for active expansion in similar repeat disorders (Fu et al. 1991). We hypothesized that repeat size would be age-dependent in unexpanded individuals with large repeat sizes. We defined “large, unexpanded” repeats as the upper quintile of repeat sizes in the haplotype carrier group (upper 47 of 244 individuals). Indeed, within this group, we observed a significant positive correlation with age at sample collection (**Fig 5.3G-H**).

poly(GP) dipeptides are large and CSF levels negatively correlate with DNA repeat size

We measured relative concentrations of poly(GP) in CSF of C9ALS and non-ALS carriers from the prospective cohort, as well as SALS individuals from WU, confirming C9-specificity of the immunoassay (**Fig 5.4A**). Next, we examined poly(GP) in longitudinal CSF draws from a subset of individuals and observed that poly(GP) levels remain steady over time (**Fig 5.4B**), consistent with previous reports (Gendron et al. 2017). We found no significant correlations between poly(GP) CSF levels and ALS history measures, such as age at onset, survival, and ALSFRS-R rate of change (**Fig 5.4C-E**). Interestingly, we observed a significant negative correlation between DNA repeat size and poly(GP) levels (**Fig 5.4F**).

Finally, we investigated DPR size in human CNS autopsy tissue. We fractionated soluble protein from C9ALS CNS autopsy tissues by size and measured relative poly(GP) concentration

in various size ranges. This assay efficiently separated protein sizes covering the expected range of DPR repeats (**Fig 5.4G**). We observed poly(GP) signal only in large fractions of both cerebellum and frontal cortex (**Fig 5.4H**), indicative of full length repeat peptides.

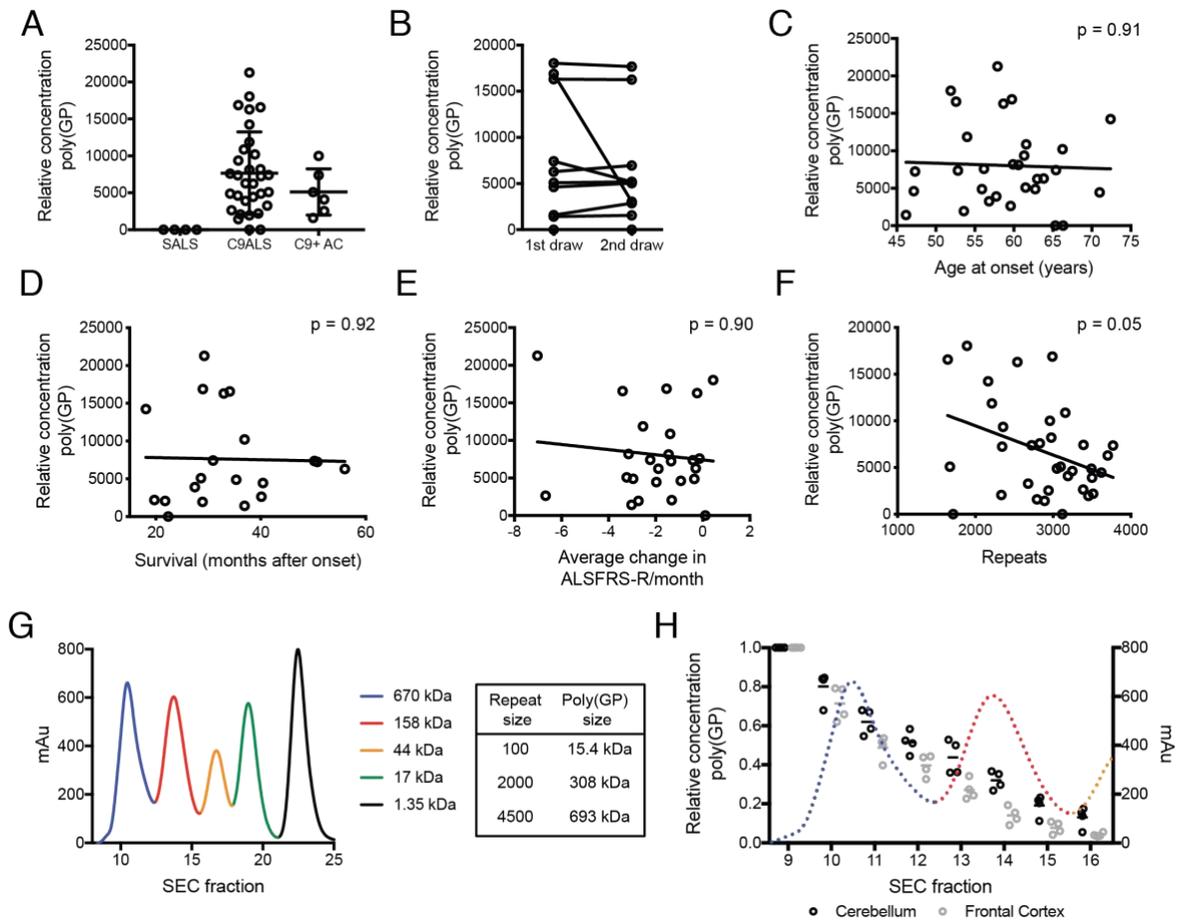


Figure 5.4. poly(GP) in C9ALS CSF and size in autopsy tissue.

(A) poly(GP) in CSF is highly C9-specific, with signal only observed in full expansion carriers. ($n = 32$ C9ALS; $n = 4$ SALS; $n = 6$ C9+ non-ALS carriers (AC)). (B) poly(GP) levels are consistent between draws for 9 out of 10 individuals with longitudinal CSF draws. (C-F) CSF poly(GP) correlations with (C) ALS age at onset ($n = 30$ individuals; $R = 0.040$; $p = 0.91$), (D) survival ($n = 20$ individuals; $R = 0.023$; $p = 0.92$), (E) ALSFRS-R average monthly rate of change ($n = 25$ individuals; $R = 0.111$; $p = 0.90$), and (F) blood DNA repeat size, as measured by Southern blot ($n = 34$ individuals; $R = .371$; $p = 0.05$). (G) SEC standard proteins (Bio-Rad, Hercules, CA, USA) separated by a Superdex 10/300 GL SEC column used in poly(GP) size assessments, demonstrating efficient separation of relevant protein sizes. (H) poly(GP) is observed in large SEC-separated C9ALS autopsy CNS samples. All samples are normalized to fraction 9 (the largest SEC fraction after void volume). Dotted line represents SEC standard proteins as shown in “G”. Matched patient autopsies were used for cerebellum and frontal cortex analyses. All p-values are corrected for multiple comparisons (FDR).

Discussion

Immense progress has been made in understanding the pathobiology of C9ALS since its discovery in 2011. Of particular importance has been the discovery of RNA foci and DPR pathologies, which numerous laboratory studies have now linked to toxicity in human and animal models (Gendron and Petrucelli 2017). Correspondingly, much focus has been recently placed on repeat RNA-lowering therapeutics such as antisense oligonucleotides (ASOs), which effectively reduce both RNA foci and DPR pathologies and have been shown to alleviate defects in C9ALS cell and animal models (Donnelly et al. 2013; Jiang et al. 2016; Lagier-Tourenne et al. 2013; Sareen et al. 2013). These promising treatments are rapidly progressing toward clinical trials, following the path of similar ASO therapies for spinal muscular atrophy (SMA) (Finkel et al. 2017) and *SOD1* ALS (Miller et al. 2013). The data herein will help facilitate design of these trials by providing an evidence-based description of the natural history and pathological features of C9ALS and further supporting poly(GP) as a pharmacodynamic biomarker.

In our prospective analysis of C9ALS clinical characteristics, we found a mean age at onset of 57.9 ± 8.3 years and a median survival of 36.9 months (**Fig 5.1A-B**). Interestingly, our retrospective C9ALS cohort presented with similar onset (59.0 ± 9.3 years) but a shorter survival (29.9 months) than the prospective cohort. While the reasons for these disparities are unclear, this may reflect inherent differences in prospective and retrospective data collection. A number of retrospectively-collected C9ALS natural history datasets have been previously reported, however clinical characteristics in these studies have been notably variable, for reasons difficult to pinpoint (van Blitterswijk et al. 2013; Boeve et al. 2012; Byrne et al. 2012; Irwin et al. 2013; Majounie et al. 2012; Van Mossevelde et al. 2017; Stewart et al. 2012; van der Zee et al. 2013). Additionally, one small prospective study has described C9ALS disease progression (Floeter et

al. 2017), however, this study included only 21 symptomatic C9ALS individuals, of which only 9 reached a survival endpoint during the study period. By prospectively collecting clinical data from a large, multi-center C9ALS population, this report definitively describes C9ALS natural history, which will aid upcoming trials in accurately determining study length and other parameters, such as power calculations, that are dependent on expectations of participants' survival.

In our retrospective analysis, we identified less variance in age at onset and fewer very slowly-progressing individuals in C9ALS compared to SALS (**Fig 5.1E-F**). In measures of disease progression from prospective C9ALS individuals, we observed an ALSFRS-R rate of decline of -1.8 ± 1.7 and SVC of $-1.4\% \pm 3.24\%$ of predicted (**Fig 5.2A-B**). In comparison, the PRO-ACT dataset consisting of over 8600 ALS individuals compiled from 16 ALS clinical trials demonstrated more variation in rates of decline, at -1.02 ± 2.3 and $-2.3\% \pm 6.9\%$ for ALSFRS-R and forced vital capacity, respectively (Atassi et al. 2014), however why differences are observed between the mean values of these measures between these two datasets is not yet understood. In summary, the C9ALS population presented here displays a relatively more homogenous clinical description than non-C9 ALS. Finally, we assessed utility of ALS-CBS and ALSCBQ in evaluating longitudinal cognitive changes in the prospective C9ALS cohort. In either measure, we were unable to observe changes over time (**Fig 5.2C-D**), though it is unclear whether these results represent a true lack of decline or an inability of these measures to accurately assess longitudinal cognitive function. With all measures of disease progression presented here, it is important to consider possible ascertainment bias due to unavailability of follow up and longitudinal data, which may be more likely to occur in quickly-progressing individuals. Additionally, while rare, a small percentage of the C9ALS individuals may carry a second,

unrelated disease-linked mutation, which may affect progression rates in an undetermined way. This was not assessed in this population. Nevertheless, collectively these data provide a template for C9ALS clinical trial design moving forward, and it is expected that the relative homogeneity of the C9ALS population will decrease the number of participants required to observe a treatment effect.

Importantly, during our prospective study, we collected C9ALS biofluids which are now housed in the NEALS Biorepository and are available for *C9orf72*-related research (<https://www.neals.org/for-als-researchers/neals-sample-repository/>). We analyzed blood DNA repeat size in full expansion carriers, which we found to correlate with individuals' ages of sample collection (**Fig 5.3B-C**), a result that mirrors several previously published reports in C9ALS populations (Beck et al. 2013; van Blitterswijk et al. 2013; Fournier et al. 2018; Hübers et al. 2014; Nordin et al. 2015; Suh et al. 2015), including one recent report showing a similar relationship in a large cohort of presymptomatic expansion carriers (Fournier et al. 2018). Interestingly, we also observed a positive correlation of repeat size and age in a separate population of unexpanded non-ALS individuals (**Fig 5.3G-H**). One possible interpretation of these data is that the *C9orf72* repeat is dynamic and able to expand over time. Similar genomic instability has been demonstrated in other expansion disorders (Cossée et al. 1997; Ditch et al. 2009; Du et al. 2012; Fu et al. 1991; McMurray 2010; Møllersen et al. 2010). Further, an abundance of SALS is attributed to *C9orf72* expansions (Majounie et al. 2012) and somatic mosaicism is commonly observed in C9ALS expansions (van Blitterswijk et al. 2013; Dols-Icardo et al. 2014; Nordin et al. 2015), both suggesting dynamic expansion is possible. Alternatively, it is possible that presence of larger repeats is in some way protective and is thus over-represented in older populations. Indeed, we observed a positive correlation between repeat

size and age at onset in the C9ALS prospective cohort (**Fig 5.3A**). In support of this, it is known that presence of large repeats decreases RNA levels of *C9orf72* (DeJesus-Hernandez et al. 2011; Haeusler et al. 2014; Renton et al. 2011; Sareen et al. 2013), which could potentially provide a protective effect by lowering levels of nascent RNA foci or DPRs. Additionally, sequestration of larger repeat RNAs into RNA foci in the nucleus could result in less cytoplasmic RAN translation, as supported by one study that found RNA foci and poly(GP) inclusions rarely occurring in the same cells (Gendron et al. 2013). Consistent with this, in this study we observed a negative correlation between DNA repeat size and poly(GP) levels in C9ALS individuals (**Fig 5.4F**), which could be attributed to these or other yet unexplored mechanisms. It should be noted that these results were obtained from blood-derived DNA and may not accurately represent CNS expansion characteristics. It is also important to note that one previous report found a significantly older age at onset and lower methylation of an upstream CpG island in “short” expansion carriers (55-100 repeats) than in full expansion carriers, supporting a possible inverse relationship of repeat size and onset age (Gijssels et al. 2016), while two other reports were unable to find any relationship of repeat size with onset age (Chen et al. 2015; Dols-Icardo et al. 2014), albeit with smaller sample sizes than in this study. Future studies in cell culture and animal models will help to refine the interpretation of these interesting correlations.

As C9ALS-targeted therapies begin to move toward clinical trials, it becomes increasingly important to develop methods to measure drug target engagement. To this end, C9ALS pharmacodynamic biomarkers have begun to be explored, such as poly(GP) (Gendron et al. 2017). We sought to provide insight into the characteristics of poly(GP) in both CSF and autopsy tissue to further investigate use of this DPR as a C9ALS biomarker. While we found no correlations between CSF poly(GP) and ALS natural history measures (**Fig 5.4C-E**), we were

able to confirm that CSF poly(GP) is C9-specific and its levels are consistent over time (**Fig 5.4A-B**), supporting its continued use as a pharmacodynamics biomarker for expansion-targeted therapeutics. However, the lack of correlation between poly(GP) levels and clinical measures, which is consistent with previous findings (Gendron et al. 2017), suggests that this measure likely does not inform on disease status and perhaps other pathologies are involved in disease pathogenesis, though these analyses could be confounded by low sample sizes in some cases. Lastly, we sought to examine DPR size in C9ALS individuals. In human CNS samples, we observed very large poly(GP) dipeptides, including those upwards of 2000 repeats (**Fig 5.4H**). While we did not observe smaller poly(GP) species here, it is possible that these are below the sensitivity limits of our immunoassay; as the epitope for poly(GP) antibodies is inherently repetitive, this assay likely has greater sensitivity for larger species. Additionally, while we only used soluble lysates in these analyses and ran this experiment under highly denaturing conditions, we cannot rule out the possibility that these sizes could be obscured by some degree of secondary structure or protein interactions. Nevertheless, this experiment highlights two important findings: large DPRs are synthesized in human brains; and poly(GP) being measured with immunoassays is likely only representing large species. Our results suggest that when possible, large DPRs should be used in mechanistic studies.

In summary, we have presented a comprehensive clinical description of C9ALS and provided a template for upcoming C9ALS clinical trials. Additionally, through collection of longitudinal biofluid samples and comparison to our prospectively-collected clinical and demographic datasets, we have highlighted important pathobiological correlations within this population. Together, these results establish baseline clinical and pathological characteristics for C9ALS and provide a reliable resource for future clinical and translational studies.

Acknowledgements

We thank the participants, families, and caregivers involved in this study. We thank members of the Northeast ALS Consortium (NEALS) for referrals to the study. Funding for the study was provided by Biogen Inc., the ALS Association, the Muscular Dystrophy Association, the Knight Alzheimer's Disease Research Center (ADRC) at Washington University, and C9ID foundation. Other support provided by U.S.P.H.S. grant 5UL1 RR024992-02. Unexpanded control data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the

University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. AJC was supported by grant 5T32GM008151-32. RHBa was supported by NIH R01 NS097545, and the Robert and Louis Schwab Family. RHBr receives support from NINDS (R01 NS088689, R01 NS073873, R01 NS104022), the ALS Association, ALS Finding a Cure, ALS ONE, the Angel Fund for ALS Research and Project ALS. TMM was supported by NINDS R01 NS078398. We gratefully acknowledge the study coordinators at each site: Mark Levine-Weinberg, Jessica Zhu, Leah Miller, Julia Yasek, Tiina Jie Huan Xu, Daniela Grasso, Helen Mejia-Santana, Jessica Singleton, Natalia Leontovich, Kristen Riley, Carolyn Prina, Dana Fine, Peggy Allred, Vy Nguyen, and Koral Wheeler. We also thank the group from NCRI, Leonard Friedman, Priyadarshini Vader, Lucia Alvarado-Balderrama, Amanda Nichols, Haining Li, and Ysura Wahab.

Author contributions

Study designed by AJC, NA, CDr, JJB, AMc, TF, MC, and TMM. Clinical study coordination by TH, CDr, JJB, DMY, AMa, MAE, JHV, BSV, MAO, CDo, and TMM. Clinical data management by NA, TH, SB, CDr, JJB, AVS, MAO and HY. NA, LHvdB, MH, RHBa, RHBr, JDR, MAO, and TMM enrolled individuals for the prospective C9ALS study. CC and The Knight ADRC collected, sequenced, and characterized unexpanded DNA samples. AJC, NA, TH, SB, and TMM analyzed clinical data. AJC, TH, NW, and TJE generated and analyzed data from biological samples. Manuscript written and edited by AJC, NA, TH, LHvdB, MH, RHBa, RHBr, JDR, JJB, AMc, MAE, JHV, AS, AVS, TF, CC, MC, and TMM.

Declaration of interests

NA has consulted for MT Pharma, Neuropore, Chronos, Boston Pharmaceuticals, Denali, GSK, and Anelixis. LHvdB has received personal fees from Shire, Biogen, Cytokinetics, and Treeway, outside the submitted work. MH has grant funding from Biogen. RHBa has consulted for Kite Pharmaceuticals, Maze Therapeutics, and Acurastem. JDR is a consultant for Glaxo Smith Kline and Expansion Therapeutics. AS has received consulting fees for statistical reviews in Circulation: Cardiovascular Imaging. AMc and TF are employees of Biogen. CC receives research support from Biogen, Eisai, Alector and Parabon and is a member of the advisory board of ADx Healthcare, Halia therapeutics and Vivid Genomics. MC is a consultant for Biohaven, Takeda, Avexis, and Biogen and was chair of DSMB for Lilly. TMM has served on medical advisory boards for Biogen and Ionis Pharmaceuticals and is a consultant for Cytokinetics. TMM/Washington University have a licensing agreement with Ionis Pharmaceuticals and with C2N. All other authors report no disclosures.

Bibliography

- Ash, Peter E a et al. 2013. "Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to C9FTD/ALS." *Neuron* 77(4): 639–46.
<http://dx.doi.org/10.1016/j.neuron.2013.02.004>.
- Atassi, Nazem et al. 2014. "The PRO-ACT Database." *Neurology* 83(19): 1719–25.
- Beck, Jon et al. 2013. "Large C9orf72 Hexanucleotide Repeat Expansions Are Seen in Multiple Neurodegenerative Syndromes and Are More Frequent than Expected in the UK Population." *American Journal of Human Genetics* 92(3): 345–53.
<http://www.sciencedirect.com/science/article/pii/S0002929713000438> (May 14, 2015).
- van Blitterswijk, Marka et al. 2013. "Association between Repeat Sizes and Clinical and Pathological Characteristics in Carriers of C9ORF72 Repeat Expansions (Xpansize-72): A Cross-Sectional Cohort Study." *The Lancet Neurology* 12(10): 978–88.
[http://dx.doi.org/10.1016/S1474-4422\(13\)70210-2](http://dx.doi.org/10.1016/S1474-4422(13)70210-2).
- Boeve, Bradley F. et al. 2012. "Characterization of Frontotemporal Dementia and/or Amyotrophic Lateral Sclerosis Associated with the GGGGCC Repeat Expansion in C9ORF72." *Brain* 135(3): 765–83.
- Byrne, Susan et al. 2011. "Proposed Criteria for Familial Amyotrophic Lateral Sclerosis." *Amyotrophic Lateral Sclerosis* 12(3): 157–59.
- . 2012. "Cognitive and Clinical Characteristics of Patients with Amyotrophic Lateral Sclerosis Carrying a C9orf72 Repeat Expansion: A Population-Based Cohort Study." *The Lancet Neurology* 11(3): 232–40. [http://dx.doi.org/10.1016/S1474-4422\(12\)70014-5](http://dx.doi.org/10.1016/S1474-4422(12)70014-5).
- Callister, Janis Bennion et al. 2016. "Modelling C9orf72 Dipeptide Repeat Proteins of a Physiologically Relevant Size." *Human Molecular Genetics* 44(Manuscript pre-

- publication). <http://hmg.oxfordjournals.org/>.
- Chen, Yongping et al. 2015. "Large C9orf72 Repeat Expansions Are Seen in Chinese Patients with Sporadic Amyotrophic Lateral Sclerosis." *Neurobiology of Aging* 38: 217.e15-217.e22. <http://dx.doi.org/10.1016/j.neurobiolaging.2015.11.016>.
- Cossée, M et al. 1997. "Evolution of the Friedreich's Ataxia Trinucleotide Repeat Expansion: Founder Effect and Premutations." *Proceedings of the National Academy of Sciences of the United States of America* 94(14): 7452–57.
- DeJesus-Hernandez, Mariely et al. 2011. "Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS." *Neuron* 72(2): 245–56. <http://dx.doi.org/10.1016/j.neuron.2011.09.011>.
- Ditch, Scott, Mimi C. Sammarco, Ayan Banerjee, and Ed Grabczyk. 2009. "Progressive GAA-TTC Repeat Expansion in Human Cell Lines." *PLoS Genetics* 5(10): 28–30.
- Dols-Icardo, Oriol et al. 2014. "Characterization of the Repeat Expansion Size in C9orf72 in Amyotrophic Lateral Sclerosis and Frontotemporal Dementia." *Human Molecular Genetics* 23(3): 749–54.
- Donnelly, Christopher J. et al. 2013. "RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention." *Neuron* 80(2): 415–28. <http://dx.doi.org/10.1016/j.neuron.2013.10.015>.
- Du, Jintang et al. 2012. "Role of Mismatch Repair Enzymes in GAA·TTC Triplet-Repeat Expansion in Friedreich Ataxia Induced Pluripotent Stem Cells." *Journal of Biological Chemistry* 287(35): 29861–72.
- Finkel, Richard S. et al. 2017. "Nusinersen versus Sham Control in Infantile-Onset Spinal Muscular Atrophy." *New England Journal of Medicine* 377(18): 1723–32.

<http://www.nejm.org/doi/10.1056/NEJMoa1702752>.

Floeter, Mary K. et al. 2017. "Disease Progression in *C9orf72* Mutation Carriers." *Neurology* 89(3): 234–41. <http://www.neurology.org/lookup/doi/10.1212/WNL.0000000000004115>.

Fournier, Clémence et al. 2018. "Relations between *C9orf72* Expansion Size in Blood, Age at Onset, Age at Collection and Transmission across Generations in Patients and Presymptomatic Carriers." *Neurobiology of Aging* (September): 1–8. <https://linkinghub.elsevier.com/retrieve/pii/S0197458018303336>.

Freibaum, Brian D., and J. Paul Taylor. 2017. "The Role of Dipeptide Repeats in *C9ORF72*-Related ALS-FTD." *Frontiers in Molecular Neuroscience* 10(February): 1–9. <http://journal.frontiersin.org/article/10.3389/fnmol.2017.00035/full>.

Fu, Ying Hui et al. 1991. "Variation of the CGG Repeat at the Fragile X Site Results in Genetic Instability: Resolution of the Sherman Paradox." *Cell* 67(6): 1047–58.

Gendron, Tania F. et al. 2013. "Antisense Transcripts of the Expanded *C9ORF72* Hexanucleotide Repeat Form Nuclear RNA Foci and Undergo Repeat-Associated Non-ATG Translation in *C9FTD/ALS*." *Acta Neuropathologica* 126(6): 829–44.

———. 2017. "Poly(GP) Proteins Are a Useful Pharmacodynamic Marker for *C9ORF72*-Associated Amyotrophic Lateral Sclerosis." *Science Translational Medicine* 9(383). http://stm.sciencemag.org/content/9/383/eaai7866?utm_campaign=toc_stm_2017-03-29&et rid=291001628&et cid=1243883.

Gendron, Tania F, and Leonard Petrucelli. 2017. "Disease Mechanisms of *C9ORF72* Repeat Expansions." *Cold Spring Harbor Laboratory Press*: 1–22.

Gijssels, I. et al. 2016. "The *C9orf72* Repeat Size Correlates with Onset Age of Disease, DNA Methylation and Transcriptional Downregulation of the Promoter." *Molecular Psychiatry*

21(8): 1112–24.

- Haeusler, Aaron R et al. 2014. “C9orf72 Nucleotide Repeat Structures Initiate Molecular Cascades of Disease.” *Nature* 507(7491): 195–200.
<http://www.ncbi.nlm.nih.gov/pubmed/24598541>.
- Harms, Matthew B. et al. 2013. “Lack of C9ORF72 Coding Mutations Supports a Gain of Function for Repeat Expansions in Amyotrophic Lateral Sclerosis.” *Neurobiology of Aging* 34(9): 2234.e13-2234.e19. <http://dx.doi.org/10.1016/j.neurobiolaging.2013.03.006>.
- Hübers, Annemarie et al. 2014. “Polymerase Chain Reaction and Southern Blot-Based Analysis of the C9orf72 Hexanucleotide Repeat in Different Motor Neuron Diseases.” *Neurobiology of Aging* 35(5): 1214.e1-1214.e6. <http://dx.doi.org/10.1016/j.neurobiolaging.2013.11.034>.
- Irwin, David J. et al. 2013. “Cognitive Decline and Reduced Survival in C9orf72 Expansion Frontotemporal Degeneration and Amyotrophic Lateral Sclerosis.” *Journal of Neurology, Neurosurgery and Psychiatry* 84(2): 163–69.
- Jiang, Jie et al. 2016. “Gain of Toxicity from ALS/FTD-Linked Repeat Expansions in C9ORF72 Is Alleviated by Antisense Oligonucleotides Targeting GGGGCC-Containing RNAs.” *Neuron*: 1–16. <http://dx.doi.org/10.1016/j.neuron.2016.04.006>.
- Lagier-Tourenne, Clotilde et al. 2013. “Targeted Degradation of Sense and Antisense C9orf72 RNA Foci as Therapy for ALS and Frontotemporal Degeneration.” *Proceedings of the National Academy of Sciences of the United States of America* 110(47): E4530-9.
<http://www.pnas.org/content/110/47/E4530.full> (May 7, 2015).
- Lee, Youn Bok et al. 2013. “Hexanucleotide Repeats in ALS/FTD Form Length-Dependent RNA Foci, Sequester RNA Binding Proteins, and Are Neurotoxic.” *Cell Reports* 5(5): 1178–86.

- Lehmer, Carina et al. 2017. "Poly-GP in Cerebrospinal Fluid Links C9orf72-Associated Dipeptide Repeat Expression to the Asymptomatic Phase of ALS/FTD." *EMBO Molecular Medicine*: 1–10.
- Majounie, Elisa et al. 2012. "Frequency of the C9orf72 Hexanucleotide Repeat Expansion in Patients with Amyotrophic Lateral Sclerosis and Frontotemporal Dementia: A Cross-Sectional Study." *Lancet Neurology* 11: 323–30.
[http://www.thelancet.com/pdfs/journals/lanneur/PIIS1474-4422\(12\)70043-1.pdf](http://www.thelancet.com/pdfs/journals/lanneur/PIIS1474-4422(12)70043-1.pdf) (May 3, 2015).
- McMurray, Ct. 2010. "Mechanisms of Trinucleotide Repeat Instability during Human Development." *Nature Reviews Genetics* 11(11): 786–99.
<http://www.nature.com/nrg/journal/vaop/ncurrent/full/nrg2828.html>.
- Miller, Timothy M. et al. 2013. "An Antisense Oligonucleotide against SOD1 Delivered Intrathecally for Patients with SOD1 Familial Amyotrophic Lateral Sclerosis: A Phase 1, Randomised, First-in-Man Study." *The Lancet Neurology* 12(5): 435–42.
- Mizielinska, Sarah et al. 2014. "C9orf72 Repeat Expansions Cause Neurodegeneration in *Drosophila* through Arginine-Rich Proteins." *Science* 345(9): 1192–95.
<http://www.ncbi.nlm.nih.gov/pubmed/19556136>
<http://www.ncbi.nlm.nih.gov/pubmed/25103406>.
- Mok, Kin et al. 2012. "The Chromosome 9 ALS and FTD Locus Is Probably Derived from a Single Founder." *Neurobiology of Aging* 33(1): 209.e3-209.e8.
<http://dx.doi.org/10.1016/j.neurobiolaging.2011.08.005>.
- Møllersen, Linda et al. 2010. "Continuous and Periodic Expansion of CAG Repeats in Huntington's Disease R6/1 Mice." *PLoS Genetics* 6(12): 1–11.

- Mori, Kohji et al. 2013. “The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTL D/ALS.” *Science* 339(March): 1335–39.
<http://www.sciencemag.org/content/suppl/2013/02/06/science.1232927.DC1.full> (May 7, 2015).
- Van Mossevelde, Sara et al. 2017. “Clinical Evidence of Disease Anticipation in Families Segregating a C9orf72 Repeat Expansion.” *JAMA Neurology*: 1–8.
<http://archneur.jamanetwork.com/article.aspx?doi=10.1001/jamaneurol.2016.4847>.
- Nguyen, Lien et al. 2019. “Antibody Therapy Targeting RAN Proteins Rescues C9 ALS/FTD Phenotypes in C9orf72 Mouse Model.” *Neuron* 105(February 19): 1–18.
<https://linkinghub.elsevier.com/retrieve/pii/S0896627319309705>.
- Nordin, A. et al. 2015. “Extensive Size Variability of the GGGGCC Expansion in C9orf72 in Both Neuronal and Non-Neuronal Tissues in 18 Patients with ALS or FTD.” *Human Molecular Genetics* 24(February): 1–10.
<http://www.hmg.oxfordjournals.org/cgi/doi/10.1093/hmg/ddv064>.
- Renton, Alan E. et al. 2011. “A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD.” *Neuron* 72(2): 257–68.
- Sareen, Dhruv et al. 2013. “Targeting RNA Foci in iPSC-Derived Motor Neurons from ALS Patients with a C9ORF72 Repeat Expansion.” *Science Translational Medicine* 5(208): 208ra149.
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4090945&tool=pmcentrez&rendertype=abstract>.
- Smith, Bradley N et al. 2012. “The C9ORF72 Expansion Mutation Is a Common Cause of ALS+/-FTD in Europe and Has a Single Founder.” *European Journal of Human Genetics*

- (February 2012): 102–8.
- Stewart, Heather et al. 2012. “Clinical and Pathological Features of Amyotrophic Lateral Sclerosis Caused by Mutation in the C9ORF72 Gene on Chromosome 9p.” *Acta Neuropathologica* 123(3): 409–17.
- Su, Zhaoming et al. 2014. “Discovery of a Biomarker and Lead Small Molecules to Target r(GGGGCC)-Associated Defects in C9FTD/ALS.” *Neuron*: 1–8.
<http://dx.doi.org/10.1016/j.neuron.2014.07.041>.
- Suh, Eun Ran et al. 2015. “Semi-Automated Quantification of C9orf72 Expansion Size Reveals Inverse Correlation between Hexanucleotide Repeat Number and Disease Duration in Frontotemporal Degeneration.” *Acta Neuropathologica* 130(3): 363–72.
- Turner, Martin R. et al. 2017. “Genetic Screening in Sporadic ALS and FTD.” *Journal of Neurology, Neurosurgery and Psychiatry* 88: 1042–44.
- Wen, Xinmei et al. 2014. “Antisense Proline-Arginine RAN Dipeptides Linked to C9ORF72 - ALS / FTD Form Toxic Nuclear Aggregates That Initiate In Vitro and In Vivo Neuronal Death.” *Neuron* 84(6): 1213–25. <http://dx.doi.org/10.1016/j.neuron.2014.12.010>.
- Wijsman, Ellen M. et al. 2011. “Genome-Wide Association of Familial Late-Onset Alzheimer’s Disease Replicates BIN1 and CLU and Nominates CUGBP2 in Interaction with APOE.” *PLoS Genetics* 7(2).
- van der Zee, Julie et al. 2013. “A Pan-European Study of the C9orf72 Repeat Associated with FTL D: Geographic Prevalence, Genomic Instability, and Intermediate Repeats.” *Human Mutation* 34(2): 363–73.
- Zhou, Qihui et al. 2017. “Antibodies Inhibit Transmission and Aggregation of C9orf72 Poly-GA Dipeptide Repeat Proteins.” *EMBO Molecular Medicine*: 1–16.

———. 2019. “Active Poly-GA Vaccination Prevents Microglia Activation and Motor Deficits in a C9orf72 Mouse Model.” *EMBO molecular medicine*: e10919.

<http://www.ncbi.nlm.nih.gov/pubmed/31858749>.

Zu, Tao et al. 2013. “RAN Proteins and RNA Foci from Antisense Transcripts in C9ORF72 ALS and Frontotemporal Dementia.” *Proceedings of the National Academy of Sciences of the United States of America* 110(51): E4968-77.

<http://www.pnas.org/content/110/51/E4968.full> (May 3, 2015).

CHAPTER 6

Conclusions and future directions

Summary of the dissertation

This dissertation details four separate projects: Chapter 2: development of adeno-associated virus (AAV) calling cards for TF profiling in the brain; Chapter 3: investigation of BRD4 binding patterns predictive of phenotypic outcomes in acute seizure; Chapter 4: enhancer and miRNA profiling in *C9orf72* ALS (C9ALS) mouse models and human tissues; Chapter 5: defining the natural history, biomarkers, and pathobiology of C9ALS.

In Chapter 2, we described the development of AAV calling cards and demonstrated proof-of-principle experiments showing utility of this technique in cell type-specific and longitudinal TF recording. These advances allow for insights into TF-mediated epigenetic regulation that could not be ascertained with previously available profiling techniques.

In Chapters 3 and 4, we utilized AAV calling cards as well as other established epigenomic profiling methodologies to investigate epigenetic misregulation in neurological disease models. In Chapter 3, we probed pre-insult BRD4 binding in an acute seizure model and identified enhancers for which BRD4 occupancy predicts eventual phenotypic outcomes. Then, in Chapter 4, we explored misregulation of transcriptional and translational epigenetic control through enhancers and microRNAs (miRNAs), respectively, in C9ALS. We found a broad pattern of disrupted epigenetic regulation in both mouse models and postmortem human tissues, suggesting that *C9orf72* mutations drive epigenetic aberrations which may be involved in disease pathogenesis.

Finally, in Chapter 5, we detail our efforts to understand pathological and clinical features of C9ALS patients. As part of a large natural history study of C9ALS, we probed size and abundance of key *C9orf72*-related pathologies, including dipeptide repeat products (DPRs) and DNA repeat expansions, in patient blood, cerebrospinal fluid (CSF), and postmortem CNS

tissues. We found the presence of large DPRs in human brain and uncovered novel pathobiological correlations between DNA repeat size and age, suggesting that such expansions may increase in size over time.

Below, we discuss in detail the advances, limitations, and future directions associated with each of these studies.

Advances, limitations, and future directions

Development of AAV calling cards

AAV calling cards is a TF-tagging technique, in which binding events of a TF of interest are marked by transposon insertion via a C-terminally-fused *hyperPiggyBac* (hypPB) transposase. One innate advantage of this system over traditional TF profiling techniques, such as ChIP-seq, is non-requirement of TF-specific antibodies, increasing the number of TFs that can be potentially profiled with our method. Further, this advance eliminates individual assay optimization for each TF and may potentially allow for the profiling of low-abundance TFs that would not be accessible with ChIP-seq due to high input requirements (Aughey and Southall 2016; Southall et al. 2013). In future versions of the AAV calling cards technique, it may even be possible to multiplex the profiling of multiple TFs at the same time using different transposase species (Gogol-Doring et al. 2016), while AAV delivery of this system extends its potential application to other AAV-tractable animal models (e.g. rats, primates). However, exogenous TF expression also carries inherent risks that should be considered when contemplating which methodology to choose for an experiment. It is possible that overexpression of the TF of interest may induce artifactual binding patterns, or even transcriptional or phenotypic cellular changes in some cases. However, in our experience, calling card and ChIP-seq data largely align (Cammack, Moudgil, et al. 2019; Moudgil et al. 2019; Wang et al. 2012), suggesting that factors other than TF expression level may regulate site-specific binding (e.g. surrounding chromatin state (Guertin and Lis 2010; John et al. 2011; X. Liu et al. 2006; Neph et al. 2012), local sequence context (White et al. 2013; Yanez-Cuna et al. 2012), co-expression of co-factors (Slattery et al. 2011), etc.). Moreover, we have demonstrated here that even truncated TFs can largely recapitulate binding profiles of full-length TFs; thus, some of these overexpression

concerns could be alleviated in future experiments by exclusion of active protein domains (other than DNA-binding domains) when creating TF-hypPB fusions. Further, calling cards applications are limited to live tissues, where TFs can be actively recorded; thus, profiling in human tissues is not currently possible with our technique. Finally, until AAV calling cards becomes more widely established, it remains important to validate functionality of each TF fusion via orthogonal means, as not all TFs are likely to redirect hypPB insertions to the same degree. This may be achieved by comparing calling cards datasets to reference ChIP-seq datasets, which are often publicly available for *in vitro* cell lines, or alternatively, through motif discovery analyses.

The Flip-Excision (FLEX) versions of AAV calling cards allow for the profiling of TFs in targeted cellular populations via Cre-dependent conditional expression. While other TF profiling techniques have also been modified for cell type-specific use, (Bonn et al. 2012; Deal and Henikoff 2010; Girdhar et al. 2018; Mo et al. 2015; Zhou et al. 2017), these methods have been limited to highly abundant targets and may be complicated by disassociation-related artifacts (van den Brink et al. 2017). FLEX calling cards requires no physical sorting or breeding to tagged-nuclei transgenic animals, negating these concerns and streamlining cell type-specific profiling by utilizing the same library preparation protocol for all cell types. Further, this method has the potential for use in rare cell populations for which starting material may be too little for ChIP-seq and related analyses. However, while we have engineered a novel FLEX variant (“Frontflip”) which dramatically reduces Cre(-) background in FLEX calling cards systems, it is unclear whether the remaining signal from rarer Cre(+) populations would be sufficient for TF profiling, and sensitivities of our system are likely to vary greatly between TFs. Further, FLEX calling cards is inherently limited to cell populations for which Cre driver mouse lines exist,

however this may be circumvented in some cases through the use of cell type-specific promoters (Graybuck et al. 2019; Jüttner et al. 2019) in the TF-hypPB AAV construct. Alternatively, a separate publication from our group details the adaptation of AAV calling cards for bimodal single cell TF and RNA profiling (Moudgil et al. 2019). While the sensitivity of this method is currently limited by transposon insertion rate, this technique may eventually allow for TF profiling in many brain cell types in parallel as single cell technologies gradually improve.

Finally, we demonstrated the ability of AAV calling cards to record and integrate TF binding information across time, and, in doing so, recovered transient TF binding events which were available for readout weeks later. This advantage has the capability to link downstream molecular or behavioral phenotypic outcomes with historical epigenetic information in a uniquely powerful way. One can imagine future experiments wherein the underlying drivers of cell fates can be retrospectively identified. Similarly, there are opportunities to link antecedent TF binding patterns to behavioral outcomes, such as in neurological disease models. As one direct biological example of this type of analysis, in Chapter 3 we described the use of pre-insult, longitudinal recording to identify predictive susceptibility and vulnerability enhancers in an acute seizure model. However, integrative recording comes with potential disadvantages as well. Firstly, there is the possibility that some epigenetic changes could be overwritten or canceled out by events occurring later in the recording period. Secondly, for differences that are observed between two conditions, it would be difficult or impossible to determine the exact timepoint during the recording period at which a TF binding event occurred, and similarly, one may not be able to distinguish singular, high-magnitude events from integration of multiple small-magnitude events at the same locus. Finally, there is the potential for loss of some transient events due to transposon reintegration and/or missed recording of recent events due to transposon exhaustion,

particularly after long periods of recording. Improvements in temporal control of AAV calling cards could help mitigate many of these concerns. While not explored in this dissertation, one could envision addition of drug-inducible mechanics to the AAV system, such as destabilization domains or estrogen receptor fusions, both of which have been shown to confer temporal control to calling cards systems *in vitro* (Qi et al. 2017). Alternatively, the construction of calling cards transgenic mouse lines may allow for coordinated use with conventional doxycycline-inducible systems.

Despite the drawbacks of AAV calling cards detailed above, we believe this system is capable of uncovering new biology in previously inaccessible contexts. Further, we consider this to be “version 1.0” of the system and expect that future improvements may allow its use in yet even more settings. Nonetheless, it remains important to consider all advantages and caveats when choosing a TF profiling system for future experiments.

Retrospective BRD4 profiling to identify pre-seizure susceptibility and vulnerability loci

In Chapter 3, we utilized longitudinal AAV calling cards to test whether BRD4 binding patterns could predict eventual phenotypic outcomes in an acute seizure model. We hypothesized that BRD4 may regulate seizure susceptibility, as previous work had demonstrated that BRD4 mediates transcriptional responses to neuronal activation, which in turn, influences mouse behavior. Perhaps most critically, one study demonstrated that pharmacological BRD4 inhibition in the week prior to induction of acute seizure is sufficient to reduce peak seizure severity in live mice (Korb et al. 2015). We sought to identify specific loci through which BRD4 is acting to mediate these phenotypes.

In our work, we identified a number of loci for which pre-insult BRD4 binding is predictive of seizure severity. Many of these loci were located near to genes with clear roles in epilepsy and/or seizure, supporting that epigenetic mechanisms may have a role in conferring vulnerability to severe seizure. Importantly, this study was made possible by the ability of AAV calling cards to record prodromal epigenetic states and is an example of the power of retrospective TF profiling *in vivo*. This is particularly important in the brain, where epigenetic and transcriptional changes often occur immediately following neuronal stimulation (Stavreva et al. 2015). In the context of seizures, it has been well established that immediate-early genes, such as *c-Fos*, rapidly increase in mRNA and protein expression within minutes of seizure induction (Peng and Houser 2005); thus, endpoint capture methods, such as ChIP-seq or RNA-seq would likely not inform on pre-seizure states. In contrast, by permanently recording BRD4 binding pre-seizure, AAV calling cards allows for unique access to binding information that was present pre-insult, which can then be linked to downstream phenotypic outcomes.

However, while these results generally support that BRD4 regulates seizure vulnerability, individual loci require further validation. We demonstrated that many of the genes proximal to differential BRD4-bound loci have known roles in seizure susceptibility, however the impact of the nearby enhancers identified in our study remains undetermined. This could be investigated through CRISPR-Cas9 mediated deletion of these regions, perhaps in primary or induced pluripotent stem cell (iPSC)-derived neuronal models, followed by assaying of neuronal stimulation responses. Similarly, selective activation, perhaps through dCas9-p300 fusion enzymes (Hilton et al. 2015), may inform on the contributions of individual enhancers to seizure phenotypes.

Next, it possible that some of the observed loci were altered during the seizure itself rather than being present pre-seizure. While we recorded BRD4 over a 28-day pre-seizure period and induced seizures for only 15 minutes prior to euthanizing animals, we have not empirically tested the transposition rate of hypPB *in vivo* and do not know how it may be altered during states of intense neuronal activity, such as seizures. This should be tested in the future by delivering AAV calling cards to multiple cohorts of mice, followed by sacrifice and transposon sequencing at various timepoints to determine the rate of insertion. Further, one could also compare the average insertion total between mice receiving pentylenetetrazole (PTZ) or saline control injections to determine whether insertion rate increases during seizure. However, ultimately this study would benefit the most from the implementation of temporal control, such that AAV calling cards recording may be turned off entirely prior to seizure induction.

Lastly, this work was carried out exclusively in the PTZ acute seizure model, which may not accurately represent human epilepsy. Follow up studies in human iPSC neuron models or in other mouse seizure models would help to corroborate the results depicted here (Löscher 2011; Rogawski 2006). However, it is worth noting that the PTZ model has been used previously in the discovery and testing of antiepileptic drugs (AEDs), many of which have translated successfully to the clinic and are currently used to treat epileptic patients (Löscher 2011; Rogawski 2006). Thus, this model has been proven useful and relevant for mechanistic studies of seizures in the past.

Epigenomic profiling in C9ALS

Chapter 4 explores epigenetic contributions to ALS caused by repeat expansion mutations in *C9orf72* (C9ALS). This work began by profiling miRNA expression levels in the

cervical spinal cords of *C9orf72* mutation-harboring transgenic mice ($C9_{exp}$), which led to the uncovering of a pattern of dysregulation amongst many miRNA species. One advantage of the line we chose to use in this study is that it is accompanied by two separate control lines: its own non-transgenic littermates (which we called “wt”) and a separate transgenic model harboring an unexpanded (15-mer) repeat in *C9orf72* ($C9_{non-exp}$) (O’Rourke et al. 2015). As microarray screens such as this are often underpowered for the identification of statistically significant differences after multiple comparison correction, utilization of both controls adds another layer of confidence when identifying hits for follow up study. We found a significant correlation between the fold changes in miRNA expression levels when comparing $C9_{exp}$ to either control line. Further, these differences appeared to be more pronounced when comparing $C9_{exp}$ to wt than $C9_{non-exp}$, suggesting even small expansions may induce changes in miRNA expression. When selecting individual miRNAs for follow up study in the future, one could choose to focus on miRNAs displaying pronounced up or downregulation when compared to either line, as we reason these are more likely to be informative hits. Importantly, the mice used here were only 14 weeks of age, when *C9orf72* pathologies, such as DPRs, are beginning to accumulate, and further, while these transgenic lines do express human ALS-like pathologic features, they do not develop motor neuron degeneration or ALS phenotypes, suggesting that the miRNA changes observed here are not likely to be driven by inflammation or other insults associated with post-onset ALS. However, it is important to recognize that screening assays such as this are confounded by low sample sizes and large numbers of multiple comparisons, thus rendering all miRNA changes observed here not significant after multiple comparison correcting. While valid for identification of targets for further investigation and for uncovering broad trends, it will be critical to continue to validate the interesting hits of this screen in mice and humans with larger

sample sizes, as well as to investigate functional ramifications of altered miRNA expression for motor neuron health in the future.

We were also interested in investigating enhancer activity in C9ALS, as these elements have been shown to harbor genetic variants associated with increased risk of neurological disease (Corradin and Scacheri 2014; Ernst et al. 2011; Girdhar et al. 2018; Hannon et al. 2019; Hnisz et al. 2013; Maurano et al. 2012; Wells et al. 2015). We first quantified BRD4 binding across the genome in the cerebral cortices of 7.5-8.5 month old *C9orf72* transgenic mice (C9(+); a different line than was used for miRNA profiling (Liu et al. 2016)) and their non-transgenic littermates (C9(-)) with AAV calling cards. In doing so, we discovered hundreds of significantly altered binding sites reaching statistical significance surpassing Bonferroni-corrected thresholds, the majority of which were also significantly altered in post-hoc Student's t-tests, suggesting that variability within genotype was low. We also profiled mRNA levels with RNA-seq in a subset of these animals to investigate consequences of this aberrant epigenetic regulation on gene expression. While no mRNAs were significantly altered (false discovery rate (FDR)-corrected), only 3 mice per group were assayed here, and some genes proximal to differential BRD4 peaks did show trends toward up or downregulation consistent with nearby BRD4 occupancy. Further, gene ontology (GO) of all proximal genes identified neuronal pathways known to be involved in C9ALS, such as neuronal excitability (Donnelly et al. 2013). Similar to those used for our miRNA profiling, these mice do not get motor neuron disease in our hands (despite previously reported phenotypes (Liu et al. 2016)), and these mice were profiled at a later timepoint (7.5-8.5 months) where DPR pathology is abundant ((Liu et al. 2016) and in-house data not shown), suggesting that these changes could be upstream of motor neuron dysfunction. However, it is important to note that these results are limited by the serotype of AAV9, which was used in this

study for AAV calling cards profiling and only transduces neurons and astrocytes (Cammack, Moudgil, et al. 2019). While these two cell types are historically the most heavily implicated in ALS pathogenic mechanisms (Boillee et al. 2006; Haidet-Phillips et al. 2011; Nagai et al. 2007; Yamanaka et al. 2008), other glial types, such as microglia (Boillee et al. 2006) and oligodendrocytes (Ferraiuolo et al. 2016; Kang et al. 2013), have also been shown to alter ALS pathogenesis and progression in animal models. Future studies of BRD4 in C9ALS could be improved through implementation of FLEX calling cards for cell type-specific profiling. It would be exciting to observe, for example, motor neuron-specific epigenetic signatures in these animals. Further, extending these studies to the spinal cord, perhaps through intrathecal AAV delivery, would probe cell types directly affected in C9ALS (though it should be noted that *C9orf72* mutations also cause a significant portion of frontotemporal dementia (FTD) cases, and are thus also worth exploring in cortical regions).

Given our findings in *C9orf72* transgenic mice, we then sought to investigate enhancer activation in human ALS postmortem tissues. To do so, we profiled enhancers and super enhancers via H3K27ac ChIP-seq in lumbar spinal cord tissues from age and sex matched autopsy samples from C9ALS, sALS, and non-ALS (Con) individuals. We found a broad pattern of enhancer and super enhancer dysregulation in the spinal cord which correlated between C9ALS and sALS, suggesting that these alterations may extend beyond just *C9orf72*-associated disease. These results help to corroborate and extend the results we observed in the C9ALS mouse model. We also found that spinal cord enhancers harbor known ALS risk-associated genetic variants, suggesting that misregulation at these regions may be an upstream pathogenic mechanism. Interestingly, in addition to the spinal cord, these enhancers were found to be active nearly exclusively in brain and muscle/muscle-containing tissues (from a reference H3K27ac

ChIP-seq dataset). While this is simply observational at this point, it will be interesting to investigate these variants further in future studies to determine their roles in conferring susceptibility to motor neuron disease.

It is important to note that the findings observed here are descriptive and somewhat preliminary in nature. Whether these epigenetic aberrations contribute to C9ALS disease mechanisms is yet unclear and requires further follow up study. This could be done through selective disruption or activation of enhancer loci via CRISPR Cas9-related methodologies (Hilton et al. 2015), as discussed above in our seizure study, or broad pharmacological or genetic inhibition of epigenetic co-factors, including BRD4 in C9ALS cell and animal models. One challenge with studies such as this is choosing a disease-relevant phenotypic outcome to assay. Some studies have demonstrated that primary and patient-derived iPSC neurons can recapitulate C9ALS disease survival phenotypes (Shi et al. 2018; Wen et al. 2014), suggesting they may be useful in determining functional effects of targeting enhancers. A number of *C9orf72* transgenic and virally induced mouse models have been developed for the study of C9ALS *in vivo* (Chew et al. 2015; Hao et al. 2019; Jiang et al. 2016; Liu et al. 2016; O'Rourke et al. 2015; Peters et al. 2015; Zhang et al. 2019), however only a select few display disease-relevant phenotypes (Jiang et al. 2016; Liu et al. 2016; Zhang et al. 2019), with one having not yet been replicated outside of the depositing laboratory (Liu et al. 2016). As we hypothesize that DPRs may be drivers of the epigenetic aberrations, we have begun to develop our own cell models of DPR-mediated toxicity and hope to utilize these systems in the future for functional investigation of enhancers in C9ALS.

Finally, and perhaps most critically, we are interested in understanding the mechanisms underlying the miRNA and enhancer activity changes we have observed here. One interesting

hypothesis is that miRNA and/or enhancer regulation could be impaired through altered liquid-liquid phase separation (LLPS) of their enzymatic components. Indeed, evidence has been demonstrated that all of these biological processes may involve LLPS. Enzymes involved in both miRNA biogenesis as well as mRNA repression mechanisms naturally undergo LLPS (Jiang et al. 2017; Sheu-Gruttadauria and MacRae 2018), while TFs (Boija et al. 2018), co-factors such as BRD4 (Sabari et al. 2018), and even chromatin itself (Gibson et al. 2019) are thought to use LLPS as a means for increasing their local concentrations at enhancers (Hnisz et al. 2017). LLPS has recently become a major focus in the ALS field as well, as many ALS-linked genes have been shown to undergo this process, while ALS-causing mutations disrupt the dynamicity of phase separated bodies (Dao et al. 2019; Mann et al. 2019; Molliex et al. 2015; Patel et al. 2015). In C9ALS, it has been shown that the arginine-containing DPRs, poly(GR) and poly(PR) themselves phase separate (Boeynaems et al. 2017), co-localize to other phase separated condensates (Lee et al. 2016), and disrupt natural condensate function (Boeynaems et al. 2017; Lee et al. 2016). Interestingly, one recent study demonstrated the poly(PR) co-phase separates with heterochromatin protein 1 (HP1), an enzyme which interacts with methylated histone tails and participates in gene silencing, and in turn, induces alterations in histone H3 posttranslational modifications, both in the poly(PR)-expressing mouse model, as well as C9ALS/FTD postmortem tissue (Zhang et al. 2019). However, whether similar LLPS changes are induced by DPRs for miRNA or enhancer-associated proteins has not yet been explored. We have begun to develop assays to test the hypothesis that presence of DPRs disrupts of LLPS of BRD4, chromatin, and miRNA processing enzymes. These include “optodroplet”-like assays, as well as *in vitro* experiments utilizing recombinant proteins, as has been done for similar proteins in the past (Boija et al. 2018; Lee et al. 2016; Mann et al. 2019; Sabari et al. 2018; Zhang et al. 2019).

However, this remains, at this point, a hypothesis; it is equally possible that other mechanisms may be responsible for driving epigenetic aberrations in C9ALS. For example, TDP-43 has been shown to impair nucleosomal dynamics through the chromatin modifier CHD1 (Berson et al. 2017). Further, global histone hypoacetylation has been observed in mutant *FUS* transgenic mouse models which can be reversed with pharmacological inhibition of histone deacetylases (HDACs) (Rossaert et al. 2019). Future studies will help to elucidate whether these, or other yet unexplored mechanisms may be responsible for the observations described here.

Investigation of pathological characteristics, biomarkers, and natural history of C9ALS

In the penultimate Chapter of this dissertation, we present our C9ALS natural history study (Cammack, Atassi, et al. 2019), wherein we characterize the clinical and pathological characteristics of the C9ALS patient population. This Chapter is the aggregate of years of effort by clinical and basic science research staff and represents a substantial step forward in the path to develop C9ALS-targeted therapeutics. By thoroughly and definitively describing the clinical demographics of C9ALS patients, future clinical trials can now make evidence-based decisions on treatment regimens and timelines.

However, for the purposes of this dissertation, perhaps the most relevant findings to come from this work are from the investigations into DNA expansion and DPR size in human tissues. Given that presence of DPRs is one of the few consistent pathological species between C9ALS transgenic mice and humans, we believe they are a likely candidate for causing the epigenetic aberrations seen in Chapter 4 (of note, this is yet unproven). However, despite intense interest in DPRs in the C9ALS field, it was previously unknown how large DPR peptides are in human CNS tissues. Because DPRs form via repeat associated, non-ATG (RAN) translation

mechanisms (Ash et al. 2013; Gendron et al. 2013; Mori et al. 2013; Zu et al. 2013), it is unclear where translation begins for these protein species. In the majority of mechanistic studies, small (<100-mer) DPRs are utilized, largely for convenience; indeed, the creation of large DPRs is technically challenging, and short DPRs appear to confer similar toxicities as longer DPRs in some contexts (Gendron and Petrucelli 2017). However, DPRs of different sizes may behave differently in biological settings. In our study, we provided the first characterization of DPR size in humans, and in doing so, identified the existence of large poly(GP) species which extend to lengths similar to their DNA repeat derivatives. In addition, we quantified poly(GP) abundance in human CSF samples, as well as size of DNA repeat expansions, both of which we later compared with our prospectively collected clinical and demographic datasets. We found that poly(GP) abundance in CSF was negatively correlated with DNA expansion size, suggesting that large expansions may decrease RAN translation levels. Further, poly(GP) in CSF was specific to C9ALS, helping to solidify its status as a pharmacodynamic biomarker in future clinical trials (Gendron et al. 2017; Lehmer et al. 2017; Su et al. 2014). However, our investigations into DPR characteristics were somewhat constrained in this study by technical limitations; because the antibodies used in our poly(GP) quantification assay were targeted to repetitive epitopes, they likely have increased sensitivity for large species. Thus, we could not eliminate the possibility that small species exist as well but are below the sensitivity thresholds of our assay. Further, our finding that poly(GP) in CSF was not correlated to clinical features of the C9ALS population indicates that this DPR is likely not a valid biomarker of disease progression, which is consistent with previously published findings (Gendron et al. 2017; Lehmer et al. 2017; Su et al. 2014).

For DNA expansions, we found positive correlations between repeat size and age in both C9ALS and unexpanded control individuals, suggesting that these repeats may be unstable and

grow over time. While this hypothesis is appealing and suggests the possibility of *de novo* mutations which could account for the abundance of *C9orf72* mutations described in sALS (DeJesus-Hernandez et al. 2011; Majounie et al. 2012; Renton et al. 2011), our measurements of DNA repeat size in this study were static and do not directly demonstrate change in repeat size over time. Further, because DNA samples were collected near the same time as ALS disease onset, we found that expansion size is also correlated with this measure. Thus, an alternative explanation of these data is that large expansions could be protective and overrepresented in the older onset population. Future mechanistic studies examining whether repeat size changes over time in cell or animal models will be critical for the interpretation of these interesting pathobiological correlations.

Finally, it should be noted that because this study was focused on C9ALS, it did not test for possible double mutation carriers. Further, some patients in our survival analysis had not reached a survival endpoint by study closing, and thus we may be underrepresenting survival estimates in this patient population. Nonetheless however, the work presented here provides a template for C9ALS clinical trials and highlights novel pathobiological correlations within this population, providing new insights and a reliable resource for future translational and mechanistic studies.

Impact, innovation, and final thoughts

In this dissertation, we have detailed here four studies which help to elucidate the mechanistic underpinnings of epigenetic regulation in the brain. Each Chapter tells a unique story of gene expression control, either in the healthy brain or in disease states.

Most importantly, what we have provided here is a set of resources that will lead to innovative new studies in the future. AAV calling cards is a novel method that may unlock original biology in a number of different contexts moving forward. Meanwhile, our epigenetic profiling in disease models, beyond implicating these mechanisms as potentially involved in disease pathogenesis, provide a catalog of specific regulatory processes displaying misregulation in disease. Finally, our pathological and clinical characterization of the C9ALS population will facilitate the design and implementation of future clinical trials.

It is my hope that the results presented here will continue to be a valuable resource to the neuroscience community in the future, as we continue to strive to understand the human brain in health and disease.

Bibliography

- Ash, Peter E a et al. 2013. “Unconventional Translation of C9ORF72 GGGGCC Expansion Generates Insoluble Polypeptides Specific to C9FTD/ALS.” *Neuron* 77(4): 639–46.
<http://dx.doi.org/10.1016/j.neuron.2013.02.004>.
- Aughey, Gabriel N., and Tony D. Southall. 2016. “Dam It’s Good! DamID Profiling of Protein-DNA Interactions.” *Wiley Interdisciplinary Reviews: Developmental Biology* 5(1): 25–37.
- Berson, Amit et al. 2017. “TDP-43 Promotes Neurodegeneration by Impairing Chromatin Remodeling.” *Current Biology*: 3579–90.
- Boeynaems, Steven et al. 2017. “Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics.” *Molecular Cell* 65(6): 1044-1055.e5.
<http://linkinghub.elsevier.com/retrieve/pii/S1097276517301284>.
- Boija, Ann et al. 2018. “Transcription Factors Activate Genes through the Phase-Separation Capacity of Their Activation Domains.” *Cell*: 1–14.
<https://linkinghub.elsevier.com/retrieve/pii/S0092867418313989>.
- Boillee, S. et al. 2006. “Onset and Progression in Inherited ALS Determined by Motor Neurons and Microglia.” *Science* 312(5778): 1389–92.
<http://www.sciencemag.org/cgi/doi/10.1126/science.1123511>.
- Bonn, Stefan et al. 2012. “Cell Type-Specific Chromatin Immunoprecipitation from Multicellular Complex Samples Using BiTS-Chip.” *Nature Protocols* 7(5): 978–94.
<http://dx.doi.org/10.1038/nprot.2012.049>.
- van den Brink, Susanne C et al. 2017. “Single-Cell Sequencing Reveals Dissociation-Induced Gene Expression in Tissue Subpopulations.” *Nature Methods* 14(10): 935–36.
- Cammack, Alexander J., Arnav Moudgil, et al. 2019. “A Viral Toolkit for Recording

- Transcription Factor-DNA Interactions in Live Mouse Tissues.” *BioRxiv [Preprint]* (<https://doi.org/10.1101/538504>).
- Cammack, Alexander J., Nazem Atassi, et al. 2019. “Prospective Natural History Study of C9orf72 ALS Clinical Characteristics and Biomarkers.” *Neurology* 93(17): e1605–17.
- Chew, Jeannie et al. 2015. “C9ORF72 Repeat Expansions in Mice Cause TDP-43 Pathology, Neuronal Loss, and Behavioral Deficits.” *Science (New York, N.Y.)* 348(6239): 1151–54.
- Corradin, Olivia, and Peter C. Scacheri. 2014. “Enhancer Variants: Evaluating Functions in Common Disease.” *Genome Medicine* 6(10): 1–14.
- Dao, Thuy P. et al. 2019. “ALS-Linked Mutations Affect UBQLN2 Oligomerization and Phase Separation in a Position- and Amino Acid-Dependent Manner.” *Structure* 27(6): 937-951.e5.
- Deal, Roger B., and Steven Henikoff. 2010. “A Simple Method for Gene Expression and Chromatin Profiling of Individual Cell Types within a Tissue.” *Developmental Cell* 18(June 15): 1030–40. <http://dx.doi.org/10.1016/j.devcel.2010.05.013>.
- DeJesus-Hernandez, Mariely et al. 2011. “Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS.” *Neuron* 72(2): 245–56. <http://dx.doi.org/10.1016/j.neuron.2011.09.011>.
- Donnelly, Christopher J. et al. 2013. “RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention.” *Neuron* 80(2): 415–28. <http://dx.doi.org/10.1016/j.neuron.2013.10.015>.
- Ernst, Jason et al. 2011. “Mapping and Analysis of Chromatin State Dynamics in Nine Human Cell Types.” *Nature* 473(7345): 43–49. <http://dx.doi.org/10.1038/nature09906>.
- Ferraiuolo, L. et al. 2016. “Oligodendrocytes Contribute to Motor Neuron Death in ALS via

- SOD1 Dependent Mechanism.” *Proceedings of the National Academy of Sciences*: 1–10.
- Gendron, Tania F. et al. 2013. “Antisense Transcripts of the Expanded C9ORF72 Hexanucleotide Repeat Form Nuclear RNA Foci and Undergo Repeat-Associated Non-ATG Translation in C9FTD/ALS.” *Acta Neuropathologica* 126(6): 829–44.
- . 2017. “Poly(GP) Proteins Are a Useful Pharmacodynamic Marker for C9ORF72-Associated Amyotrophic Lateral Sclerosis.” *Science Translational Medicine* 9(383). http://stm.sciencemag.org/content/9/383/eaai7866?utm_campaign=toc_stm_2017-03-29&et rid=291001628&et cid=1243883.
- Gendron, Tania F, and Leonard Petrucelli. 2017. “Disease Mechanisms of C9ORF72 Repeat Expansions.” *Cold Spring Harbor Laboratory Press*: 1–22.
- Gibson, Bryan A et al. 2019. “Organization of Chromatin by Intrinsic and Regulated Phase Separation.” *Cell* 179(2): 470-484.e21. <http://www.ncbi.nlm.nih.gov/pubmed/31543265>.
- Girdhar, Kiran et al. 2018. “Cell-Specific Histone Modification Maps in the Human Frontal Lobe Link Schizophrenia Risk to the Neuronal Epigenome.” *Nature Neuroscience* 21(8): 1126–36. <http://dx.doi.org/10.1038/s41593-018-0187-0>.
- Gogol-Doring, Andreas et al. 2016. “Genome-Wide Profiling Reveals Remarkable Parallels between Insertion Site Selection Properties of the MLV Retrovirus and the PiggyBac Transposon in Primary Human CD4+ T Cells.” *Molecular Therapy* 24(3): 592–606. <http://dx.doi.org/10.1038/mt.2016.11>.
- Graybuck, Lucas T et al. 2019. “Prospective, Brain-Wide Labeling of Neuronal Subclasses with Enhancer-Driven AAVs.” *bioRxiv [Preprint]*: doi.org/10.1101/525014. <https://www.biorxiv.org/content/early/2019/01/20/525014>.
- Haidet-Phillips, Amanda M et al. 2011. “Astrocytes from Familial and Sporadic ALS Patients

- Are Toxic to Motor Neurons.” *Nature Biotechnology* 29(9): 824–28.
- Hannon, Eilis, Sarah J Marzi, Leonard S Schalkwyk, and Jonathan Mill. 2019. “Genetic Risk Variants for Brain Disorders Are Enriched in Cortical H3K27ac Domains.” *Molecular Brain* 12(7): 1–6.
- Hao, Zongbing et al. 2019. “Motor Dysfunction and Neurodegeneration in a C9orf72 Mouse Line Expressing Poly-PR.” *Nature Communications* 10(1): 2906.
<http://www.nature.com/articles/s41467-019-10956-w>.
- Hilton, Isaac B. et al. 2015. “Epigenome Editing by a CRISPR-Cas9-Based Acetyltransferase Activates Genes from Promoters and Enhancers.” *Nature Biotechnology* 33(5): 510–17.
- Hnisz, Denes et al. 2013. “Super-Enhancers in the Control of Cell Identity and Disease.” *Cell* 155(4): 934–47. <http://linkinghub.elsevier.com/retrieve/pii/S0092867413012270>.
- . 2017. “A Phase Separation Model for Transcriptional Control.” *Cell* 169: 13–23.
<http://dx.doi.org/10.1016/j.cell.2017.02.007>.
- Jiang, Jie et al. 2016. “Gain of Toxicity from ALS/FTD-Linked Repeat Expansions in C9ORF72 Is Alleviated by Antisense Oligonucleotides Targeting GGGGCC-Containing RNAs.” *Neuron* 90(3): 535–50. <http://dx.doi.org/10.1016/j.neuron.2016.04.006>.
- Jiang, Li et al. 2017. “NEAT1 Scaffolds RNA-Binding Proteins and the Microprocessor to Globally Enhance Pri-MiRNA Processing.” *Nature Structural and Molecular Biology* 24(10): 816–24.
- Jüttner, Josephine et al. 2019. “Targeting Neuronal and Glial Cell Types with Synthetic Promoter AAVs in Mice, Non-Human Primates and Humans.” *Nature Neuroscience*.
- Kang, Shin H. et al. 2013. “Degeneration and Impaired Regeneration of Gray Matter Oligodendrocytes in Amyotrophic Lateral Sclerosis.” *Nature Neuroscience* 16(5): 571–79.

<http://dx.doi.org/10.1038/nn.3357>.

- Korb, Erica et al. 2015. “BET Protein Brd4 Activates Transcription in Neurons and BET Inhibitor Jq1 Blocks Memory in Mice.” *Nature Neuroscience* 18(10): 1464–73.
- Lee, Kyung-Ha et al. 2016. “C9orf72 Dipeptide Repeats Impair the Assembly, Dynamics, and Function of Membrane-Less Organelles.” *Cell* 167(3): 774-788.e17.
<http://linkinghub.elsevier.com/retrieve/pii/S0092867416313836>.
- Lehmer, Carina et al. 2017. “Poly-GP in Cerebrospinal Fluid Links C9orf72-Associated Dipeptide Repeat Expression to the Asymptomatic Phase of ALS/FTD.” *EMBO Molecular Medicine*: 1–10.
- Liu, Yuanjing et al. 2016. “C9orf72 BAC Mouse Model with Motor Deficits and Neurodegenerative Features of ALS/FTD.” *Neuron* 90: 1–14.
<http://dx.doi.org/10.1016/j.neuron.2016.04.005>.
- Löscher, Wolfgang. 2011. “Critical Review of Current Animal Models of Seizures and Epilepsy Used in the Discovery and Development of New Antiepileptic Drugs.” *Seizure* 20(5): 359–68. <http://dx.doi.org/10.1016/j.seizure.2011.01.003>.
- Majounie, Elisa et al. 2012. “Frequency of the C9orf72 Hexanucleotide Repeat Expansion in Patients with Amyotrophic Lateral Sclerosis and Frontotemporal Dementia: A Cross-Sectional Study.” *Lancet Neurology* 11: 323–30.
[http://www.thelancet.com/pdfs/journals/lanneur/PIIS1474-4422\(12\)70043-1.pdf](http://www.thelancet.com/pdfs/journals/lanneur/PIIS1474-4422(12)70043-1.pdf) (May 3, 2015).
- Mann, Jacob R. et al. 2019. “RNA Binding Antagonizes Neurotoxic Phase Transitions of TDP-43.” *Neuron* 0(0): 1–18. [https://www.cell.com/neuron/fulltext/S0896-6273\(19\)30075-3?platform=hootsuite](https://www.cell.com/neuron/fulltext/S0896-6273(19)30075-3?platform=hootsuite).

- Maurano, Matthew T. et al. 2012. "Systematic Localization of Common Disease-Associate Variation in Regulatory DNA." *Science* 337(September): 1190–95.
- Mo, Alisa et al. 2015. "Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain." *Neuron* 86(6): 1369–84. <http://dx.doi.org/10.1016/j.neuron.2015.05.018>.
- Molliex, Amandine et al. 2015. "Phase Separation by Low Complexity Domains Promotes Stress Granule Assembly and Drives Pathological Fibrillization." *Cell* 163(1): 123–33. <http://dx.doi.org/10.1016/j.cell.2015.09.015>.
- Mori, Kohji et al. 2013. "The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTL/ALS." *Science* 339(March): 1335–39. <http://www.sciencemag.org/content/suppl/2013/02/06/science.1232927.DC1.full> (May 7, 2015).
- Moudgil, Arnav et al. 2019. "Self-Reporting Transposons Enable Simultaneous Readout of Gene Expression and Transcription Factor Binding in Single Cells." *BioRxiv [Preprint]* (<https://doi.org/10.1101/538553>).
- Nagai, Makiko et al. 2007. "Astrocytes Expressing ALS-Linked Mutated SOD1 Release Factors Selectively Toxic to Motor Neurons." *Nature Neuroscience* 10(5): 615–22.
- O'Rourke, Jacqueline G. et al. 2015. "C9orf72 BAC Transgenic Mice Display Typical Pathologic Features of ALS/FTD." *Neuron* 88(5): 892–901. <http://dx.doi.org/10.1016/j.neuron.2015.10.027>.
- Patel, Avinash et al. 2015. "A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation." *Cell* 162(5): 1066–77. <http://dx.doi.org/10.1016/j.cell.2015.07.047>.
- Peng, Zechun, and Carolyn R. Houser. 2005. "Temporal Patterns of Fos Expression in the

- Dentate Gyrus after Spontaneous Seizures in a Mouse Model of Temporal Lobe Epilepsy.”
Journal of Neuroscience 25(31): 7210–20.
- Peters, Owen M et al. 2015. “Human C9ORF72 Hexanucleotide Expansion Reproduces RNA Foci and Dipeptide Repeat Proteins but Not Neurodegeneration in BAC Transgenic Mice.”
Neuron 88(5): 902–9. <http://dx.doi.org/10.1016/j.neuron.2015.11.018>.
- Qi, Zongtai et al. 2017. “An Optimized, Broadly Applicable *PiggyBac* Transposon Induction System.” *Nucleic Acids Research*: gkw1290. <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkw1290>.
- Renton, Alan E. et al. 2011. “A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD.” *Neuron* 72(2): 257–68.
- Rogawski, Michael A. 2006. “Diverse Mechanisms of Antiepileptic Drugs in the Development Pipeline.” *Epilepsy Research* 69(3): 273–94.
- Rossaert, Elisabeth et al. 2019. “Restoration of Histone Acetylation Ameliorates Disease and Metabolic Abnormalities in a FUS Mouse Model.” *Acta Neuropathologica Communications* 7(1): 1–19.
- Sabari, Benjamin R. et al. 2018. “Coactivator Condensation at Super-Enhancers Links Phase Separation and Gene Control.” *Science* 361(6400): eaar3958.
<http://science.sciencemag.org/>.
- Sheu-Gruttadauria, Jessica, and Ian J. MacRae. 2018. “Phase Transitions in the Assembly and Function of Human MiRISC.” *Cell* 173(4): 946-957.e16.
<https://doi.org/10.1016/j.cell.2018.02.051>.
- Shi, Yingxiao et al. 2018. “Haploinsufficiency Leads to Neurodegeneration in C9ORF72 ALS/FTD Human Induced Motor Neurons.” *Nature Medicine* (January 2017).

<http://www.nature.com/doi/10.1038/nm.4490>.

Slattery, Matthew et al. 2011. “Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins.” *Cell* 147(6): 1270–82.

<http://dx.doi.org/10.1016/j.cell.2011.10.053>.

Southall, Tony D. et al. 2013. “Cell-Type-Specific Profiling of Gene Expression and Chromatin Binding without Cell Isolation: Assaying RNA Pol II Occupancy in Neural Stem Cells.”

Developmental Cell 26(1): 101–12. <http://dx.doi.org/10.1016/j.devcel.2013.05.020>.

Stavreva, Diana A. et al. 2015. “Dynamics of Chromatin Accessibility and Long-Range

Interactions in Response to Glucocorticoid Pulsing.” *Genome Research* 25(6): 845–57.

Su, Zhaoming et al. 2014. “Discovery of a Biomarker and Lead Small Molecules to Target r(GGGGCC)-Associated Defects in C9FTD/ALS.” *Neuron*: 1–8.

<http://dx.doi.org/10.1016/j.neuron.2014.07.041>.

Wang, Haoyi et al. 2012. “‘Calling Cards’ for DNA-Binding Proteins in Mammalian Cells.”

Genetics 190(3): 941–49.

Wells, Alan et al. 2015. “The Anatomical Distribution of Genetic Associations.” *Nucleic Acids*

Research 43(22): 10804–20.

Wen, Xinmei et al. 2014. “Antisense Proline-Arginine RAN Dipeptides Linked to C9ORF72 -

ALS / FTD Form Toxic Nuclear Aggregates That Initiate In Vitro and In Vivo Neuronal Death.” *Neuron* 84(6): 1213–25. <http://dx.doi.org/10.1016/j.neuron.2014.12.010>.

White, M. A., C. A. Myers, J. C. Corbo, and B. A. Cohen. 2013. “Massively Parallel in Vivo Enhancer Assay Reveals That Highly Local Features Determine the Cis-Regulatory

Function of ChIP-Seq Peaks.” *Proceedings of the National Academy of Sciences* 110(29): 11952–57. <http://www.pnas.org/cgi/doi/10.1073/pnas.1307449110>.

- Yamanaka, Koji et al. 2008. “Astrocytes as Determinants of Disease Progression in Inherited Amyotrophic Lateral Sclerosis.” *Nature Neuroscience* 11(3): 251–53.
<http://www.nature.com/doi/10.1038/nn2047>.
- Yanez-Cuna, J. Omar et al. 2012. “Uncovering Cis-Regulatory Sequence Requirements for Context-Specific Transcription Factor Binding.” *Genome Research*: 2018–30.
- Zhang, Yong-Jie et al. 2019. “Heterochromatin Anomalies and Double-Stranded RNA Accumulation Underlie C9orf72 Poly(PR) Toxicity.” *Science* 363(6428): eaav2606.
<http://science.sciencemag.org/content/363/6428/eaav2606/tab-pdf>.
- Zhou, Pingzhu et al. 2017. “Mapping Cell Type-Specific Transcriptional Enhancers Using High Affinity, Lineage-Specific Ep300 BioChIP-Seq.” *eLife* 6: 1–29.
- Zu, Tao et al. 2013. “RAN Proteins and RNA Foci from Antisense Transcripts in C9ORF72 ALS and Frontotemporal Dementia.” *Proceedings of the National Academy of Sciences of the United States of America* 110(51): E4968-77.
<http://www.pnas.org/content/110/51/E4968.full> (May 3, 2015).

Curriculum Vitae
Alexander J. Cammack

Washington University in St. Louis | (409)338-6826 | ajcammack@wustl.edu

Education

Washington University in St. Louis, St. Louis, MO, USA 08/2014-present
PhD Candidate (Expected May 2020), Neurosciences

Tulane University, New Orleans, LA, USA 08/2010-05/2014
B.S. in Cell and Molecular Biology
Magna Cum Laude, with Honors in Cell and Molec. Biol, GPA: 3.78

Research Experience

“Investigating transcription factor-mediated epigenetic regulation
in the healthy brain and neurological disease” 05/2015-present
Laboratory of Dr. Timothy Miller, Washington University,
St. Louis, MO, USA
PhD Dissertation Research

“Role of the CNGB3 subunit in cone light response and adaptation” 02/2015-05/2015
Laboratory of Dr. Vladimir Kefalov, Washington University,
St. Louis, MO, USA
Graduate Research Rotation

“Circadian control of neuroinflammation and astrocyte activation” 11/2014-02/2015
Laboratory of Dr. Erik Musiek, Washington University,
St. Louis, MO, USA
Graduate Research Rotation

“Effects of ALS-associated genetic variation on miRNA-mediated
gene regulation” 08/2014-11/2014
Laboratory of Dr. Timothy Miller, Washington University,
St. Louis, MO, USA
Graduate Research Rotation

“Role of RNA•DNA hybrid formation and resolution in a FRDA
model” 06/2013-05/2014
Laboratory of Dr. Ed Grabczyk, LSUHSC,
New Orleans, LA, USA
Undergraduate Honors Thesis Research

“Role of oxygen in mammalian bone regeneration” 01/2012-05/2013
Laboratory of Dr. Ken Muneoka, Tulane University,
New Orleans, LA, USA
Undergraduate Research

Publications

Cammack AJ, ... , Miller TM. Retrospective BRD4 occupancy profiling reveals predictive vulnerability and susceptibility loci in an acute seizure model. Manuscript in preparation.

Cammack AJ, ... , Miller TM. Brain and spinal cord enhancers are misregulated in *C9orf72* ALS mouse models and human tissues and harbor ALS risk-associated non-coding genetic variation. Manuscript in preparation.

Cammack AJ, ... , Dougherty JD. A viral toolkit for recording transcription factor-DNA interactions in live mouse tissues. *PNAS*, in press. (*bioRxiv* 539504, preprint).

Cammack AJ*, Atassi N*, ... , Miller TM. Prospective natural history study of *C9orf72* ALS clinical characteristics and biomarkers. *Neurology*. 2019 Oct 2.

Simkin J, Sammarco MC, Marrero L, Dawson LA, Yan M, Tucker C, **Cammack AJ**, Muneoka K. Macrophages are required to coordinate mouse digit tip regeneration. *Development*. 2017 Nov 1.

Kanekura K, Yagi T, **Cammack AJ**, ... , Urano F. Poly-dipeptides encoded by the *C9ORF72* repeats block global protein translation. *Hum. Mol. Genetics*. 2016 May 1.

Sammarco MC, Simkin J, **Cammack AJ**, ... , Muneoka K. Hyperbaric oxygen promotes proximal bone regeneration and organized collagen composition during digit regeneration. *PLoS One*. 2015 Oct 9.

Sammarco MC, Simkin J, Fassler D, **Cammack AJ**, ... , Muneoka K. Endogenous bone regeneration is dependent upon a dynamic oxygen event. *J Bone Miner Res*. 2014 Apr 18.

Awards and Recognitions

Graduate:

O'Leary Prize competition (winner), Washington Univ.	03/2020
Neuroscience program retreat poster competition (finalist), Washington Univ.	10/2019
Neuroscience program retreat seminar competition (finalist), Washington Univ.	10/2018
Hope Center poster competition (winner), Washington Univ.	04/2018

Undergraduate:

Senior Honors Scholar in Cell and Molecular Biology, Tulane Univ.	05/2014
Best Honors Thesis Award in Cell and Molec. Biol. (winner), Tulane Univ.	05/2014
Summer research internship poster competition (winner), LSUHSC	08/2013
Judith and Morris Henkin Memorial Scholarship, Tulane Univ.	02/2013
Goldwater Scholarship (nominee), Tulane Univ.	01/2013
UAN travel award, ASBMB.	10/2012
BEAST undergraduate poster competition (winner), Tulane Univ.	10/2012
Tulane Honors Program Summer Research Grant, Tulane Univ.	05/2012

Invited talks

Retrospective transcription factor recording in the healthy brain and neurological disease” March 2020. O’Leary Prize symposium and competition, Washington University, St. Louis, MO, USA.

***selected from applications; prize winner**

“Investigating aberrant epigenetic regulation in *C9orf72* ALS/FTD” November 2019. Knight ADRC Seminar Series, Washington University, St. Louis, MO, USA.

“Transposon-mediated, cell type-specific transcription factor profiling in the mouse brain.” October 2018. Neuroscience Program Annual Retreat, Washington University, St. Louis, MO, USA. ***selected from abstracts**

“*C9orf72* G₄C₂ repeat expansions are dynamic and result in large dipeptide repeat proteins.” June 2018. Hope Center Seminar Series, Washington University, St. Louis, MO, USA. ***invited talk after poster award**

Poster presentations

Cammack AJ, et al. Mar. 2020. “Investigating aberrant epigenetic regulation in C9ALS/FTD” Packard Center for ALS Research Annual Meeting, Baltimore, MD, USA.

Cammack AJ, et al. Oct. 2019. “A viral toolkit for recording transcription factor-DNA interactions in live mouse tissues.” Society for Neuroscience Annual Meeting, Chicago, IL, USA.

Cammack AJ, et al. Oct. 2019. “Investigating aberrant epigenetic regulation in ALS/FTD with *in vivo* enhancer profiling” Neuroscience Program Annual Retreat, Washington University, St. Louis, MO, USA.

Cammack AJ, et al. Oct. 2019. “A viral toolkit for recording transcription factor-DNA interactions in live mouse tissues.” American Academy of Neurology Annual Meeting, St. Louis, MO, USA.

Cammack AJ, et al. Jul. 2019. “Transposon-mediated, cell type-specific transcription factor profiling in the mouse brain.” Gordon Research Conference, Mount Snow, VT, USA.

Cammack AJ, et al. Nov. 2018. “Transposon-mediated, cell type-specific transcription factor profiling in the mouse brain.” Society for Neuroscience Annual Meeting, San Diego, CA, USA.

Cammack AJ, et al. Apr. 2018. “*C9orf72* G₄C₂ repeat expansions in ALS patients are dynamic and result in large dipeptide repeat proteins.” Hope Center Annual Retreat, Washington University, St. Louis, MO, USA.

Cammack AJ, et al. Oct 2017. “*In vivo*, retrospective gene expression analysis in ALS motor neurons.” Neuroscience Program Annual Retreat, Washington University, St. Louis, MO, USA.

Cammack AJ, et al. May 2017. “Dipeptide repeat protein size in *C9orf72* amyotrophic lateral sclerosis.” Hope Center Annual Retreat, Washington University, St. Louis, MO, USA.

Cammack AJ, et al. Aug. 2013. “Contrasting effects of different length TTC oligomers on transcription in a Friedreich ataxia model.” LSUHSC Summer Internship Poster Competition, New Orleans, LA, USA.

Cammack AJ, et al. Apr. 2013. “Increased oxygen enhances bone degradation and promotes bone regeneration after digit amputation in mice.” Experimental Biology Annual Meeting, Boston, MA, USA.

Cammack AJ, et al. Apr. 2013. “Digits exposed to increased oxygen tension demonstrate 1) enhanced bone degradation and 2) subsequent regeneration from a regeneration-incompetent site.” Tulane University SSE Research Day, New Orleans, LA, USA.

Cammack AJ, et al. Apr. 2013. “Digits exposed to increased oxygen tension demonstrate 1) enhanced bone degradation and 2) subsequent regeneration from a regeneration-incompetent site.” Tulane University HS Research Day, New Orleans, LA, USA.

Cammack AJ, et al. Oct. 2012. “Increased oxygen tension both enhances bone degradation and augments bone regeneration *in vivo*.” Tulane University BEAST Poster Competition, New Orleans, LA, USA.

Experimental Skills

- Molecular biology techniques (molecular cloning, PCR, sequencing library preparation)
- Biochemical techniques (Western/Southern blots, detergent fractionation, column chromatography)
- Tissue/cell culture (neuronal lines, stem cell lines, immortalized lines)
- Animal methods (mouse and rat handling, stereotactic injection, survival surgery, tissue collection)
- Viral production, handling, and application to cell and animal models (AAV, lentivirus)
- Immunohistochemical techniques and microscopy (light, confocal)
- Bioinformatics/genomics (analysis of next gen. seq., including CHIP-seq, calling cards, RNA-seq)

Teaching Experience

Coursemaster and lecturer, Molecular Biol. at the Cutting-Edge Washington University, St. Louis, MO, USA Spring 2018-2020

Graduate Student Mentor to a Washington Univ. undergrad. Washington University, St. Louis, MO, USA 01/2016-05/2019

Graduate Teaching Assistant, Principles of the Nervous System Washington University, St. Louis, MO, USA 08/2015-12/2015

References

Timothy (Tim) Miller, MD, PhD

PhD advisor

miller.t@wustl.edu

Joseph (Joe) Dougherty, PhD

Collaborator and thesis committee member

jdougherty@wustl.edu

Robi (Rob) Mitra, PhD

Collaborator and thesis committee member

rmitra@wustl.edu

Conrad (Chris) Wehl, MD, PhD

Thesis committee member and monthly joint lab meetings

weihlc@wustl.edu