Epigenomic organization and activation of brain sex differences

A dissertation presented by

Bruno Gegenhuber

to the

School of Biological Sciences

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Biological Sciences

at

Cold Spring Harbor Laboratory

April 11, 2022

Abstract

The “Organizational-Activational Hypothesis”, the central dogma of neuroendocrinology, states that early-life (“organizational”) and adult (“activational”) gonadal hormone signaling drives sex differences in the mammalian brain. Early-life hormones irreversibly sculpt the development of the same neural circuits that adult hormones act upon to promote the display of sex-typical behaviors. Gonadal hormones bind nuclear hormone receptors, which principally act as transcription factors (TFs). Hence, the regulation of gene expression by hormone receptors in the brain comprises the molecular basis of the “Organizational-Activational Hypothesis”, and, accordingly, ensures the survival and propagation of most mammalian species. However, the genomic targets of gonadal hormone receptors in the brain, and their regulation during “organizational” and “activational” windows, have not been identified. Here, I utilize low-input and single-cell chromatin and transcriptomic profiling approaches to identify the genomic mechanisms specifying sex differences in the developing and adult mouse bed nucleus of the stria terminalis (BNST) – a central brain region in the limbic circuitry controlling sex-typical behaviors. I find that male-specific activation of estrogen receptor α (ERα) on the day of birth drives sustained sex differences in gene regulation in the developing brain. Subsequently, testosterone (T) largely controls sex differences in gene regulation following puberty. Of note, adult gonadectomy ablates sex differences in chromatin accessibility, while estradiol (E2) replacement after gonadectomy leads to sex-shared genomic responses, revealing brain “organization” by early-life hormone signaling is reversible at the level of gene regulation. Together, I reveal the first genomic targets of a gonadal hormone receptor in the brain and define the molecular principles of the “Organizational-Activational Hypothesis”. I provide a compendia of sex-biased enhancers, genes, and neuron types across life stages that serve the basis for future functional investigation into the genomic encoding of social behavior.

Acknowledgements

I have many to thank for helping me complete this thesis. First, I would like to thank my advisor, Jessica Tollkuhn, for providing insight, support, and discussion along the journey. She went above and beyond to support my career at every opportunity and to train me not only in research, but in navigating life in academia. I also thank current and former members of the Tollkuhn lab for their intellectual and scientific contributions to this work: Melody Wu, who helped with countless dissections, Rob Bronstein, Kat Denney, Simón(e) Sun, and Matthew Johnson.

I am grateful to my academic mentor, John Inglis, and my thesis committee chairs, Rob Martienssen and Z. Josh Huang, for their guidance, support, and mentorship. I am particularly grateful to Josh, who co-sponsored my F31 application and chose to stay on the committee after moving his lab to Duke. I also thank my other committee members, Stephen Shea and Lucas Cheadle, for supporting my career and maintaining a balance between neuroscience and molecular biology in our meetings, as well as my external committee member and collaborator, Holly Ingraham, for constantly championing my science.

I would not be able to complete this research without the support of many core facilities at CSHL, particularly Jon Preall, Claire Regan, and Brian He in the Single Cell Core; Pam Moody in the Flow Cytometry Core; and Kristin Cox and Alexander Aguilera in the LAR Core. Likewise, I would not be here today without the people running the School of Biological Sciences – Alex Gann, Alyson Kass-Eisler, Monn Monn Myat, Kim Creteur, and Kim Graham – who provided me with an education conducive to scientific success.

Lastly, I thank my partner, Rena Yan, as well as my family, classmates, and friends (perhaps Ben Harris) for their love and constant support.

This work received funding from the School of Biological Sciences at CSHL and the National Institute of Mental Health Ruth L. Kirschstein National Research Service Award.

Table of contents

List of Figures and Tables

List of Abbreviations

5hmC	5-Hydroxymethylcytosine
5mC	5-Methylcytosine
AF1	Activation function 1
AP-1	Activator protein-1
AR	Androgen receptor
ARC	Arcuate nucleus of the hypothalamus
ARE	Androgen response element
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
AVPV	Anteroventral periventricular hypothalamus
BICCN	Brain Initiative Cell Census Network
BNST	Bed nucleus of the stria terminalis
BNSTp	BNST posterior
CBP	CREB-binding protein
CGE	Caudal ganglionic eminence
ChIP-seq	Chromatin Immunoprecitation using sequencing
CREB	cAMP response element-binding protein
CRISPRa	CRISPR activation
CRISPRi	CRISPR interference
CUT&RUN	Cleavage Under Targets & Release Under Nuclease
CUT&Tag	Cleavage Under Targets & Tagmentation
DNMT	DNA methyltransferase
DO	Disease Ontology
E2	Estradiol or estradiol benzoate
ERα	Estrogen receptor α
ERβ	Estrogen receptor β
ERE	Estrogen response element
GDX	Gonadectomized
GO	Gene Ontology
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HGNC	Human Gene Nomenclature Committee
HSP90	Heat shock protein 90
LBD	Ligand-binding domain
LH	Luteinizing hormone
MBH	Medial-basal hypothalamus
Mc4r	Melanocortin receptor 4
mCA	Methyl CA
mCG	Methyl CG
MeA	Medial amygdala
MeApd	MeA posterior dorsal
MeApv	MeA posterior ventral
MGE	Medial ganglionic eminence
mGluR	Metabotropic glutamate receptor
MPOA	Medial pre-optic area of the hypothalamus
MSH	Melanocyte-stimulating hormone
PF	Pioneer factor
POA	Pre-optic area

POMC	Pro-opiomelanocortin
PR	Progesterone receptor
PTM	Post-translational modification
PVH	Paraventricular hypothalamus
scRNA-seq	Single-cell RNA-seq
SDN-POA	Sexually dimorphic nucleus of the POA
SNB	Spinal nucleus of the bulbocavernosus
snRNA-seq	Single-nucleus RNA-seq
SP1	Specificity protein 1
SRC	Steroid receptor coactivator
T	Testosterone
TET	Tet-eleven translocation
TF	Transcription factor
TH	Tyrosine hydroxylase
TSS	Transcription start site
VMH	Ventromedial hypothalamus
VMHvl	VMH ventrolateral
WGBS	Whole genome bisulfite sequencing

Chapter 1

Introduction to sex differences in the brain

Sections of this chapter were previously published in Genes (2019) under the title "Sex Differences in the Epigenome: A Cause or Consequence of Sexual Differentiation of the Brain?" by Bruno Gegenhuber and Jessica Tollkuhn and in WIREs Developmental Biology (2020) under the title "Signatures of sex: Sex differences in gene expression in the vertebrate brain" by Bruno Gegenhuber and Jessica Tollkuhn.

1.1 Introduction

Sex differences in the brain arise from contributions of gonadal steroid hormones. The classic "Organization and Activation Hypothesis", first articulated sixty years ago, states that early-life hormone signaling specifies sex differences in the brain, which are later activated by adult hormones to regulate sex-typical reproductive and territorial behaviors (McCarthy, Wright, et al., 2009; Phoenix et al., 1959). At birth, the testes briefly activate, releasing a surge of testosterone that undergoes local conversion to E2 in the brain by aromatase, a P450 enzyme (Balthazart & Ball, 1998; Clarkson & Herbison, 2016). E2 binds its canonical receptors estrogen receptor alpha and beta (ERα/β), which are nuclear receptor TFs that bind DNA in response to ligand (McKenna & O'Malley, 2002). E2, and other estrogens, can also act rapidly at the neuronal membrane to increase firing (Heimovics et al., 2015; Micevych & Kelly, 2012; Stincic et al., 2018; Woolley, 2007), potentially initiating activity-dependent transcriptional programs that differ from those directed by nuclear ERα.

ERα is considered the master regulator of sexual differentiation of the rodent brain. In mice, genetic deletion of this receptor attenuates male-typical sexual and territorial behaviors and feminizes

the expression of ERβ and AR, the receptor for testosterone (Ogawa et al., 1997; Rissman et al., 1999; Scordalakes & Rissman, 2003; Wersinger et al., 1997; M. V. Wu & Tollkuhn, 2017a). Treating female mice or rats at birth with E2 effectively masculinizes the brain, affecting cell number and neuronal wiring days after the hormone surge has subsided (McCarthy, 2008a; Turano et al., 2018). For instance, in males, neonatal E2 production induces neurite outgrowth and promotes cell survival in certain brain areas, such as the medial amygdala (MeA) and BNST (Forger et al., 2004b; M. V. Wu et al., 2009a), while simultaneously initiating apoptosis in the anteroventral periventricular hypothalamus (AVPV), a region that regulates ovulation in females (Forger et al., 2004b; Kelly et al., 2013). These neurodevelopmental events contribute to sex differences in the display of innate behaviors. For instance, males and females engage in different mating routines, and only males urine-mark their territory and aggressively defend it, although lactating females also attack intruding conspecifics to defend their offspring (Beach, 1981; Hashikawa et al., 2018; Ishii & Touhara, 2019; Lonstein & Gammie, 2002). Aside from mating and aggression, there are extensive sex differences in stress responses and motivated behaviors, for which the underlying molecular and cellular mechanisms remain poorly understood (Bangasser & Wicks, 2017; Becker & Chartoff, 2019; Laman-Maharg & Trainor, 2017).

A parallel logic exists for sexual differentiation of the primate brain, although less is known about this process. Testosterone is considered to be the principal regulator of sexual differentiation of the primate brain (Hines, 2008). The testes become active around week 7 of human gestation and secrete testosterone until around week 24, after which they become largely quiescent until birth (Figure 1.1) (Hines, 2006b; Reyes et al., 1974). Similar to rodents, there is no hormonal activity of the ovaries at this time, although fetuses of both sexes are exposed to estrogens and progesterone through the maternal circulation. Testosterone level sharply rises again during infancy then drops until the onset of puberty. Behavioral studies in children demonstrate an association between mid-gestation testosterone level and male-typical toy preferences and social behaviors (Hines, 2006a; Hines et al., 2016). Both testosterone

surges coincide with major developmental events in the brain: the prolonged fetal bolus overlaps with the peak of neurogenesis and onset of neuronal migration, whereas the infant testosterone surge occurs during synaptogenesis and myelination (Figure 1.1). Similar to the rodent brain, it remains unclear how testosterone intersects with these developmental processes to direct male-typical brain development, although advancements in human brain organoid culture systems may lead to mechanistic insights (Kelava et al., 2022).

Here, I review neural circuits and molecular mechanisms that define sex differences in the rodent brain. I then discuss mechanisms by which steroid hormone receptors regulate gene expression, primarily informed by non-neuronal *in vitro* systems, and highlight prior studies examining sex differences in the neuronal epigenome. Lastly, I introduce fundamental principles of neuroendocrinology, with a focus on understanding molecular, cellular, and transcriptomic sex differences during early life, puberty, and adulthood.

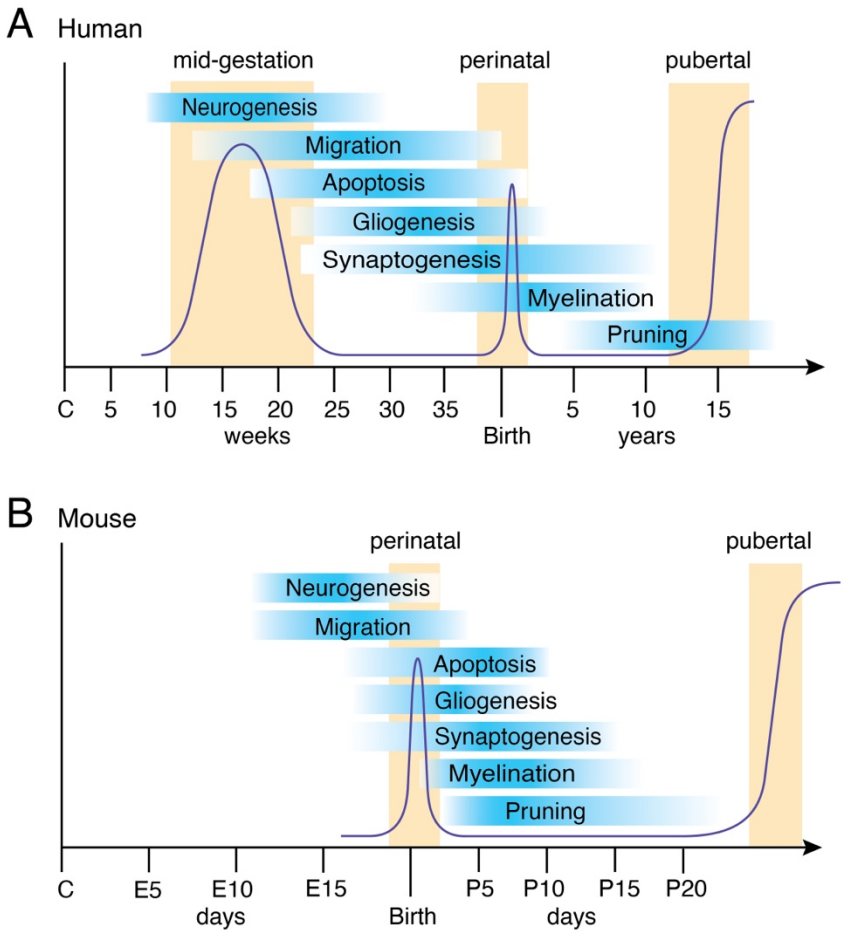

Figure 1.1: Intersection of testosterone surges with neurodevelopmental events.
Male humans (**A**) and mice (**B**) experience developmental T surges that delineate critical periods for sexual differentiation of the brain. Human testes secrete T at approximately week 8 of gestation, peaking around week 16, and then declining until week 24. Fetal T levels are nearly as high as those during puberty. There is also elevated T during months 1-3 of infancy. Mouse testes produce a T surge on the day of birth. The testes are then inactive until puberty, which begins around postnatal week 4. At birth, the mouse brain is still undergoing neuronal migration, particularly of inhibitory interneurons, while gliogenesis and synaptogenesis are also increasing. Human brain development occurs on a much longer timescale with the bulk of neuronal proliferation occurring during the fetal T surge. Events that occur during T surges are thought to give rise to sex differences in brain function, as female ovaries do not undergo significant hormone release until puberty.

1.2 Hormone receptor expression in the brain

To understand how sex differences in the brain arise, it is essential to know which brain regions and cell types express gonadal hormone receptors. Hormone receptors are nuclear transcription factors (TFs) that directly bind DNA to regulate gene expression: progesterone binds to progesterone receptor (PR/*Pgr*), testosterone to androgen receptor (AR/*Ar*) and estrogens to estrogen receptors alpha and beta (ERα/*Esr1*, ERβ/*Esr2*). Steroid hormones travel via blood circulation and diffuse through cell membranes to bind their receptors, which upon activation, undergo conformational changes to enable DNA binding. Therefore, differential gonadal hormone levels between sexes may lead to sex differences in gene regulation within the brain and other tissues. Importantly, while analysis of postmortem human brain tissue has revealed the expression of hormone receptors and aromatase in adulthood (Azcoitia et al., 2011; Yague et al., 2006), it is still not known when and where the genes encoding these factors are expressed during brain development. The BrainSpan database reports almost undetectable levels (RPKM <1) for the four gonadal hormone receptors; however, this is likely due to their low and sparse expression pattern in the brain.

In contrast, hormone receptor expression is well-characterized in the developing and adult rodent brain (Manoli & Tollkuhn, 2018b; Stincic et al., 2018). In vertebrates, all four receptors are highly expressed in limbic brain regions, such as the MeA, BNST, medial pre-optic area of the hypothalamus (MPOA) and the ventromedial hypothalamus (VMH) (Fig. 1.2). These regions comprise an interconnected network that responds to chemosensory cues from conspecifics to regulate innate reproductive and territorial behaviors, often referred to as the vomeronasal network or social behavior network (Y. Li & Dulac, 2018; Newman, 1999; O'Connell & Hofmann, 2011). These receptors are also sparsely expressed in the hippocampus as well as areas that mediate motivated behaviors (Manoli & Tollkuhn, 2018a), in addition to specific cortical layers and subregions (M. V. Wu & Tollkuhn, 2017a; Zuloaga et al., 2014b). While ERα expression is most abundant in the hypothalamus, ERβ is the primary estrogen receptor expressed in cortex (Clemens et al., 2019; González et al., 2007; Kritzer, 2002;

Shughrue et al., 1997). ERβ mutant animals have deficits in cortical development, consistent with a role for ERβ in cell survival and migration (Fan et al., 2010; Panda et al., 2018; L. Wang et al., 2003).

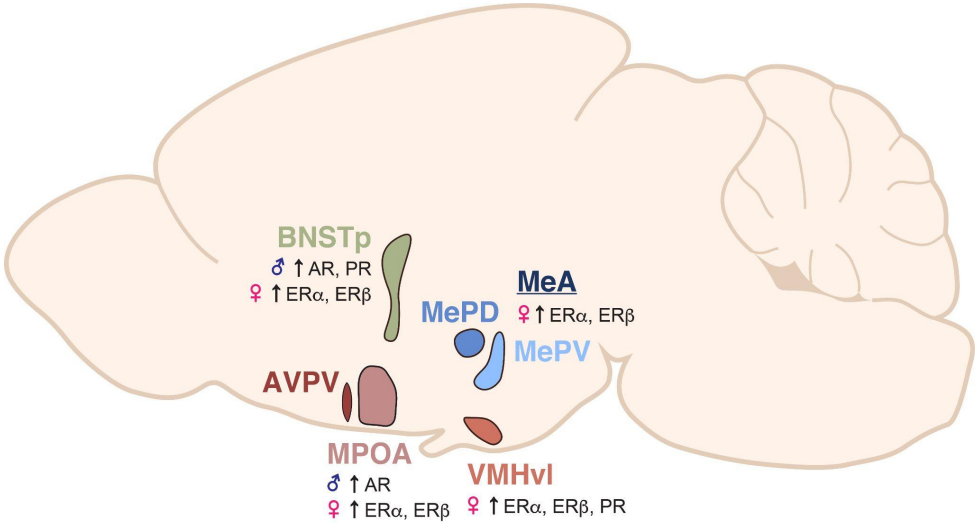

Figure 1.2: Sex differences in steroid hormone receptor expression in the mouse brain. Sex differences in steroid hormone receptor expression have been reported in several brain areas involved in innate social behaviors. These areas include the AVPV, MeA posterior dorsal (MeApd) and posterior ventral (MeApv) regions, MPOA, BNST posterior (BNSTp) region, and VMHvl. AR androgen receptor; ERα estrogen receptor α; ERβ estrogen receptor β; PR progesterone receptor.

1.3 Mechanisms of hormone receptor gene regulation

Prior to this thesis, the genomic targets of gonadal hormone receptors in the brain had not been determined, due largely to their low and sparse expression. Thus, our knowledge of hormone receptor gene regulatory mechanisms in the brain derives from human cancer cell lines and peripheral mouse tissues, such as liver and uterus, from which sufficient material can be collected for chromatin immunoprecipitation sequencing (ChIP-seq) (Carroll et al., 2005; Droog et al., 2016; Gertz et al., 2013; Gordon et al., 2014; Hewitt et al., 2012; Holding et al., 2018; Hurtado et al., 2011; Métivier et al., 2003; Palierne et al., 2016; Swinstead et al., 2016). In this section, I discuss mechanisms of ERα, as this receptor has been studied extensively in multiple biological contexts and is the master regulator of brain sexual differentiation.

Nuclear hormone receptors are activated by hormone binding to their ligand-binding domain (LBD), which facilitates receptor dimerization, followed by either tethering to other transcription factors or direct DNA binding to a consensus motif sequence (Brzozowski et al., 1997; Maggi, 2011). Extracellular molecules, such as growth factors, cytokines, and peptide hormones, have also been shown to activate hormone receptors through ligand-independent signaling pathways (Bennesch & Picard, 2015). For instance, dopamine activates ERα in certain *in vitro* systems (Gangolli et al., 1997; Power et al., 1991). Typically, ligand-independent signaling involves activation of intracellular kinases, such as ERK, AKT, or PKA, which subsequently phosphorylate serine or tyrosine residues in the ERα activation function 1 (AF1) domain and release the receptor from its inhibitory complex with heat shock protein 90 (HSP90).

Following ligand-binding, hormone receptors associate with coactivators and core transcriptional machinery and bind directly to specific sequence motifs at gene regulatory elements – ERs bind the estrogen response element (ERE) and AR binds the androgen response element (ARE) (Fig. 1.3). Historically, investigators wishing to identify hormone-responsive genes performed bioinformatic searches for EREs/AREs in the vicinity of gene promoters. However, more recently,

genome-wide approaches have revealed that fewer than 10% of receptor binding sites in the genome are at promoters and that receptors are often recruited to DNA that is co-bound by other transcription factors associating with their own canonical binding motifs at distal or intronic enhancers (Fig. 1.3). Specificity protein 1 (SP1) and activator protein-1 (AP-1), which consists of a FOS/JUN heterodimer, can also tether nuclear hormone receptors to the genome (Safe & Kim, 2008). Hormone receptors may also regulate transcription independent of direct DNA binding by controlling the phosphorylation of cAMP binding protein (CREB) – a factor that regulates *Fos* and other genes in response to neural activity (Boulware et al., 2005; Fix et al., 2004; Nguyen et al., 2009; Shaywitz & Greenberg, 1999). In primary hippocampal culture and striatal neurons, ERα localized to the postsynaptic membrane signals through group 1 or 2 metabotropic glutamate receptors (mGluRs) to promote or inhibit CREB phosphorylation, respectively (Figure 1.3) (Boulware et al., 2005; Grove-Strawser et al., 2010).

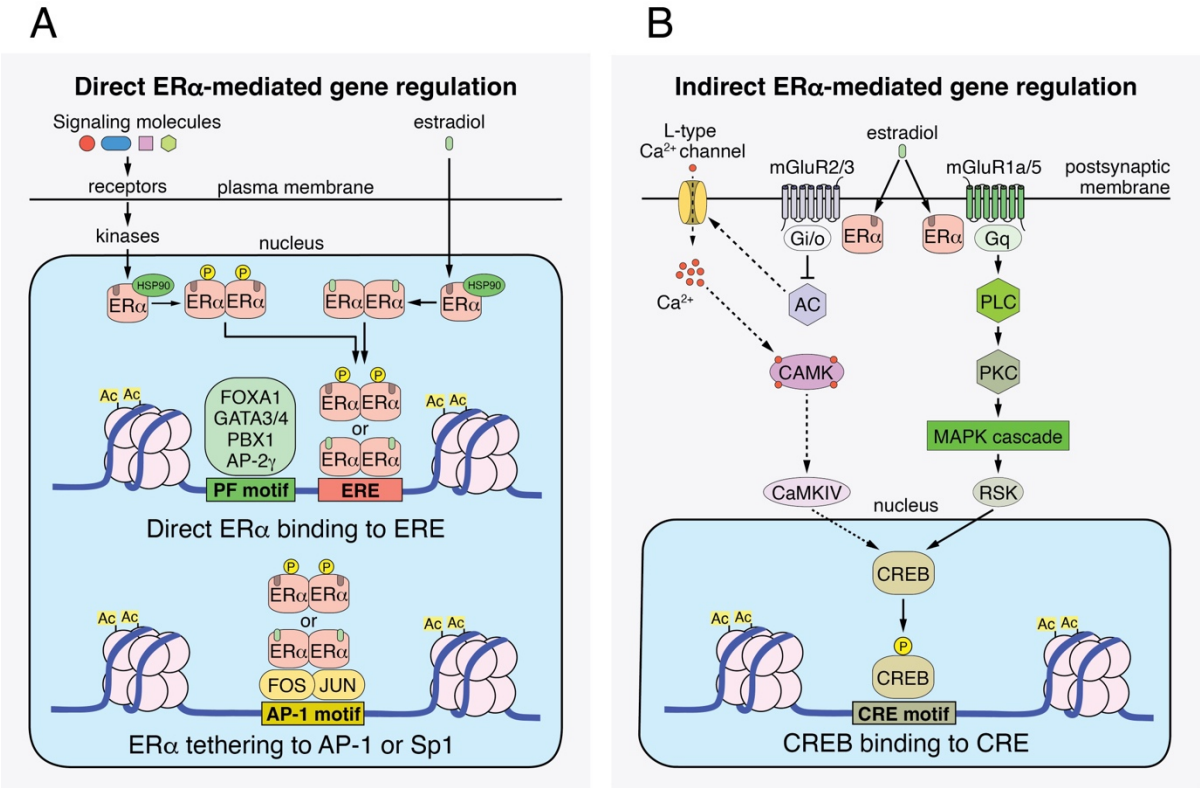

Figure 1.3: Mechanisms of ERα-mediated gene regulation.
ERα acts through different binding partners and intracellular signaling pathways to regulate gene expression in a tissue- and cell type-specific manner. (**A**) Following ligand binding, ERα can directly influence transcription by binding canonical EREs, activator protein 1 AP-1 transcription factor complex, or pioneer factors (PFs), which infiltrate heterochromatin and impart cell type-specificity to ERα genomic binding. ERα also receives PTMs from diverse signaling pathways independently of ligand. Many extracellular molecules, such as growth factors, cytokines, and peptide hormones, can initiate ligand-independent activation of ERα. (**B**) ERα can also indirectly regulate gene expression by controlling the phosphorylation of CREB. Phosphorylated CREB binds CREB-binding protein (CBP), a transcriptional coactivator that serves a dual function in gene regulation as a histone acetyltransferase and scaffold for transcriptional machinery.

1.4 Organization of neural circuits by steroid hormones

Although there are few examples of direct genomic targets of gonadal hormone receptors in the brain, there is ample evidence that neonatal estradiol establishes sex differences in the brain (Forger et al., 2016). Four regions of the rodent brain are consistently described as sexually dimorphic in cell number, and each are all involved in reproductive behaviors. The AVPV, MPOA, BNSTp, and MeApd all contain more cells, occupy a larger volume, and exhibit more neural projections in males than in females. In contrast, the AVPV controls the release of luteinizing hormone (LH) from the pituitary, which promotes ovulation in females, and has more cells in females than in males as a result of the neonatal surge. Many of these neurons express tyrosine hydroxylase (TH), an enzyme that is necessary for dopamine synthesis, and the number of TH+ neurons is also higher in mothers than in virgin females or males (Scott et al., 2015; Simerly et al., 1985). Conversely, the MPOA is larger in males, and male rats, in particular, have a pronounced sexually dimorphic nucleus (SDN) of the POA (SDN-POA) – the first sexually dimorphic brain region to be identified in mammals (Gorski et al., 1978). The MeApd and BNSTp are also larger in males (Hines et al., 1992a; Raznahan et al., 2015), and a male-bias in BNSTp volume has been described in humans (Allen & Gorski, 1990a). A post-mortem study of cisgender men, cisgender women, and transgender women found that BNSTp volume is similar in cisgender and transgender women, which is approximately 50% of the volume in cisgender men (J. N. Zhou et al., 1995). BNSTp size was not influenced by adult hormone status or sexual preference. Although this study was performed with small sample sizes (42 total subjects), this result suggests gender identity may be influenced by the organizational effects of developmental hormone signaling (Bao & Swaab, 2011; Hines, 2010).

The role of steroid hormones in promoting postnatal neuronal survival was first observed in the spinal nucleus of the bulbocavernosus (SNB), which innervates penile muscles and contains more motor neurons in males than in females (Forger et al., 1992). Females typically lose up to 70% of SNB motor neurons, and females treated with testosterone during the perinatal period showed a reduction in

cell death and male-typical cell numbers at P10 (Nordeen et al., 1985). The BNSTp, MPOA and MeApd all contain higher numbers of dying cells in postnatal females compared to males, and perinatal E2 masculinizes cell number in these regions (M. V. Wu et al., 2009a). Abolishing programmed cell death by deletion of the pro-apoptotic gene *Bax* leads to similar cell numbers in the BNSTp of females and males (Forger et al., 2004a). Some of these cells express aromatase, as adult females have fewer aromatase-expressing neurons in the BNSTp than males (M. V. Wu et al., 2009a). Conversely, in the AVPV, there is increased cell death in males, accompanied by reduced expression of the cell survival signal *Bcl-2* (Tsukahara et al., 2006), revealing region-specific responses to neonatal estradiol signaling (Krishnan et al., 2009). Intriguingly, similar patterns of sex-specific cell death occur in the developing brains of female *Drosophila* and hermaphroditic *C. elegans*, suggesting conserved logic controlling the specification of sex-typical neural circuitry across certain branches of the animal kingdom (Conradt & Horvitz, 1999; Kimura et al., 2005; Peden et al., 2007; Sanders & Arbeitman, 2008; Schwartz & Horvitz, 2007).

What is the identity of surviving neurons that give rise to sexual dimorphism in the mammalian brain? In 2018, the Dulac lab made unprecedented progress toward characterizing the cellular repertoire of the POA. By combining single-cell RNA-sequencing (scRNA-seq) with multiplexed error-robust fluorescent in situ hybridization (MERFISH), Moffit, Bambah-Mukku and colleagues created a molecular and functional cell atlas of the POA (Moffitt et al., 2018a). The authors obtained single-cell gene expression profiles for ~31,000 POA cells and integrated these data with the spatial expression of 155 marker genes across ~1 million POA cells to identify 70 neuronal populations. Moreover, by co-labeling for *Fos* in the MERFISH assay, the authors determined which of these populations are active during mating, aggression, and parenting behaviors. In particular, one aromatase-expressing inhibitory neuron type in the SDN-POA was activated in males, but not in females, during all three behaviors. This type expressed marker genes *Moxd1* (Tsuneoka et al., 2017) and *Cplx3*, a complexin family member that regulates phasic exocytosis of neurotransmitters (Mortensen et al., 2016). Taken together,

these results suggest that sexual dimorphism in the SDN-POA is defined by the co-expression of these two markers, gonadal hormone receptors, and aromatase. In the future, it will be important to determine how this population arises during development and whether it plays a causal role in the display of male-typical behaviors.

In addition to controlling cell number, early-life gonadal hormones organize neural circuits by regulating axon guidance and outgrowth throughout the social behavior network, giving rise to sex differences in neuronal connectivity. For instance, neurons projecting from the BNSTp to the AVPV have more fibers in males than in females; this connection develops prior to postnatal day 10 (P10) (Hutton et al., 1998) and requires neonatal E2 (G. Gu et al., 2003). Co-culturing neonatal BNSTp and AVPV explants demonstrated that the male, but not female, AVPV drives BNSTp axon outgrowth, suggesting regulation by target-secreted axon guidance molecules (Ibanez et al., 2001). Male BNSTp neurons also have more fibers projecting to the ventral premammillary nucleus (PMV) and MeA than female neurons (G. Gu et al., 2003). Unlike the BNSTp, the MPOA has few sex differences in the strength of its projections, as determined from classic fiber tract-tracing or modern viral approaches (Kohl et al., 2018; Simerly & Swanson, 1988); however, it is possible that targeting specific neuron types within the MPOA may reveal populations with sexually dimorphic projections.

1.5 Sex differences in the epigenome

I begin this section by defining and distinguishing between the terms “epigenetic” and “epigenomic”, in line with the writings of Mark Ptashne (Ptashne, 2007, 2013), Steven Henikoff and John Greally (Henikoff & Greally, 2016). The “epigenome” refers to chemical and/or structural modifications to the genome that influence gene expression, such as DNA methylation, post-translational modifications (PTMs) to histone proteins, and higher-order chromatin organization, such as topologically associating domains (TADs) (Gibcus & Dekker, 2013; Jaenisch & Bird, 2003; Shilatifard, 2006). “Epigenetic”, as coined by C.H. Waddington in 1942, refers to any self-perpetuating mechanism by which a cell retains

a memory of a transient event. While many, but not all, epigenetic processes involve the epigenome, it is not always the case that epigenomic modifications will persist following a transient signal, and under circumstances in which these features do persist, often the epigenome, in itself, is not self-perpetuating. Rather, the regulators that maintain persistent epigenomic modifications, principally TFs and their cofactors, drive epigenetic processes. Henceforth I use the term "epigenetic" sparingly and only to describe gene regulatory mechanisms that persist following a transient signal.

While hormone receptors have been shown to regulate epigenomic features in cell culture, particularly breast (ERα) and prostate (AR) cancer lines, few studies have examined how hormone receptors modify chromatin in the brain. Moreover, the longevity and maintenance of epigenomic modifications in neurons, which are post-mitotic, remains an open area of research. In this section, I review studies on sex differences in the neuronal epigenome in the context of fundamental principles of neuronal gene regulation.

Whole genome bisulfite sequencing (WGBS) of neurons has revealed a striking degree of DNA methylation on cytosine residues that are not followed by a guanine but rather an adenine (mCA) (Lister et al., 2013). This modification is found in other tissues, but it increases in the brains of mice and humans during postnatal development, concomitant with synaptogenesis and experience-dependent neural activity (He & Ecker, 2015; Lister et al., 2013; Stroud et al., 2017). In contrast, mCG patterns are established prenatally and appear stable over time; although it is possible their dynamics across individual cell types have been masked in analyses of bulk tissue. WGBS of genetically-defined neuron types, along with Dnmt3a ChIP-seq, revealed that Dnmt3a deposits mCA over gene bodies during postnatal development to maintain cell type-specific gene repression (Stroud et al., 2017), suggesting mCA contributes to neuronal maturation.

Can neonatal E2 alter DNA methylation state and expression of associated genes? Prior reports of sex differences in DNA methylation in the brain have assessed either individual methylated cytosines at promoters (Kurian et al., 2010; Schwarz et al., 2010; Westberry et al., 2010) or reported percent

methylation across the entire genome (Ghahramani et al., 2014; Nugent et al., 2015). There remain no examples of differentially-methylated regions causally linked to sex differences in gene expression. To date, the best example of the epigenetic effects of neonatal E2 comes from *Esr1* itself. *Esr1* is expressed in the brains of both sexes beginning in mid-gestation (Tobet et al., 1999) and is persistently downregulated by neonatal E2 via an unknown mechanism (DonCarlos et al., 1995), resulting in female-biased expression in certain brain areas (DonCarlos & Handa, 1994; X. Xu et al., 2012).

There are many reports of sex differences in DNA methylation at the *Esr1* promoter (Edelmann & Auger, 2011; Kurian et al., 2010; Schwarz et al., 2010; Westberry et al., 2008, 2010; Westberry & Wilson, 2012; Wilson et al., 2008, 2011), although the residues identified as differentially hypo- or hyper-methylated vary across studies. In humans, the proximal promoter of human *ESR1* is hypomethylated across tissues, suggesting that variation in methylation at individual promoter CpGs is not causal for *Esr1* expression. In fact, across the genome, most CpGs at promoters are hypomethylated, regardless of the expression level of the associated gene (He & Ecker, 2015). However, CpG methylation at upstream, alternative *Esr1* promoters is tissue-specific, consistent with prior work demonstrating that *Esr1* is expressed from alternate promoters in distinct tissues and tumor types (Kos et al., 2001; Maekawa et al., 2016; Westberry et al., 2010; Wilson et al., 2008). WGBS across multiple developmental time points may reveal distal regulatory elements that maintain sex-biased expression of *Esr1*, and other estradiol-regulated genes, in the brain (Hon et al., 2013). While there is little evidence of an interaction between ERα and DNMT enzymes, TET2, a member of the ten-eleven translocation (TET) family of enzymes that catalyzes hydroxylation of 5mC, has recently been shown to be recruited to the genome by ERα *in vitro* (Broome et al., 2021).

In addition to DNA methylation, chemical modification of amino acid residues within the N-terminal tail of histone proteins has long been linked to epigenetic regulation of gene expression (Jenuwein & Allis, 2001). Such modifications physically alter chromatin structure via biophysical

interactions as well as attract histone "reader" proteins, such as BRD4, that remodel chromatin and recruit transcriptional machinery (Yun et al., 2011).

The neonatal hormone surge has been hypothesized to impart such histone PTMs, resulting in chromatin states that maintain sex-specific gene expression programs (Gagnidze et al., 2010; McCarthy & Nugent, 2015). Although such chromatin states have not been described, ERα has been found to bind histone acetyltransferases (HATs), methyltransferases, and deacetylases (HDACs), the SWI/SNF nucleosome remodeling complex, and Mediator protein in cancer cell lines (Perissi & Rosenfeld, 2005). ERα also binds steroid receptor coactivators (SRCs), such as p160 coactivator, as well as p300, which is a HAT that occupies active enhancers (Foulds et al., 2013; McKenna & O'Malley, 2002; Yi et al., 2015). Broad manipulation of coregulator or HDAC function disrupts brain development (Apostolakis et al., 2002; Matsuda et al., 2011; Molenda et al., 2002; Murray et al., 2009), but as with DNA methylation, there are little to no examples of persistent sex differences in histone modifications or transcription factor occupancy at specific loci in the brain (Ratnu et al., 2017).

1.6 Activation of neural circuits by steroid hormones following puberty

Puberty is considered a second critical period for sexual differentiation of the brain, in which developmentally-programmed circuits undergo refinement and maturation (Schulz & Sisk, 2016). Puberty refers to hormonal changes required for reproduction, while adolescence, in humans, is defined as the maturation of social and cognitive skills that can occur independently of gonadal hormones. Although many sex differences are established during the early postnatal critical period, behaviors regulated by the social behavior network (mating, aggression, social preference, parental care) are not displayed until after adult patterns of circulating hormones are established during puberty. Accordingly, sex-specific neuronal firing in the MeA in response to opposite-sex odors does not occur in pre-pubertal animals (Bergan et al., 2014).

During adolescence, the brain undergoes major synaptic pruning, which extends into the twenties in humans and P35-45 in rodents (Fig. 1.1). Connectivity within the mesolimbic reward pathway continues to develop during this time and is modulated by gonadal hormone signaling (reviewed in (Delevich et al., 2019; Walker et al., 2017)). This pathway intersects with social behavior circuitry to modulate sex differences in motivated behaviors such as parenting, sexual behavior, threat avoidance, and intake of drugs of abuse (Becker, 2009; Gee et al., 2018; O'Connell & Hofmann, 2011). During this time, the brain is uniquely vulnerable to stress, and many psychiatric diseases first show symptoms (Kessler et al., 2007; Paus et al., 2008). Early puberty can exacerbate the onset and symptoms of these conditions. Indeed, there are sex differences in the trajectory of puberty itself, with puberty onset occurring in girls before boys and at increasingly earlier ages in recent years (Piekarski, Johnson, et al., 2017; Sisk & Foster, 2004).

It is thought that pubertal hormones end the plasticity of the juvenile brain, locking-in learned skills and impeding the ability to acquire new ones. Songbird research has demonstrated a direct relationship between testosterone level and plasticity associated with vocal learning. Many male songbirds acquire their unique song during a juvenile critical period, in which both sensory experience and sensorimotor learning must occur (Brainard & Doupe, 2002). Similar to the role of neonatal E2 in rodent territorial behaviors, testosterone in early life is required for the appropriate onset of singing, whereas high levels of pubertal testosterone are required to crystallize song structure into a permanent stereotyped adult song. Exogenous administration of testosterone during the quiescent period between these two developmental windows impairs song learning, resulting in abnormal songs with altered syntax and decreased complexity (Korsia & Bottjer, 1991; Whaling et al., 1995). Electrophysiological data also demonstrate that early testosterone accelerates the maturation of neuronal firing patterns in song-related brain areas (Sizemore & Perkel, 2011; White et al., 1999). Taken together, this work suggests pubertal hormones reduce plasticity, and early puberty may lead to the premature closure of developmental critical periods.

In support of this idea, Piekarski and colleagues found that inhibition onto excitatory neurons in the cingulate cortex increases in female mice following puberty (Piekarski, Boivin, et al., 2017). This increase in inhibition was associated with a decrease in behavioral flexibility, as pre-pubertal animals demonstrated better performance on a reversal learning task compared to post-pubertal animals. Moreover, early puberty induction by treatment with ovarian hormones resulted in adult-like task performance. Parallel conclusions drawn from male songbirds and female mice point strongly to a conserved role for pubertal hormones in establishing adult-typical patterns of neural activity.

What are the gene expression programs that drive neuronal maturation during puberty? Do the genomic targets of gonadal hormones differ between birth and puberty? These questions remain open. Although there has been extensive investigation into both behavioral transitions and anatomical correlates of cortical maturation during puberty (reviewed in (Gee et al., 2018)), there are relatively few studies on the dynamic changes in gene expression that underlie these events. Walker et al. assessed gene expression in the POA and medial-basal hypothalamus (MBH) from birth to adulthood and identified *Tac2* as a novel female-biased gene in the MBH, beginning at P15 (Walker et al., 2012). Shi et al. extensively analyzed the BrainSpan database, which includes up to 15 brain regions staged from 8pcw to 40 years of age (Shi et al., 2016); the authors found that although sex differences in expression first arose during the second trimester, when the fetal testosterone surge occurs, most sex differences were not detected until puberty (8-19 years). However, the results from this study were highly variable due to low sample size, and only 3 sex-biased autosomal genes showed a consistent bias across developmental stages. Transcriptomic analyses of the maturing mesolimbic circuitry in rodents may reveal candidate factors that alter neuronal firing or synaptic maturation, such as cell adhesion molecules, ion channels, and vesicle release machinery (Favuzzi et al., 2019; Paul et al., 2017a).

1.7 Conclusion

Given the many sex differences in the brain that emerge during early-life and adulthood, there is a clear need to characterize sex differences in gene regulation that occur during these life stages. In this thesis, I describe experiments that I performed to address this question. In Chapter 2, I introduce the low-input TF profiling method, Cleavage Under Targets & Release Under Nuclease (CUT&RUN), and describe how I used this technology to profile ERα genomic binding in the mouse social behavior network. In Chapter 3, I demonstrate how activation of a single ERα genomic target, melanocortin receptor 4 (*Mc4r*), in the ventrolateral VMH (VMHvl) drives female physical activity. In Chapter 4, I characterize the activational effects of gonadal hormones by examining sex differences in gene expression and chromatin state in the adult BNST. In Chapter 5, I describe how the neonatal hormone surge influences ERα activation, resulting in a male-typical gene regulatory program that persists throughout neonatal BNST development. In Chapter 6, I demonstrate that single-nucleus multiome sequencing, which captures both gene expression and chromatin accessibility from the same individual cell, improves the prediction of TFs regulating neuron identity. In Chapter 7, I review the primary conclusions from this work and explore future directions. Lastly, in the Appendix, I include a manuscript on a novel, single-cell sequencing method, called Balls of Acrylamide Gel sequencing (BAG-seq), which has the potential to reduce the cost and increase the throughput of single-cell transcriptomic and epigenomic experiments relative to current commercial options.

Chapter 2

Genomic targets of ERα in the social behavior network

Sections of this chapter were previously uploaded to bioRxiv (2020) under the title "Regulation of neural gene expression by estrogen receptor alpha" by Bruno Gegenhuber, Melody V. Wu, Robert Bronstein, and Jessica Tollkuhn. This manuscript is now in press at Nature (2022) under the title "Gene regulation by gonadal hormone receptors defines neuronal sex differences" by Bruno Gegenhuber, Melody V. Wu, Robert Bronstein, and Jessica Tollkuhn. I thank Melody V. Wu for collecting the RNA-seq and in situ hybridization data shown in this chapter.

2.1 Abstract

Many actions of E2 are exerted through its canonical nuclear receptor, which is rapidly recruited to DNA upon activation. To date, the genomic targets of ERα in the brain remain unknown, due largely to its low and sparse expression. Here I apply the low-input TF profiling method CUT&RUN to characterize ERα genomic binding within three brain regions of the social behavior network in which ERα expression is required for male-typical behaviors: the BNSTp, MeA, and MPOA. In addition, I use genetic targeting approaches to study how E2 influences chromatin accessibility and ribosome-bound transcript levels in BNSTp *Esr1*+ neurons. Together, these experiments reveal ERα binds unique genomic targets in the brain important for synaptic communication, and can drive rapid changes in chromatin state and gene expression across both sexes, independently of prior exposure to gonadal hormone signaling.

2.2 Introduction

Over the last 40 years, mechanisms of transcriptional control by ERα have been extensively characterized in the context of breast cancer, leading to diagnostic and therapeutic breakthroughs (Carroll, 2016; Evans, 1988; Ross-Innes et al., 2012). However, the genes regulated by ERα in neurons remain obscure, due largely to technical challenges posed by the heterogeneity of the mammalian brain. For many years, ChIP-seq has been considered the gold standard method for profiling TF genomic binding. However, one limitation of ChIP-seq is that it typically requires a large number of cells (>1 million) to achieve high-quality data, because of signal loss arising from epitope masking and destruction during fixation and sonication, respectively. Given that ERα, and other gonadal hormone receptors (AR, PR), are expressed only in a few thousand cells across the entire mouse brain, ChIP-seq has not been considered a suitable approach for identifying neural genomic targets of gonadal hormone receptors.

To bypass the input requirement of ChIP-seq, the Henikoff lab developed a modified approach to TF profiling, called CUT&RUN (Skene et al., 2018; Skene & Henikoff, 2017). CUT&RUN achieves near-quantitative recovery of TF-bound DNA fragments at single-nucleotide resolution by deviating from ChIP-seq at two critical stages:

1) rather than chromatin fixation and solubilization, which masks and/or damages the epitope, CUT&RUN is performed on unfixed, permeabilized cells or nuclei attached covalently to magnetic beads and

2) rather than TF immunoprecipitation, which often pulls down non-specific DNA fragments, CUT&RUN uses a protein A-micrococcal nuclease (pA-MNase) fusion enzyme to bind and cleave DNA selectively around primary antibodies targeted to the TF of interest. Cleaved DNA fragments diffuse out of the nuclear and/or plasma membrane, and these fragments are isolated from bead-bound nuclei/cells by centrifugation or magnetic separation.

Here I use CUT&RUN to characterize ERα genomic binding in the brain, after first optimizing the method in MCF-7 breast cancer cells. I then examine the effects of E2 on chromatin accessibility and gene expression in BNSTp *Esr1*+ cells, by performing Assay for Transposase-Accessible Chromatin using Sequencing (ATAC-seq) on *Esr1*^{Cre}; *Sun1-GFP*^{lx} mice and Translating Ribosome Affinity Purification (TRAP), followed by RNA-seq, on *Esr1*$^{Cre/+}$; *Rpl22*$^{HA/+}$ mice, respectively. Lastly, we validate ERα regulation of specific genes within the BNSTp using *in situ* hybridization. Collectively, these experiments reveal the first genomic targets of a gonadal hormone receptor in the brain and demonstrate mechanisms of E2-dependent gene regulation in the BNSTp.

2.3 Results

2.3.1 ERα CUT&RUN validation in MCF-7 breast cancer cells

As CUT&RUN has not yet been applied to ERα, I first validated the approach in MCF-7 breast cancer cells, which are a model system for studying ERα genomic regulation (Carroll et al., 2005). I performed ERα CUT&RUN on 25,000 MCF-7 cells, using 2 different antibodies, and compared to a published ERα ChIP-seq dataset (Franco et al., 2015) generated from ~2.5 million MCF-7 cells that underwent a similar E2 treatment paradigm. Despite using ~1% of the input material, CUT&RUN detected robust and reproducible recruitment of ERα to peaks identified in the ChIP-seq dataset, as well as a clear MNase footprint surrounding EREs within these peaks (Fig. 2.1a-b). Differential peak-calling of ERα CUT&RUN data revealed significant ERα binding to 6363 (antibody #1) or 7530 (antibody #2) genomic regions (Fig. 2.1c-d). Comparison of individual genomic loci between CUT&RUN and ChIP-seq revealed near-identical ERα binding at canonical MCF-7 targets, *TFF1* and *GREB1* (Fig. 2.1e). Moreover, ERα CUT&RUN and ChIP-seq peaks shared near-identical enriched and *de novo* motifs, including not only the ERE, but also motifs of known ERα binding partners, such as forkhead box (FOX) and homeodomain (HD) proteins (Fig. 2.1f).

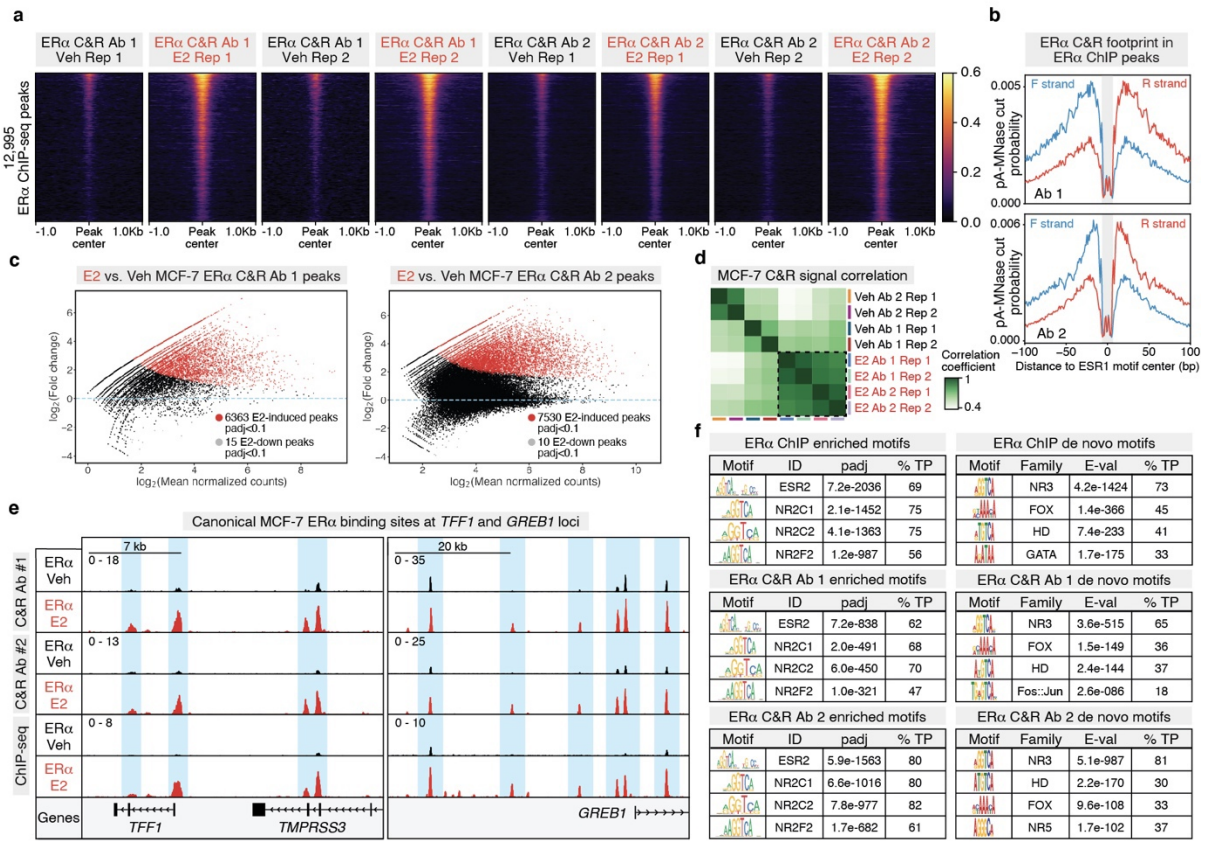

Figure 2.1. Validation of ERα CUT&RUN in MCF-7 cells. (**a**) Heatmap of mean MCF-7 ERα CUT&RUN CPM ±1Kb around 12,995 17β-estradiol (E2)-induced MCF-7 ERα ChIP-seq peaks (DiffBind, DESeq2, padj<0.01) for individual replicates (n=2 per condition and antibody). (**b**) pA-MNase-cut footprint (CUT&RUNTools) around ESR1 motif sites (FIMO) detected in ERα ChIP-seq peaks. (**c**) MA plots of differential ERα CUT&RUN peaks (DiffBind, DESeq2, padj<0.1) for ERα antibody #1 (Santa Cruz sc-8002) and ERα antibody #2 (EMD Millipore Sigma 06-935). (**d**) Pearson correlation coefficient of CPM-normalized CUT&RUN signal within the consensus peak matrix across ERα CUT&RUN samples. Red text indicates E2 treatment group. (**e**) ERα CUT&RUN (both antibodies) and ChIP-seq tracks at canonical MCF-7 ERα target genes (*TFF1*, *GREB1*). (**f**) (left) Top enriched motifs (AME) and (right) top *de novo* motifs (DREME) within ERα ChIP-seq peaks, Ab #1 ERα CUT&RUN peaks, and Ab #2 ERα CUT&RUN peaks. % TP=% of peaks called as positive for the indicated motif. *De novo* motifs were classified into motif families using TomTom.

2.3.2 ERα CUT&RUN in the brain

To determine the genomic targets of ERα in the brain, I used an established hormone starvation and replacement paradigm that reproducibly elicits sex-typical behaviors (McEwen, 1981) and replicates the media conditions required to detect ERα genomic binding in cell lines (Hurtado et al., 2011). Four hours after treatment with estradiol benzoate (E2) or vehicle control, I profiled ERα binding within three interconnected limbic brain regions in which ERα is genetically required for sex-typical behaviors: the BNSTp, MPOA, and MeA (Y. Li & Dulac, 2018; Simerly, 2002b; M. V. Wu & Tollkuhn, 2017b) (Fig. 2.2a). With this approach, I detected 1930 E2-induced ERα-bound loci in the brain (Fig. 2.2b, Fig 2.3a-e). The most enriched TF binding motif in these peaks was the ERE, validating the specificity of the method (Fig. 2.3d). Prior ChIP-seq studies *in vivo* have examined ERα genomic binding in mouse uterus (Gertz et al., 2013; Hewitt et al., 2012), liver (Gertz et al., 2013; Gordon et al., 2014), aorta (Gordon et al., 2014), efferent ductules (G. Yao et al., 2017), and mammary gland (Palaniappan et al., 2019). Comparison of brain ERα binding sites to those previously detected in these tissues revealed that the majority are specific to the brain (Fig. 2.2c, Fig. 2.3f). Brain-specific ERα binding events were uniquely enriched for synaptic and neurodevelopmental disease gene ontology (GO) terms, including neurotransmitter receptors, ion channels, neurotrophin receptors, and extracellular matrix genes (Fig. 2.2d, Fig. 2.3h-j). I also found evidence supporting direct crosstalk between estradiol and neuroprotection, as ERα directly binds loci for neurotrophin receptors TrkB (*Ntrk2*) and *Ntrk3* (Fig. 2.3k).

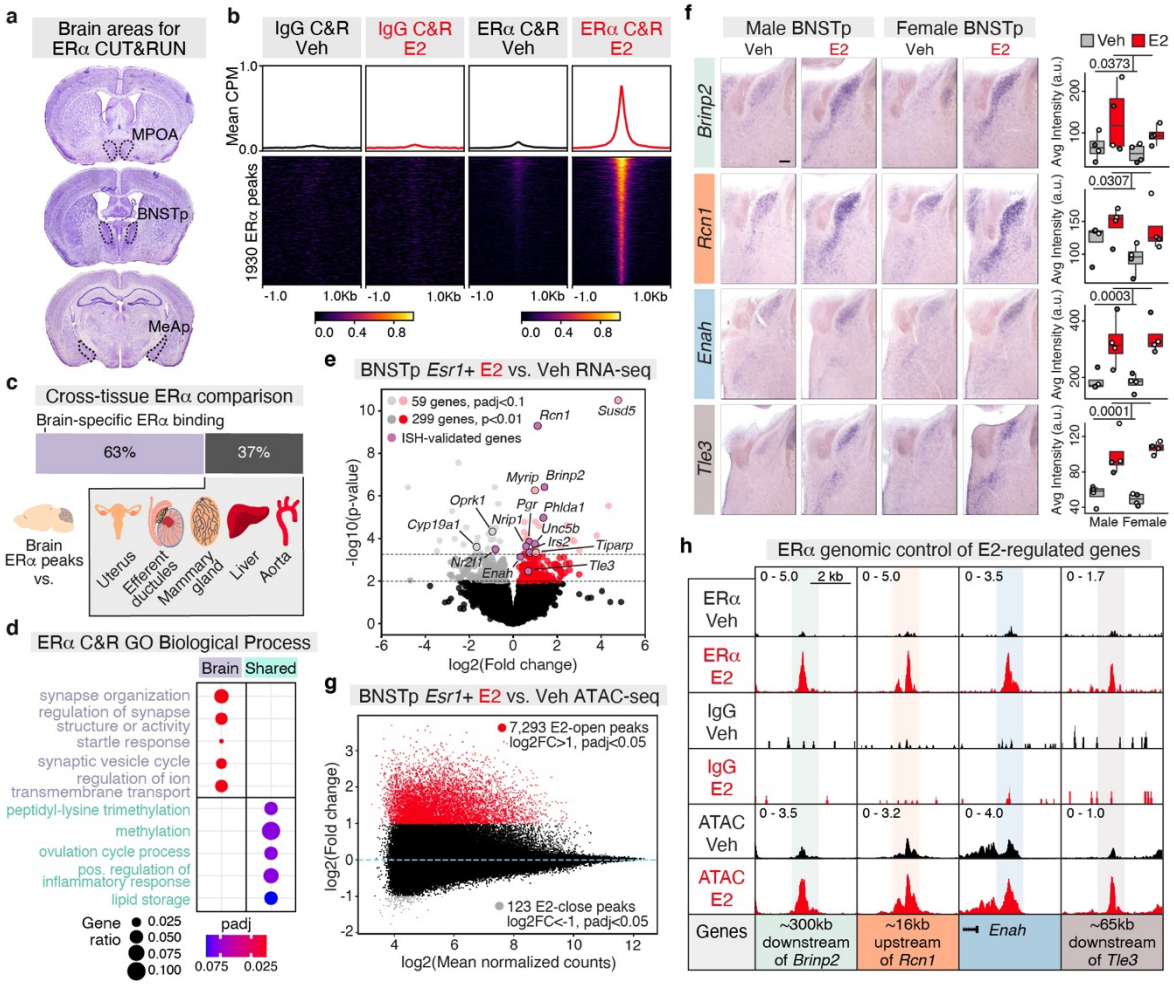

Figure 2.2. Genomic targets of ERα in sexually dimorphic neuronal populations.

(**a**) Coronal sections containing sexually dimorphic brain areas used for ERα CUT&RUN. (**b**) Heatmap of mean IgG and ERα CUT&RUN CPM ±1Kb around E2-induced ERα CUT&RUN peaks (DiffBind edgeR, padj<0.1). Heatmap sorted by E2 ERα CUT&RUN signal. (**c**) Proportion of ERα peaks detected specifically in brain. (**d**) Top GO Biological Process terms associated with genes nearest to brain-specific or shared (≥4 other tissues) ERα peaks (clusterProfiler, padj<0.1). (**e**) Combined sex E2 vs. Veh RNA-seq in BNSTp *Esr1*+ cells; light grey and red dots (DESeq2, padj<0.1), dark grey and red dots (DESeq2, p<0.01), purple dots (ISH-validated). (**f**) ISH of select genes induced by E2 in both sexes. Boxplot center=median, box boundaries=1st and 3rd quartile, whiskers=1.5*IQR from boundaries. 2-way ANOVA: *Brinp2* p=0.0373, *Rcn1* p=0.0307, *Enah* p=0.0003, *Tle3* p=0.0001; n=4 per condition, scale=200um. (**g**) MA plot of E2-regulated ATAC-seq peaks in BNSTp *Esr1*+ cells; red dots (DiffBind edgeR, log2FC>1, padj<0.05), grey dots (DiffBind edgeR, log2FC<-1, padj<0.05). (**h**) Example ERα peaks at E2-induced genes.

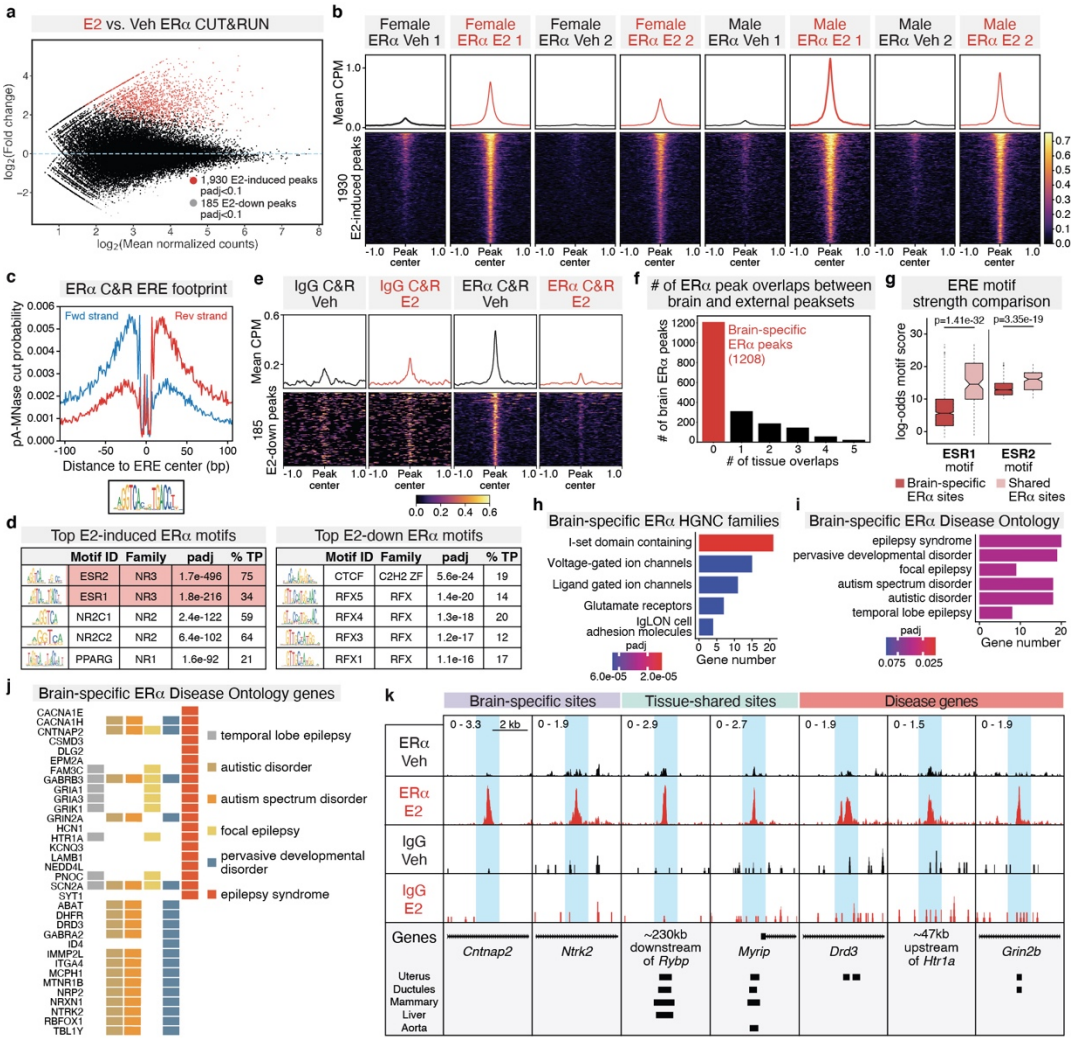

Figure 2.3. Additional analysis of adult brain ERα CUT&RUN dataset. (**a**) MA plot of differential ERα CUT&RUN peaks (DiffBind, edgeR, padj<0.1) in adult mouse brain. red dots=E2-induced peaks, grey dots=E2-down peaks. (**b**) Heatmap of mean brain ERα CUT&RUN CPM ±1Kb around 1930 E2-induced ERα CUT&RUN peaks (see also Fig. 1b) for individual replicates (n=2 per condition). (**c**) ESR1 motif footprint in ERα peaks (CUT&RUNTools). (**d**) Top enriched motifs (AME) in (left) E2-induced ERα peaks and (right) E2-down ERα peaks. (**e**) Heatmap of mean brain IgG and ERα CUT&RUN CPM ±1Kb around 185 E2-down ERα peaks. (**f**) Number of overlaps between E2-induced ERα peaks and 7 external ERα ChIP-seq peaksets: intersected peaks of uterus 1 and uterus 2, intersected peaks of liver 1 and liver 2, aorta, efferent ductules, and mammary gland. Red indicates brain-specific ERα peaks. (**g**) Log-odds motif scores (FIMO) for the ESR1 motif (MA0112.3, left) and ESR2 motif (MA0258.2, right) in brain-specific (red) and shared (pink) ERα peaks. Boxplot center=median, box boundaries=1st and 3rd quartile, whiskers=1.5*IQR from boundaries. n=1304 brain-specific ESR1, 139 shared ESR1, 1276 brain-specific ESR2, 157 shared ESR2. p-values from two-sided, Wilcoxon rank-sum test. (**h**) Top Hugo Gene Nomenclature Committee (HGNC) gene families (clusterProfiler, padj<0.1) enriched within brain-specific ERα peak-associated genes. (**i**) Top Disease Ontology terms associated with genes nearest to brain-specific ERα peaks (DOSE, padj<0.1). (**j**) Brain-specific ERα peak-associated genes within each enriched Disease Ontology (DO) term (clusterProfiler, padj<0.1), colored by term (**k**) Example brain-specific (*Cntnap2*, *Ntrk2*), shared (*Rybp*, *Myrip*), and disease-associated (*Drd3*, *Htr1a*, *Grin2b*) ERα peaks.

2.3.3 Estradiol regulation of chromatin accessibility and gene expression in the BNSTp

To identify the functional consequences of ERα genomic binding, I focused on a single brain region, the BNSTp, which is a central node in the regulation of sex-typical behaviors. The BNSTp sends projections to the POA, MeA, and hypothalamus and is 40% larger in rodent and human males compared to females (Allen & Gorski, 1990a; Hines et al., 1992a; Simerly, 2002a; Welch et al., 2019; J. N. Zhou et al., 1995). In mice, this dimorphism requires neonatal ERα activation, which promotes survival of BNSTp neurons (Tsukahara et al., 2011; M. V. Wu et al., 2009b). Our group performed translating ribosome affinity purification (TRAP) in microdissected BNSTp from *Esr1*$^{Cre/+}$; *Rpl22*$^{HA/+}$ females and males using the same treatment paradigm as above, followed by RNA-seq. I identified 358 genes with differential expression between E2 and Veh treatment groups, including known estradiol-induced genes, such as *Pgr* and *Nrip1* (Fig. 2.2e).

We validated several of the top estradiol-induced genes by *in situ* hybridization (ISH) (Fig. 2.2f). Genes involved in neuron wiring (*Brinp2*, *Unc5b*, *Enah*) and synaptic plasticity (*Rcn1*, *Irs2*) are robustly induced in the BNSTp, illustrating how estradiol signaling may sculpt sexual differentiation of BNSTp circuitry. ERα co-repressors *Tle3* and *Nrip1* increase expression, while the co-activator *Nr2f1* decreases expression, suggesting a negative transcriptional feedback mechanism.

In addition to direct genomic binding by ERα, E2 can regulate gene expression via cell membrane-initiated signaling cascades, tethering of ERα to AP-1 sites via Fos/Jun heterodimers, and activation of estrogen receptor beta (ERβ), which is also expressed in the BNSTp (Micevych & Kelly, 2012; Srivastava et al., 2013; Zuloaga et al., 2014b). To detect additional E2-responsive genomic regions in the BNSTp, I performed ATAC-seq on FACS-purified BNSTp *Esr1*+ cells collected from *Esr1*$^{Cre/+}$; *Sun1-GFP*$^{lx/+}$ mice using the same treatment paradigm. This approach revealed robust chromatin opening in response to E2. Across sexes, 7,293 chromatin regions increased accessibility (E2-open) and 123 regions decreased accessibility (E2-close) (Fig. 2.2g, Fig. 2.4a-b, d). While the number of E2-open loci far exceeded the number of E2-regulated genes, E2-open loci were similarly

enriched at genes involved in neuronal connectivity, including regulation of membrane potential and synapse organization (Fig. 2.4e-g). Motif enrichment analysis of these E2-open regions, which occurred primarily at distal enhancer elements (Fig. 2.4c), showed that 89% contain an ERE (Fig. 2.4h-i), consistent with the observation that nearly all ERα binding sites overlapped an E2-open region (Fig. 2.4j). Collectively, these results indicate direct estrogen receptor binding, rather than indirect or membrane-initiated signaling pathways, drives the majority of E2-responsive chromatin regions in the BNSTp.

After examining the relationship between estradiol-regulated chromatin loci and gene expression, I found that E2-open regions localized at both E2-upregulated and -downregulated genes (Fig. 2.5a). E2-open regions at downregulated genes contained EREs yet lacked widespread ERα binding (Fig. 2.5b-d), suggesting transient ERα recruitment may contribute to gene repression (Guertin et al., 2014). E2-upregulated genes with corresponding E2-responsive chromatin loci include *Brinp2*, *Rcn1*, *Enah*, and *Tle3* (Fig. 2.2f, h); E2-downregulated genes include *Astn2*, a regulator of synaptic trafficking, and *Nr2f1* (Fig. 2.5d).

While most estradiol regulation events were shared between sexes in our treatment paradigm, I noted certain sex-dependent effects. Pairwise comparison by sex revealed nearly 300 differential genes between females and males in our TRAP RNA-seq data (data not shown). Moreover, I observed 306 genes with a differential response to estradiol between sexes (Fig. 2.5e-f). These sex-dependent, E2-responsive genes lacked enrichment of E2-responsive chromatin regions (Fig. 2.5g), which may indicate additional estradiol regulation at the translational level, consistent with the recent discovery of ERα as an RNA-binding protein (Y. Xu et al., 2021). Likewise, across ERα CUT&RUN and ATAC-seq modalities, I observed negligible sex differences and sex-dependent, E2-responsive loci (Fig. 2.5h-j), demonstrating females and males mount a similar genomic response to exogenous estradiol upon removal of the hormonal milieu.

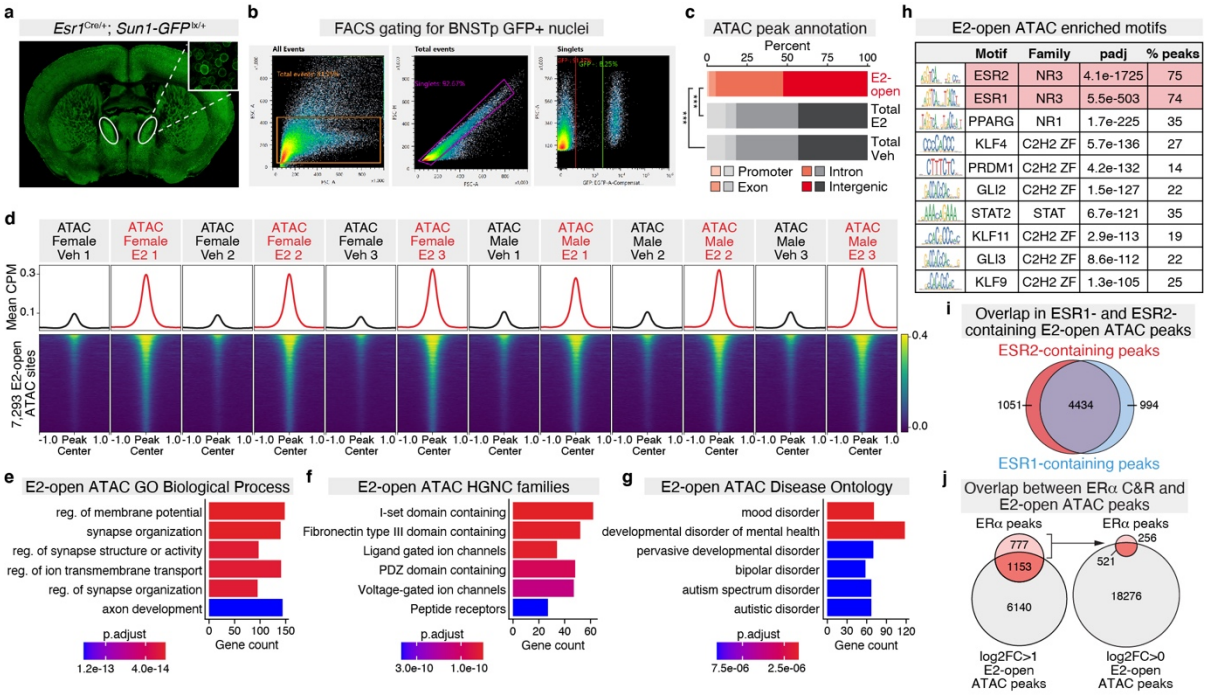

Figure 2.4. Additional analysis of adult BNSTp ATAC-seq. (**a**) GFP immunofluorescence staining in an adult male *Esr1*$^{Cre/+}$; *Sun-GFP*$^{lx/+}$ mouse, scale=1mm. Inset shows Sun1-GFP signal at nuclear membrane, scale=10um. (**b**) Fluorescence-activated cell sorting (FACS) gating strategy for isolating BNSTp GFP+ nuclei for ATAC-seq. (**c**) Proportion of E2-open ATAC peaks (red), total E2 ATAC peaks (black), and total Veh ATAC peaks (black) annotated to promoters (±1Kb around TSS), exons, introns, and intergenic regions. E2-open ATAC peaks have a significantly lower proportion of peaks annotated to gene promoters than total vehicle (11% vs 1% , Fisher's Exact Test, $p=4.6x10^{-260}$) and total E2 (11% vs 1%, Fisher's Exact Test, $p=4.3x10^{-267}$) peaks. ***p<0.001. (**d**) Heatmap of mean BNSTp *Esr1*+ ATAC CPM ±1Kb around 7293 E2-open ATAC peaks for individual female and male replicates (n=3 per condition) (see also Fig. 1g). (**e-g**), Top Gene Ontology (GO) Biological Process terms (**e**), HGNC gene families (**f**), and DO terms (**g**) enriched within E2-open ATAC peak-associated genes (clusterProfiler, padj<0.1). (**h**) Top motifs enriched in E2-open ATAC-seq peaks (AME). % of peaks containing motifs determined with FIMO. (**i**) Overlap in E2-open ATAC peaks containing the ESR1 motif (blue) and the ESR2 motif (red), identified using FIMO. The majority of peaks (4434/6479) containing either motif are the same. (**j**) (left) Overlap between brain ERα CUT&RUN peaks and E2-open ATAC peaks (log2FC>1). (right) Overlap between remaining 777 brain ERα CUT&RUN peaks and log2FC>0 E2-open ATAC peaks.

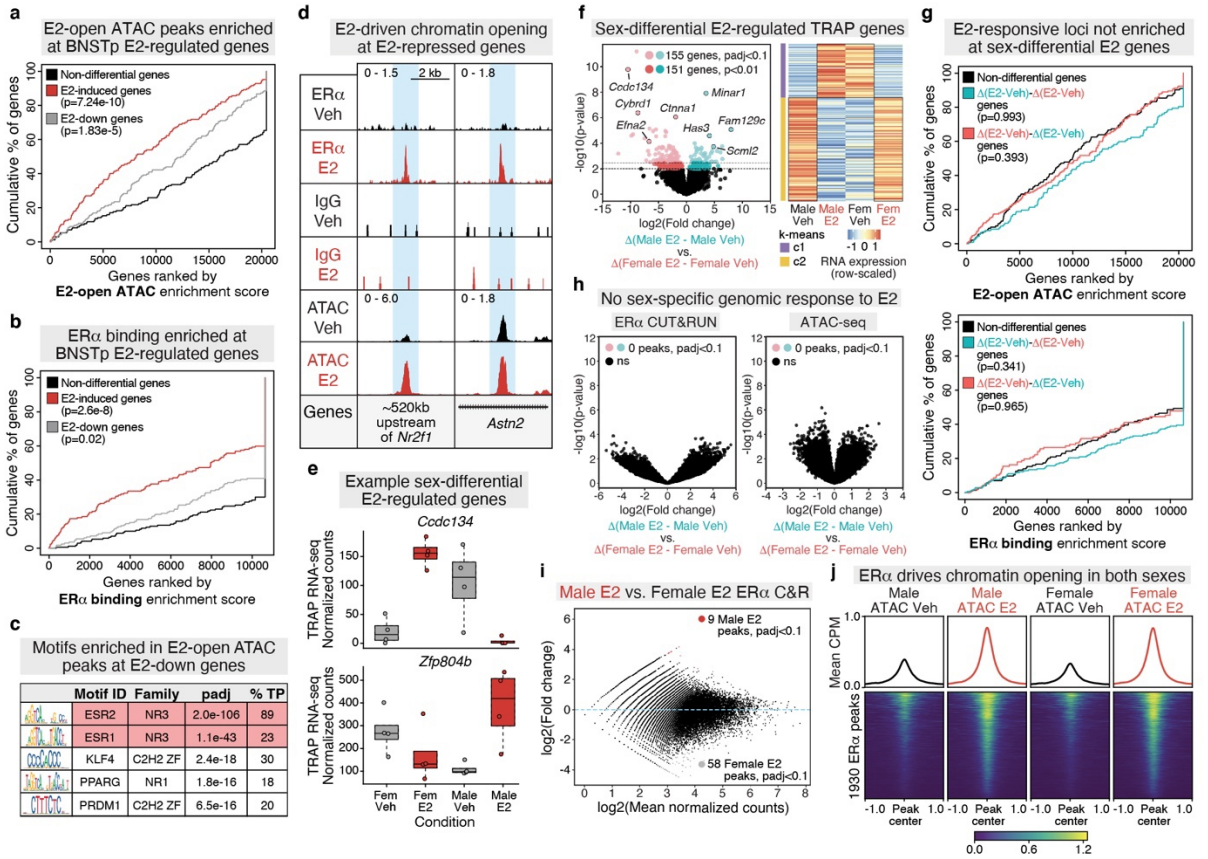

Figure 2.5. Integration of adult RNA-seq, ATAC-seq, and CUT&RUN datasets. (**a-b**) BETA enrichment of E2-open ATAC peaks (**a**) and brain ERα CUT&RUN peaks (**b**) at E2-induced and E2-down genes identified by RNA-seq (DESeq2, p<0.01) relative to a background of non-differential, expressed genes. (**c**) Top enriched motifs (AME) in E2-open ATAC peaks ±350Kb around E2-down genes (identified with BETA). (**d**) Example E2-open ATAC peaks/ERα peaks at E2-repressed genes, *Nr2f1* and *Astn2*. (**e**) Normalized counts for example genes (*Ccdc134*, *Zfp804b*) with a sex-dependent response to E2 treatment. Boxplot center=median, box boundaries=1st and 3rd quartile, whiskers=1.5*IQR from boundaries. n=4. (**f**) (Left) Volcano plot of sex-dependent, E2-responsive genes; light blue and red dots (DESeq2, padj<0.1), dark blue and red dots (DESeq2, padj<0.1). (Right) Mean, normalized expression of sex-dependent, E2-responsive genes (DESeq2, padj<0.1), grouped by k-means clustering. (**g**) Lack of significant enrichment of E2-open ATAC peaks (top) and ERα peaks (bottom) at sex-dependent, E2-reponsive genes relative to a background of non-differential, expressed genes (BETA). (**h**) Volcano plots of sex-dependent, E2-responsive ERα CUT&RUN peaks (edgeR, padj<0.1) (left) and ATAC peaks (edgeR, padj<0.1). (**i**) MA plot of male E2 vs. female E2 ERα CUT&RUN peaks (DiffBind, edgeR, padj<0.1); red dots=male E2-biased peaks, grey dots=female E2-biased peaks. (**j**) Heatmap of mean ATAC CPM, split by sex and treatment, ±1Kb around E2-induced ERα peaks.

2.4 Discussion

Here I characterize the genomic and transcriptomic response to estradiol in the adult mouse brain. Within BNSTp *Esr1*+ neurons, estradiol largely activates chromatin loci, primarily at distal or intronic enhancer elements, via direct ERα binding. This finding suggests that membrane-initiated ERα signaling, culminating in CREB phosphorylation, or modulation of BNSTp *Esr1*+ neuron activity via long-range *Esr1*+ projection neurons from other brain regions do not contribute to the genomic response. Surprisingly, I found minimal sex differences in the genomic response to estradiol, whereas the transcriptome, specifically levels of ribosome-bound mRNAs, deviated between sexes following treatment. These data reveal that prior history of hormone signaling (i.e. neonatal surge in males, post-pubertal hormones in both sexes) do not restrict the chromatin to respond to later hormones, meaning perturbation of adult hormone levels effectively alters genomic sex – a concept I return to in Chapter 5. However, as the TRAP data still showed sex differences in the response to estradiol, I suspect additional regulation occurs at the translational level. In support of this hypothesis, ERα has recently been found to act as an RNA-binding protein (Y. Xu et al., 2021). Direct comparison of BNSTp *Esr1*+ nuclear vs. ribosome-bound transcripts following estradiol treatment would be critical in distinguishing between these layers of regulation.

Consistent with one prior *in vivo* study (Gertz et al., 2013), ERα displays tissue-specific genomic binding. It is possible that such tissue-specific binding enables ERα to mediate estradiol-dependent biological processes critical for the function of each tissue, such as bone growth in osteoclasts, lipid metabolism in the liver, and cardioprotection in the heart. As the majority of ERα binding events were accessible in the BNSTp prior to treatment, tissue-specific recruitment may be driven by the pre-existing chromatin landscape specified by tissue-specific expression, and/or combinatorial expression, of PFs. Additional ERα CUT&RUN and ATAC-seq experiments across tissues, followed by comparison and motif analysis, would provide mechanistic insight into how ERα regulates diverse biological processes throughout the body.

Importantly, the experiments performed here establish a methodological framework for studying gonadal hormone receptor function in the brain. I return to this methodology in Chapters 5 and 6 to define gene regulatory mechanisms controlling sex differences in the adult ("activational") and developing ("organizational") BNSTp.

2.5 Materials and Methods

Animals. All animals were maintained on a 12:12 light cycle and provided food and water *ad libitum*. All mouse experiments were performed under strict guidelines set forth by the CSHL Institutional Animal Care and Use Committee (IACUC). *Esr1*^{Cre} (H. Lee et al., 2014), *Rpl22*^{HA} (Sanz et al., 2009), *ROSA26*$^{CAG\text{-}Sun1\text{-}sfGFP\text{-}Myc}$ (Mo et al., 2015) (abbreviated as *Sun1-GFP*), and C57Bl6/J wildtype mice were obtained from Jackson labs. Adult male and female mice were used between 8-12 weeks of age. For adult hormone treatment experiments, animals were sacrificed for tissue collection four hours after subcutaneous administration of 5μg estradiol benzoate (E2) (Sigma E8515) suspended in corn oil (Sigma C8267) or vehicle three weeks post-gonadectomy.

Cell lines. Cells were maintained in standard DMEM supplemented with 10% FBS and penicillin/streptomycin. Prior to CUT&RUN, MCF-7 cells were grown in phenol-red free DMEM media containing 10% charcoal-stripped FBS and penicillin/streptomycin for 48 hours then treated with 20 nM 17-β-estradiol or vehicle (0.002% EtOH) for 45 minutes.

Adult RNA-seq and in situ hybridization. Experiments were performed as previously described (Ahrens et al., 2018). Briefly, the BNSTp was microdissected following rapid decapitation of deeply anesthetized adult *Esr1*$^{Cre/+}$; *Rpl22*$^{HA/+}$ mice. Tissue homogenization, immunoprecipitation, and RNA extraction was performed, and libraries were prepared from four biological replicates samples (each consisting of 8-9 pooled animals) using NuGEN Ovation RNA-Seq kits (7102 and 0344). Multiplexed libraries were sequenced with 76bp single end reads on the Illumina NextSeq. Validation by in-situ

hybridization staining and quantification was performed as previously described (Ahrens et al., 2018; M. V. Wu & Tollkuhn, 2017b).

Adult ATAC-seq nuclei isolation. Adult *Esr1*$^{Cre/+}$; *Sun1-GFP*$^{lx/+}$ mice (4 pooled per condition) were deeply anesthetized with ketamine/dexmedetomidine. 500-µm sections spanning the BNSTp were collected in an adult mouse brain matrix (Kent Scientific) on ice. The BNSTp was microdissected and collected in 1 ml of cold supplemented homogenization buffer (250 mM sucrose, 25 mM KCl, 5 mM MgCl2, 120 mM tricine-KOH, pH 7.8), containing 1 mM DTT, 0.15 mM spermine, 0.5 mM spermidine, and 1X EDTA-free PIC (Sigma Aldrich 11873580001). The tissue was dounce-homogenized 15x in a 1 ml glass tissue grinder (Wheaton) with a loose pestle. 0.3% IGEPAL CA-630 was added, and the suspension was homogenized 5x with a tight pestle. The homogenate was filtered through a 40-µm strainer then centrifuged at 500 x g for 15 min at 4ºC. The pellet was resuspended in 0.5 ml homogenization buffer containing 1 mM DTT, 0.15 mM spermine, 0.5 mM spermidine, and 1X EDTA-free PIC. 30,000 GFP+ nuclei were collected into cold ATAC-RSB (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM MgCl2) using the Sony SH800S Cell Sorter (purity mode) with a 100-µm sorting chip. After sorting, 0.1% Tween-20 was added, and the nuclei were centrifuged at 500 x g for 5 min at 4ºC. The nuclei pellet was directly resuspended in transposition reaction mix.

ATAC-seq library preparation. Tn5 transposition was performed using the OMNI-ATAC protocol (Corces et al., 2017). 2.5 µl of Tn5 enzyme (Illumina 20034197) were used in the transposition reaction. Libraries were prepared with NEBNext High-Fidelity 2X PCR Master Mix (NEB M0541L), following standard protocol. After the initial 5 cycles of amplification, an additional 4 cycles were added, based on qPCR optimization. Following amplification, libraries were size-selected (0.5x-1.8x) twice with AMPure XP beads (Beckman Coulter A63880) to remove residual primers and large genomic DNA.

Individually barcoded libraries were multiplexed and sequenced with paired-end 76bp reads on an Illumina NextSeq, using either the Mid or High Output Kit.

Cell line CUT&RUN. To harvest cells for CUT&RUN, cells were washed twice with Hank's Buffered Salt Solution (HBSS) and incubated for five minutes with pre-warmed 0.5% Trypsin-EDTA (10X) at 37ºC/5% CO_2. Trypsin was inactivated with phenol-red free DMEM supplemented with 10% charcoal-stripped FBS and penicillin/streptomycin. After trypsinizing, cells were centrifuged at 500 x g in a 15ml conical tube and resuspended in fresh media. CUT&RUN was performed as previously described (Skene et al., 2018), with minor modifications. Cells were washed twice in Wash Buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 1X PIC, 0.02% digitonin). Cell concentration was measured on a Countess II FL Automated Cell Counter (Thermo Fisher). 25,000 cells were used per sample. Cells were bound to 20 µl concanavalin A beads (Bangs Laboratories, BP531), washed 2x in Wash Buffer, and incubated overnight with primary antibody (ERα antibody #1: Santa Cruz sc-8002 or antibody #2: EMD Millipore 06-935) diluted 1:100 in Antibody Buffer (Wash Buffer containing 2 mM EDTA). The following day, cells were washed 2x in Wash Buffer, and 700 ng/ml protein A-MNase (pA-MNase, prepared in-house) was added. After 1 hr incubation at 4ºC, cells were washed 2x in Wash Buffer and placed in a metal heat block on ice. pA-MNase digestion was initiated with 2 mM CaCl2. After 90 min, digestion was stopped by mixing 1:1 with 2X Stop Buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 µg/ml RNase A, 50 µg/ml glycogen, 0.02% digitonin). Digested fragments were released by incubating at 37ºC for 10 min, followed by centrifuging at 16,000 x g for 5 min at 4ºC. DNA was purified from the supernatant by phenol-chloroform extraction, as previously described (Skene et al., 2018).

Adult brain CUT&RUN. Nuclei were isolated from microdissected POA, BNSTp, and MeAp from gonadectomized C57Bl6/J mice, following anatomic designations (Paxinos & Franklin, 2019), as

described previously (Mo et al., 2015). Following tissue douncing, brain homogenate was mixed with a 50% OptiPrep solution and underlaid with 4.8 ml of 30% then 40% OptiPrep solutions, in 38.5 ml Ultra-clear tubes (Beckman-Coulter C14292). Ultracentrifugation was performed with a Beckman SW-28 swinging bucket rotor at 9200 RPM for 18 minutes at 4ºC. Following ultracentrifugation, ~1.5 ml of nuclei suspension was collected from the 30/40% OptiPrep interface by direct tube puncture with a 3 ml syringe connected to an 18-gauge needle. Nuclei concentration was measured on a Countess II FL Automated Cell Counter. For ERα CUT&RUN (1:100, EMD Millipore Sigma 06-935), 400,000 nuclei were isolated from BNST, MPOA, and MeA of 5 animals. 400,000 cortical nuclei were used for the CUT&RUN IgG control (1:100, Antibodies-Online ABIN101961). Prior to bead binding, 0.4% IGEPAL CA-630 was added to the nuclei suspension to increase affinity for concanavalin A magnetic beads. All subsequent steps were performed as described above, with a modified Wash Buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.1% BSA, 0.5 mM spermidine, 1X PIC).

CUT&RUN library preparation. Cell line CUT&RUN libraries were prepared using the SMARTer ThruPLEX DNA-seq Kit (Takara Bio R400676), with the following PCR conditions: 72ºC for 3 min, 85ºC for 2 min, 98ºC for 2 min, (98ºC for 20 sec, 67ºC for 20 sec, 72ºC for 30 sec) x 4 cycles, (98ºC for 20 sec, 72ºC for 15 sec) x 14 cycles (MCF-7). Brain CUT&RUN libraries were prepared using the same kit with 10 PCR cycles. All samples were size-selected with AMPure XP beads (0.5x-1.7x) to remove residual adapters and large genomic DNA. Individually barcoded libraries were multiplexed and sequenced with paired-end 76bp reads on an Illumina NextSeq, using either the Mid or High Output Kit.

CUT&RUN data processing. Paired-end reads were trimmed to remove Illumina adapters and low-quality basecalls (cutadapt -q 30) (Martin, 2011). Trimmed reads were aligned to mm10 using Bowtie2 (Langmead & Salzberg, 2012) with the following flags: --dovetail --very-sensitive-local --no-unal --

no-mixed --no-discordant --phred33. Duplicate reads were removed using Picard (http://broadinstitute.github.io/picard/) MarkDuplicates (REMOVE_DUPLICATES=true). Reads were filtered by mapping quality (H. Li et al., 2009) (samtools view -q 40) and fragment length (Ramírez et al., 2016) (deepTools alignmentSieve --maxFragmentLength 120). Reads aligning to the mitochondrial chromosome and incomplete assemblies were also removed using SAMtools. After filtering, peaks were called on individual replicate BAM files using MACS2 callpeak (--min-length 25 -q 0.01) (Zhang et al., 2008). To identify consensus Nfix peaks across samples, MACS2 callpeak was performed on BAM files merged across biological replicates (n=2) and subsequently intersected across treatment and sex. TF peaks that overlapped peaks called in the IgG control were removed using bedtools intersect (-v) (Quinlan & Hall, 2010) prior to downstream analysis.

CUT&RUN data analysis. CUT&RUN differential peak calling was performed with DiffBind (Stark et al., 2011). A count matrix was created from individual replicate BAM and MACS2 narrowpeak files (n=2 per condition). Consensus peaks were re-centered to ±100 bp around the point of highest read density (summits=100). Contrasts between sex and treatment were established (categories = c(DBA_TREATMENT, DBA_CONDITION)), and edgeR (Robinson et al., 2010) was used for differential peak calling. Differential ERα peaks with padj<0.1 were used for downstream analysis. To identify sex-dependent, estradiol-responsive peaks for adult brain ERα CUT&RUN, the DiffBind consensus peakset count matrix was used as input to edgeR, and an interaction between sex and treatment was tested with glmQLFTest.

Brain E2-induced ERα CUT&RUN peaks were annotated to NCBI RefSeq mm10 genes using ChIPseeker (Yu, Wang, & He, 2015). DeepTools plotHeatmap was used to plot ERα CUT&RUN, representing CPM-normalized bigwig files pooled across replicate and sex per condition, at E2-induced ERα peaks. CUT&RUNTools (Q. Zhu et al., 2019) was used to plot ERα CUT&RUN fragment ends surrounding ESR1 motifs (JASPAR MA0112.3) within E2-induced ERα ChIP-seq peaks. BETA (basic

mode, -d 500000) (S. Wang et al., 2013) was used to determine whether ERα peaks were significantly overrepresented at E2-regulated RNA-seq genes ($p<0.01$), as well as sex-dependent E2-regulated genes ($p<0.01$), compared to non-differential, expressed genes. Motif enrichment analysis of ERα peaks was performed with AME (Bailey et al., 2009) using the 2020 JASPAR core non-redundant vertebrate database. Motif enrichment analysis was performed using a control file consisting of shuffled primary sequences that preserves the frequency of *k*-mers (--control --shuffle--). The following 7 ERα ChIP-seq files were lifted over to mm10 using UCSC liftOver and intersected with E2-induced ERα peaks to identify brain-specific and shared (≥4 intersections) ERα binding sites: uterus [intersection of GEO:GSE36455 (uterus 1) (Hewitt et al., 2012) & GEO:GSE49993 (uterus 2) (Gertz et al., 2013)], liver [intersection of GEO:GSE49993 (liver 1) (Gertz et al., 2013) & GEO:GSE52351(liver 2) (Gordon et al., 2014)], aorta [(Gordon et al., 2014), GEO: GSE52351], efferent ductules [(G. Yao et al., 2017), Supplementary information], and mammary gland [(Palaniappan et al., 2019), GEO: GSE130032]. ClusterProfiler (Yu et al., 2012) was used to identify associations between brain-specific and shared ERα peak-annotated genes and Gene Ontology (GO) Biological Process terms (enrichGO, ont="BP", $padj<0.1$). For Disease Ontology (DO) and HUGO Gene Nomenclature Committee (HGNC) gene family enrichment, brain-specific ERα peak-associated gene symbols were converted from mouse to human using bioMart (Durinck et al., 2009) then analyzed with DOSE (Yu, Wang, Yan, et al., 2015) (enrichDO, $padj<0.1$) and enricher ($padj<0.1$). Log-odds ESR1 and ESR2 motif scores in brain-specific and shared ERα peaks were calculated with FIMO (Grant et al., 2011), using default parameters.

MCF-7 ERα CUT&RUN data were compared to MCF-7 ERα ChIP-seq data from [(Franco et al., 2015); GEO: GSE59530]. Single-end ChIP-seq fastq files for 2 vehicle-treated and 2 17β-estradiol (E2)-treated IP and input samples were accessed from Sequence Read Archive and processed identically as ERα CUT&RUN data, with the exception of fragment size filtering. Differential ERα ChIP-seq peak calling was performed using DiffBind DESeq2 ($padj<0.01$). DeepTools was used to plot CPM-normalized ERα CUT&RUN signal at E2-induced ERα ChIP-seq binding sites. DREME (Bailey,

2011) and AME were used to compare *de novo* and enriched motifs between E2-induced MCF-7 ERα CUT&RUN and ChIP-seq peaks.

Adult RNA-seq data processing and analysis. Reads were adapter-trimmed and quality-filtered (q>30) (http://hannonlab.cshl.edu/fastx_toolkit/), then mapped to the mm10 reference genome using STAR (Dobin et al., 2013). The number of reads mapping to the exons of each gene was counted with featureCounts (Liao et al., 2014), using the NCBI RefSeq mm10 gene annotation. Differential gene expression analysis was performed using DESeq2 (Love et al., 2014) with the following designs: effect of treatment (design = ~ batch + hormone), effect of sex (design = ~ batch + sex), two-way comparison of treatment and sex (design = ~ batch + hormone_sex), four-way comparison (design = ~ 0 + hormone_sex), and sex-treatment interaction (design = ~ batch + sex + hormone + sex:hormone).

ATAC-seq data processing. ATAC-seq data were processed using the ENCODE ATAC-seq pipeline (https://github.com/ENCODE-DCC/atac-seq-pipeline) with default parameters. To generate CPM-normalized bigwig tracks, quality-filtered, Tn5-shifted BAM files were converted to CPM-normalized bigwig files using DeepTools bamCoverage (--binSize 1 --normalizeUsing CPM).

Adult ATAC-seq data analysis. ATAC-seq differential peak calling was performed with DiffBind. A DiffBind dba object was created from individual replicate BAM and MACS2 narrowPeak files (n=3 per condition). A count matrix was created with dba.count, and consensus peaks were re-centered to ±250 bp around the point of highest read density (summits=250). Contrasts between sex and treatment were established (categories = c(DBA_TREATMENT, DBA_CONDITION)), and edgeR was used for differential peak calling. Differential peaks with an FDR<0.05 and abs(log2FC)>1 or abs(log2FC)>0 were used for downstream analysis. DeepTools computeMatrix and plotHeatmap were used to plot mean ATAC CPM at E2-open ATAC peaks. To identify sex-dependent, estradiol-responsive peaks, the

DiffBind consensus peakset count matrix was used as input to edgeR, and an interaction between sex and treatment was tested with glmQLFTest. E2-open ATAC peaks and total Veh or E2 ATAC peaks (intersected across replicate and sex for each treatment condition) were annotated to NCBI RefSeq mm10 genes using ChIPseeker. ClusterProfiler was used to calculate the enrichment of GO Biological Process terms. DO and HGNC gene family enrichment was performed on E2-open ATAC peak-associated genes, as described above for ERα CUT&RUN analysis. BETA (basic mode, -d 500000) (S. Wang et al., 2013) was used to determine whether E2-open ATAC peaks were significantly overrepresented at E2-regulated RNA-seq genes ($p<0.01$), as well as sex-dependent E2-regulated genes ($p<0.01$), compared to non-differential, expressed genes. Motif enrichment analysis of E2-open ATAC peaks was performed with AME, using the 2020 JASPAR core non-redundant vertebrate database.

Chapter 3

Estrogen engages brain MC4R signaling to drive physical activity in female mice

Sections of this chapter were previously published in Nature (2021) under the title "Oestrogen engages brain MC4R signalling to drive physical activity in female mice" by William C. Krause, Ruben Rodriguez, Bruno Gegenhuber, Navneet Matharu, Andreas N. Rodriguez, Adriana M. Padilla-Roger, Kenichi Toma, Candice B. Herber, Stephanie M. Correa, Xin Duan, Nadav Ahituv, Jessica Tollkuhn, and Holly A. Ingraham. W.C.K. designed experiments, analysed data and wrote the manuscript. R.R. performed thermal and glucose homeostasis analyses in mice. B.G. optimized, performed and analysed the CUT&RUN method for ERα binding in neurons. N.M. provided CRISPRa viral vectors and expert advice. A.N.R. performed histology experiments and quantification of expression data. A.M.P.-R. aided with chemogenetic data acquisition and analyses. C.B.H. analysed bone and plasma lipid data. S.M.C. designed experiments, provided animal models and analysed data. K.T. and X.D. provided the AAV-DIO-mYFP vector. N.A. provided key unpublished reagents related to CRISPRa constructs and helped to guide studies. J.T. optimized CUT&RUN method for ERα binding in neurons, performed analyses and wrote the manuscript. H.A.I. designed experiments, analysed data and wrote the manuscript.

3.1 Abstract

In Chapter 2, I identified the first genomic targets of ERα in the adult brain and found that the majority are unique to this tissue. As these prior experiments only examined the transcriptomic and chromatin accessibility response to estradiol, it remains unclear whether and how the genomic targets of ERα regulate the display of sex-typical behaviors. Here, in collaboration with Holly Ingraham's lab, we found that one particular ERα target, melanocortin receptor 4 (*Mc4r*), is upregulated by estradiol across

the estrous cycle in the VMHvl – a brain region that regulates maternal and male-typical aggression (Hashikawa et al., 2017; Lin et al., 2011) as well as energy expenditure (Krause & Ingraham, 2017). In particular, ERα binds *Mc4r* at a downstream ERE-containing locus as well as a half-ERE site within the promoter, in addition to other estradiol-regulated genes within this region, such as *Pgr*, *Greb1*, and *Nmur2*. Genetic deletion of *Mc4r* increased body weight and sedentary behavior in both sexes, whereas reinstating *Mc4r* selectively within the VMHvl rescued this phenotype in females but not in males. Moreover, VMHvl delivery of a CRISPR-based activator (CRISPRa) directed to the *Mc4r* promoter half ERE site increased physical activity and, in response to the additional mechanical loading, cortical bone volume and thickness. Together, these results reveal estradiol regulation of VMHvl *Mc4r* signaling drives female activity and provide a model for how genomic recruitment of ERα activates the sex-specific display of an innate behavior.

3.2 Introduction

While decades of studies have revealed the rapid effects of estradiol on spatial memory and neuronal physiology, particularly in hippocampal CA1 pyramidal neurons (Woolley, 2007), the genomic targets of ERα in the brain have only now been defined. The next logical question is to determine whether these targets influence the properties of neurons or circuits in which they are regulated and, ultimately, modify behavior. However, rather than examine each individual ERα target gene, of which several presumably do not result in a functional change at the protein level, it is first critical to examine genes implicated in the regulation of behaviors known to be influenced by estradiol. For instance, GDX followed by estradiol and progesterone replacement has been used for decades to model female-typical reproductive behaviors that occur naturally throughout the estrous cycle (McEwen, 1981). These behaviors span not only the display of lordosis but also underlying changes in food intake, physical activity, and thermogenesis that facilitate reproductive success (Krause & Ingraham, 2017).

Accordingly, a number of the ERα genomic targets identified in Chapter 2 have been implicated in the regulation of sexual receptivity and metabolism. Among these are *Pgr*, the gene encoding PR, as well as a number of neuropeptides and/or their receptors, such as *Cartpt (S. J. Lee et al., 2020)*; *Cck* and *Cckar (Cooper & Dourish, 1990)*; *Npy*, *Npy2r*, and *Npy4r* (Sohn et al., 2013), *Nmur2 (Howard et al., 2000)*; and *Mc4r* (Huszar et al., 1997). *Mc4r* is of particular interest, as it was first associated with monogenic human obesity in 1998 (Yeo et al., 1998) and is considered the most common known genetic cause of this disease (Farooqi et al., 2000; Wade et al., 2021). Consistent with the estradiol-dependent recruitment of ERα to the *Mc4r* locus, women are more susceptible to loss-of-function (LOF) mutations in this gene than men (Horstmann et al., 2013; Qi et al., 2008). Mechanistically, *Mc4r* encodes a G-protein coupled seven-transmembrane receptor expressed in discrete populations of neurons throughout the brain (Gantz et al., 1993; Mountjoy et al., 1994). α- and β-melanocyte-stimulating hormones (MSH), which derive from the neuropeptide pro-opiomelanocortin (POMC), bind to Mc4r and suppress food intake (Cowley et al., 1999). Leptin, a satiety hormone, acts upstream of this pathway by stimulating the production of POMC from neurons in the arcuate nucleus of the hypothalamus (ARC) (Cowley et al., 2001).

In this Chapter, I collaborated with the Ingraham lab to examine how ERα regulation of *Mc4r* regulates energy expenditure in mice. We found that *Mc4r* expression is induced within the VMHvl during proestrus and in response to estradiol replacement. Moreover, *Mc4r* upregulation is controlled by the genomic action of ERα, rather than indirect signaling, as ERα bound *Mc4r* at a downstream ERE-containing locus as well as a half-ERE site within the promoter. Reinstating *Mc4r* expression within the VMHvl of *Mc4r* KO animals rescued body weight and sedentary behavior in females but not in males. Finally, bypassing estradiol-ERα signaling by co-delivering dCas9-VP64 activator and a sgRNA targeting the *Mc4r* promoter half ERE site into the VMHvl increased physical activity and cortical bone volume and thickness. Together, these experiments shed light on the sex-specific role of

Mc4r in energy expenditure and provide physiological and behavioral relevance for a genomic target of ERα.

3.3 Results

3.3.1 Estradiol regulation of *Mc4r* within the VMHvl

We first tested whether ERα signaling within the VMHvl is required for the display of physical activity by stereotactically delivering AAV-Cre-GFP (ERα KO) or AAV-GFP (control) into the VMHvl or ARC of female *Esr1*$^{fl/fl}$ mice (Fig. 3.1a). VMHvl^{ERαKO}, but not ARC^{ERαKO}, females displayed increased body weight and reduced ambulatory activity (Fig. 3.1a), demonstrating ERα acts in a region-specific manner to influence female energy expenditure (Fig. 3.1a).

The activity of female mice naturally changes across the estrous cycle, peaking in proestrus, presumably as a means to enhance sexual receptivity (Della Torre & Maggi, 2017). To examine whether the activity of VMHvl ERα+ neurons changes across estrous, we stained for ERα and phosphorylated ribosomal protein S6 (pS6) – a known marker of neural activity (Fig. 3.1b). The number of VMHvl ERα+/pS6+ neurons peaked at proestrus (high estradiol state), or following injection of estradiol benzoate (EB) into GDX mice (Fig. 3.1b-c). Consistent with the lack of a role of ARC ERα+ neurons in female activity, estradiol did not influence the number of ARC ERα+/pS6+ neurons (Fig. 3.1c). Together, these findings reveal VMHvl ERα+ neurons increase activity at proestrus as a result of estradiol signaling.

To determine whether estradiol influences gene expression within the VMHvl, we performed RNA-seq in GDX females treated acutely with vehicle or EB (Fig. 3.1d). This approach revealed many estradiol-regulated genes shared across the VMHvl and BNSTp (Chapter 2), such as *Greb1*, *Pgr*, *Irs2*, and *Myrip*, as well as VMHvl-specific, estradiol-responsive genes, such as *Phf21b*, *Mad2l1*, *Gadd45a*, and *Gpr6* (Fig. 3.1d). Importantly, estradiol influenced the expression of several neuropeptide receptors involved in metabolism (*Mc4r*, *Nmur2*, *Npy1r*) (Fig. 3.1d). Among these, *Mc4r* stood out as a potential

mediator by which VMHvl ERα signaling promotes female activity, as sex differences have previously been observed for *Mc4r* LOF mutations in mice (Huszar et al., 1997; Ste Marie et al., 2000) and humans (Sina et al., 1999). Moreover, *Mc4r* increased expression in the VMHvl during proestrus compared to estrus and gonadally intact males (Fig. 3.1e-f), validating its regulation by physiological estradiol production.

To identify whether estradiol-responsive genes within the VMHvl are direct targets of ERα, I examined ERα binding sites identified using CUT&RUN and found ERα recruitment at *Mc4r*, in addition to highly-conserved loci at other estradiol target genes *Greb1*, *Pgr*, and *Nmur2* (Fig 3.1g, 3.2a-b, 3.2e). In particular, ERα bound *Mc4r* -210 kb downstream of the gene at a region containing a full ERE and at a promoter site containing a half ERE adjacent to a Sp1 motif, consistent with ERα genomic tethering (Fig. 3.1g, 3.2c-d). Together these data demonstrate a genomic pathway by which ERα upregulates *Mc4r* expression within the VMHvl and suggests dynamic induction of this pathway across estrous may play a role in promoting female physical activity.

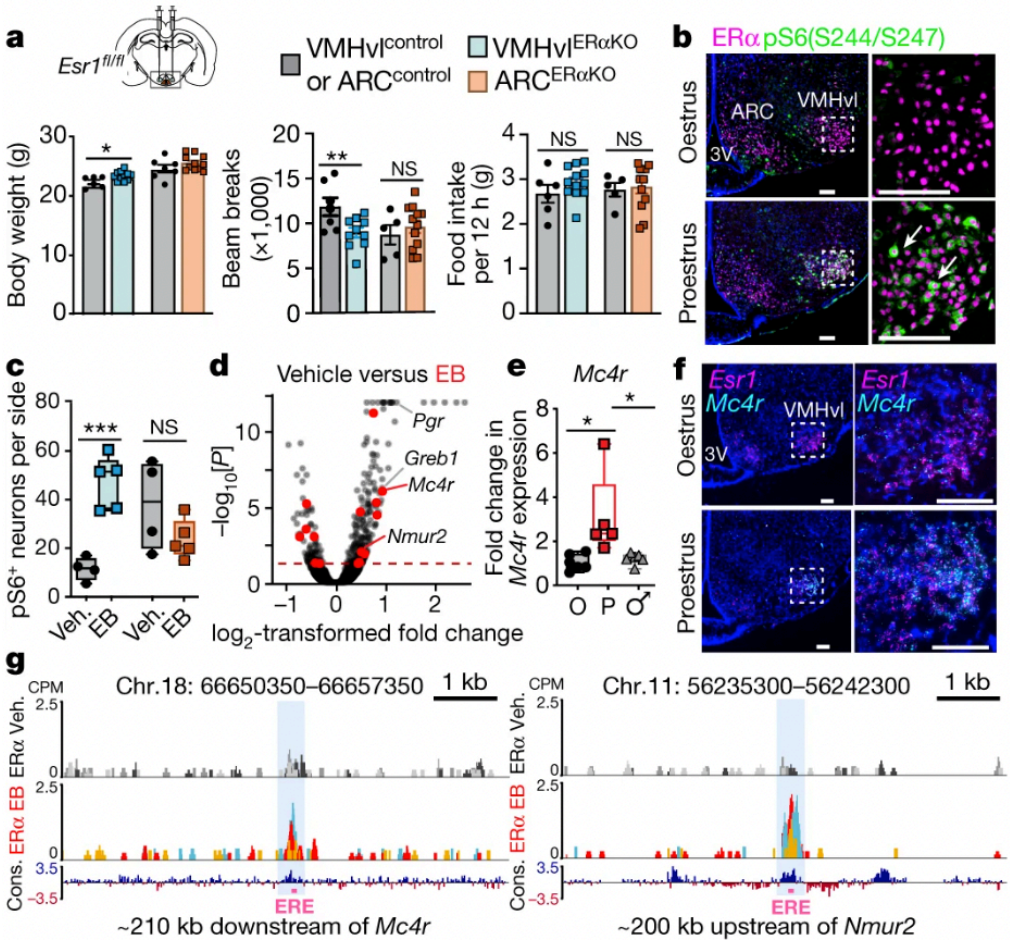

Figure 3.1. VMHvl neurons are sensitive to estradiol and maintain energy expenditure in adult female mice. (**a**) Body weight (**P* = 0.0310) at 12 weeks after infection, *x*-ambulatory activity during the dark period (***P* = 0.0080) and food intake during the dark period in VMHvl^{ERαKO} (*n* = 10), ARC^{ERαKO} (*n* = 12) and control (grey, *n* = 7 and 5) female mice. NS, not significant. (**b**) ERα and pS6(S244/S247) co-expression (arrows) in proestrus and oestrus (representative of five mice). Scale bars, 100 µm. (**c**) Number of pS6(S244/S247)-labelled VMHvl (****P* = 0.0005) and ARC cells in vehicle-treated (Veh.) (*n* = 4) or oestradiol-benzoate-treated (EB) (*n* = 5) female mice. (**d**) Enrichment of peptide ligand-binding receptors (red). Dashed line, Benjamini–Hochberg-adjusted *P* < 0.05. (**e**) VMHvl *Mc4r* expression in proestrus (P; *n* = 5), male (*n* = 6, **P* = 0.0189) and oestrus (O; *n* = 5, **P* = 0.0163) mice. (**f**) *Mc4r* and *Esr1* expression in oestrus and proestrus (representative of five mice). Scale bars, 100 µm. (**g**) Counts-per-million (CPM)-normalized coverage tracks of ERE-containing ERα-binding sites (pink boxes) within *Mc4r* and *Nmur2* loci in sub-cortical nuclei of vehicle- and oestradiol-benzoate-treated gonadectomized mice (*n* = 3; MACS2, *q* < 0.01). phyloP60wayPlacental track from UCSC shows sequence conservation (Cons.) of ERα-binding sites in placental mammals. Data are mean ± s.e.m., scatter plots or box plots. In box plots, whiskers indicate the minimum and maximum values, the edges of the box are 25th and 75th percentiles, and the centre line indicates the mean. (**a**, **c**) Unpaired two-tailed Student's *t*-test. (**e**) One-way ANOVA with Holm–Šidák multiple comparisons test.

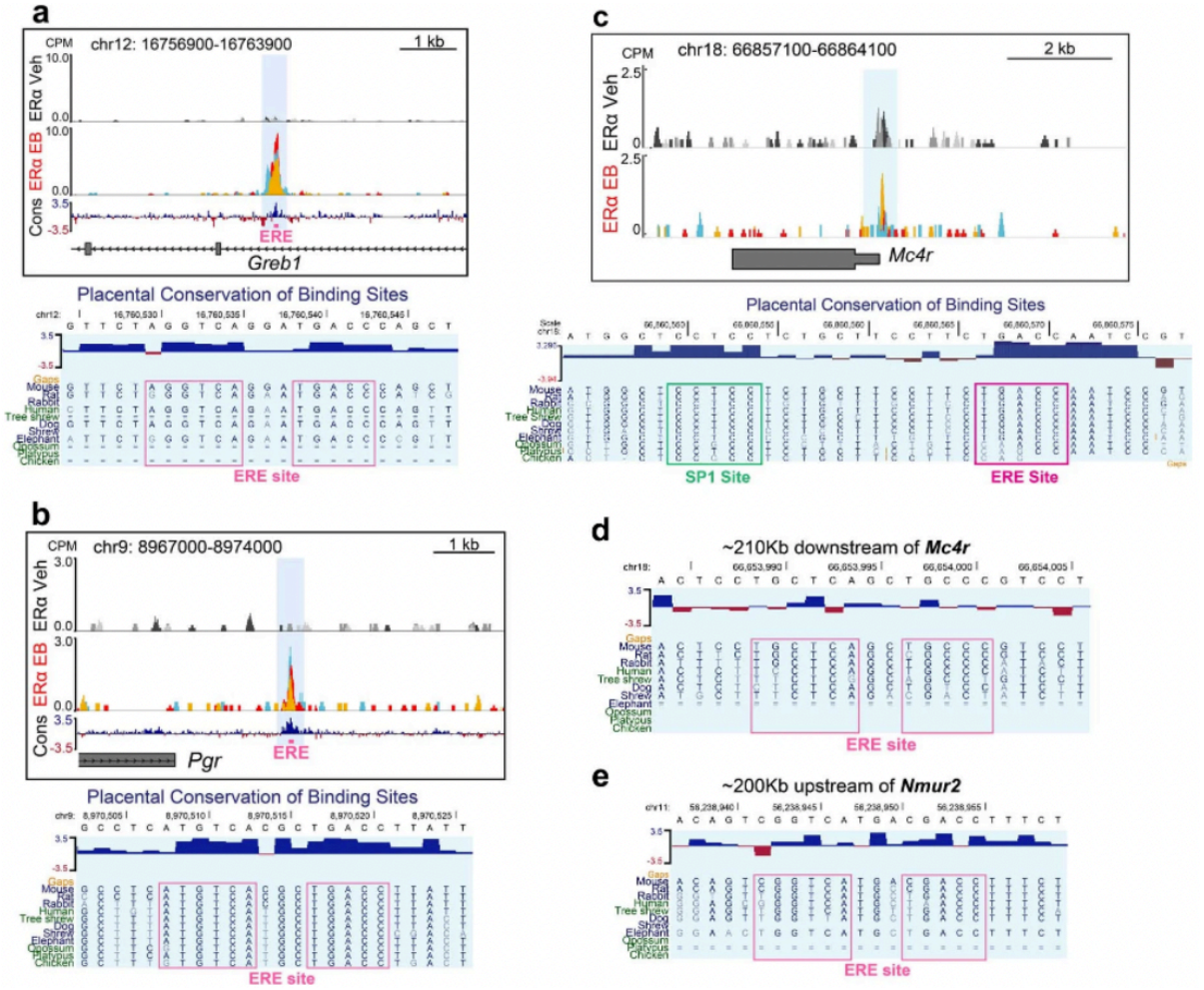

Figure 3.2. ERα-binding sites in estradiol-sensitive target genes contain conserved ERE consensus sequences. CUT&RUN CPM-normalized coverage track showing oestradiol-benzoate-specific ERα binding sites containing EREs (pink boxes) within the *Greb1* locus (**a**; 3/3 replicates) and *Pgr* locus (**b**; 3/3 replicates), and in the *Mc4r* promoter (**c**; 1 of 3 replicates) in 400,000 sub-cortical brain nuclei collected from vehicle and oestradiol benzoate (5 µg) treated gonadectomized mice. Below each track the location and sequence conservation of full (**a**, **b**) ERE and half (**c**) SP1/ERE consensus sites in target gene loci indicated by pink and green boxes. (**d**, **e**) Location and sequence conservation of ERE consensus sites within *Mc4r* and *Nmur2* loci corresponding to ERα binding sites presented in Fig. 3.1g.

3.3.2 *Mc4r* signaling in the VMHvl regulates physical activity

To examine whether *Mc4r* signaling in the VMHvl regulates physical activity, we crossed Cre-dependent *Mc4r*^{loxTB} animals, which normally are obese (Balthasar et al., 2005), to the *Sf1*^{Cre} line (Dhillon et al., 2006), which selectively restores *Mc4r* expression in the VMHvl (Fig. 3.3a). *Mc4r*^{loxTB} and *Sf1*^{Cre}; *Mc4r*^{loxTB} mice of both sexes developed obesity, as *Sf1*^{Cre} does not reinstate *Mc4r* in the paraventricular hypothalamus (PVH) – a critical node in the regulation of food intake (Balthasar et al., 2005). However, rescue of *Mc4r* within the VMHvl reduced weight gain and sedentary behavior in females but not in males (Fig. 3.3b-e), demonstrating *Mc4r* signaling within the VMHvl regulates female physical activity.

To determine whether *Mc4r* induction is sufficient to drive this behavior, we bypassed hormonal regulation of *Mc4r* by co-injecting a dual-AAV CRISPRa system, consisting of a sgRNA targeting the ERE half-site in the *Mc4r* promoter and a dCas9-VP64 activator, into the VMHvl of wildtype animals (Fig. 3.3f). We validated CRISPRa induction of *Mc4r* relative to control animals receiving only dCas9-VP64 and no sgRNA by FISH (Fig. 3.3g). *Mc4r*^{CRISPRa} animals of both sexes traveled longer distances during the dark (active) phase for weeks post-injection compared to controls (Fig. 3.3h-i). *Mc4r*^{CRISPRa} females did not lose weight (data not shown), yet this may be attributed to a modest increase in daily food intake by these animals. After weeks of elevated physical activity, and consequently mechanical loading, *Mc4r*^{CRISPRa} females increased cortical bone thickness and volume (Fig. 3.3j) Together, these results demonstrate that *Mc4r* acts within VMHvl neurons to promote physical activity in female mice and provide a model for how estradiol influences energy expenditure via ERα genomic recruitment.

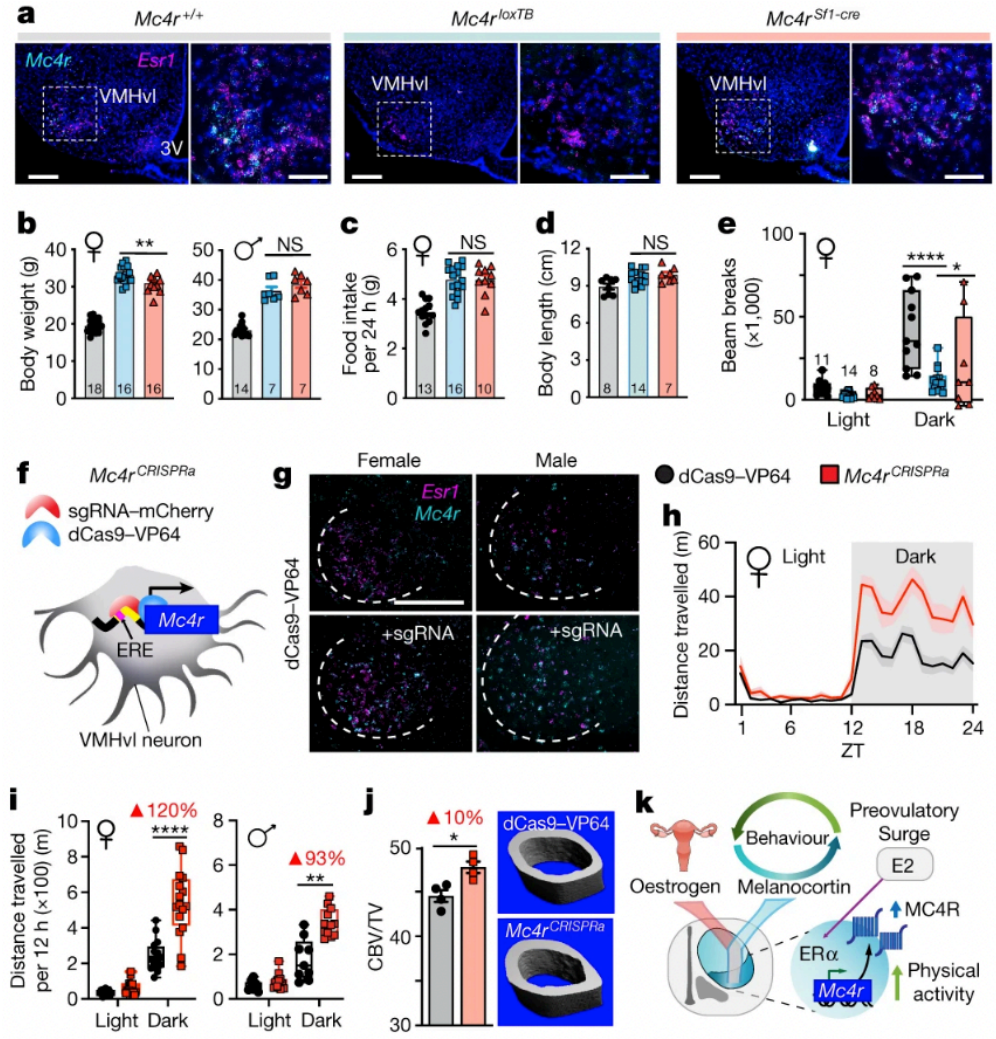

Figure 3.3. Sex-specific role for MC4R signalling in the VMHvl can be bypassed using CRISPRa. (**a**) *Esr1* and *Mc4r* expression in *Mc4r*+/+, *Mc4r*^{loxTB} and *Mc4r*$^{Sf1\text{-}cre}$ female mice. Scale bars, 200 µm (left) and 100 µm (right). (**b**) Body weights of 8-week-old female (**$P = 0.0026$) and male *Mc4r*+/+, *Mc4r*^{loxTB} and *Mc4r*$^{Sf1\text{-}cre}$ mice. (**c**) Food intake in 8-week-old female mouse cohorts. (**d**) Body length in 8-week-old female mouse cohorts. (**e**) *x*-ambulatory activity of female mice during the light and dark periods (****$P < 0.0001$, *$P = 0.0153$). (**f**) *Mc4r*^{CRISPRa} targets the *Mc4r* promoter. (**g**) *Esr1* and *Mc4r* expression in control and *Mc4r*^{CRISPRa} female and male mice. Scale bar, 200 µm. (**h**) Home-cage activity in *Mc4r*^{CRISPRa} ($n = 6$) and control ($n = 5$) female mice. (**i**) Distances for the three most active runs of *Mc4r*^{CRISPRa} and control female ($n = 6$ and 5, respectively, ****$P < 0.0001$) and male ($n = 4$ and 3, respectively) mice. (**j**) Fraction of cortical bone volume of female mice 4 months after infection ($P = 0.0129$). CBV, cortical bone volume; TV, tissue volume. (**k**) VMHvl$^{ERα/MC4R}$ neurons integrate oestrogen and melanocortin signalling to generate a specialized hormone-dependent activity node in female mice. Data are mean ± s.e.m. or box plots, and number of mice analysed are indicated on or above each bar. (**b**, **c**, **d**) One-way ANOVA with Holm–Šidák multiple comparison test. (**e**, **i**) Repeated-measures two-way ANOVA with Holm–Šidák multiple comparison test. (**j**) Unpaired two-tailed Student's *t*-tests.

3.4 Discussion

Prior studies have identified *Mc4r* signaling as a cause of monogenic human obesity (Farooqi et al., 2000; Yeo et al., 1998). Moreover, sex differences in *Mc4r* LOF mutation phenotypes have been observed in mice (Huszar et al., 1997; Ste Marie et al., 2000) and humans (Sina et al., 1999), indicating *Mc4r* undergoes differential regulation between sexes, possibly as a result of gonadal hormonal signaling. In this Chapter, we found that estradiol-responsive VMHvl ERα+/Mc4r+ neurons promote the display of physical activity in females. *Mc4r* increases expression during proestrus, a period of high estradiol production and energy expenditure. As the genomic regions bound by ERα at the *Mc4r* locus are highly conserved across vertebrate species, it is possible that a similar mechanism also regulates *MC4R* expression in the human VMHvl. Along these lines, loss of estradiol-dependent *MC4R* expression may contribute to the rise in sedentary behavior observed in post-menopausal women.

This study focused on the role of *Mc4r* in the regulation of female physical activity. However, ERα binds many other genes throughout the social behavior network (Chapter 2), several of which may also influence energy expenditure, such as *Nmur2*. In addition, the genomic targets of ERα may coordinate other estradiol-dependent processes, such as sexual differentiation of the mouse brain, which I examine in closer detail in Chapter 5. Future studies will reveal how the genes regulated by ERα, either singly or combinatorially, mediate the effects of estradiol on neural circuit function and behavior. For instance, with multiplexed CRISPRa (Savell et al., 2020), it is now possible to perturb combinations of ERα target genes and assess their collective influence on the display of sex-typical behaviors.

Besides *Mc4r*, how do physiological sex differences in gonadal hormone levels give rise to sex differences in gene regulation in the brain? In the next Chapter, I examine this question by profiling the chromatin of gonadally intact and GDX animals of both sexes.

3.5 Materials and Methods

Mice. All experiments were conducted in accordance with institutional guidelines and approved protocols for animal care and use at the University of California San Francisco (UCSF) and Cold Spring Harbor Laboratory. Mice were housed on a 12 h:12 h light:dark cycle (lights on, 06:00; lights off, 18:00) and had ad libitum access to standard chow (LabDiet, 5058) or high-fat diet (Research Diets, D12492). *Mc4r*^{loxTB} mice and the *Ai14*$^{fl/fl}$ reporter mice were purchased from Jackson Laboratories and maintained on a C57BL/6J background. *Mc4r-t2a-cre* mice (Garfield et al., 2015) were a gift from B. Lowell and were maintained on a C57BL/6J background. *Esr1*$^{fl/fl}$ mice were maintained on a mixed background, and *Sf1-cre* mice (Vong et al., 2011) were maintained on a C57BL/6N in the laboratory as previously described (Correa et al., 2015; Herber et al., 2019). Wild-type mice used for CRISPRa studies were on a pure C57BL/6J background. For *Mc4r* rescue experiments, *Sf1-cre* was contributed through female mice. CUT&RUN experiments were performed on adult male (8–12 weeks of age) gonadectomized C57Bl6/J wild-type mice obtained from Jackson Laboratories. Three weeks after gonadectomy, animals were injected subcutaneously with either corn oil (vehicle) or 5 µg of oestradiol benzoate and euthanized after 4 h. For each biological replicate, brain dissections were pooled from five animals.

Stereotaxic injections. AAV2-Cre-GFP and AAV2-GFP were purchased from the Vector Core at the University of North Carolina at Chapel Hill. AAV2-hM3Dq-mCherry and AAV2-hM4Di-mCherry were gifts from B. Roth and viral preparations were purchased from Addgene (viral prep no. 44361-AAV2, http://n2t.net/addgene:44361, RRID: Addgene_44361; and viral prep no. 44362-AAV2, http://n2t.net/addgene:44362, RRID: Addgene_44362; Addgene) (Krashes et al., 2011). For axonal tracing, AAV2-CAGs-FLEX-membrane-YFP-WPRE.hGH was re-engineered to fuse YFP with a C-terminal farnesylation tag to enhance the membrane labelling (Cai et al., 2013). AAV2 was prepared using a standard polyethylene-glycol gradient followed by a caesium-chloride density-gradient centrifugation protocol (Y. Hu et al., 2012) to reach a titre of (1×10^{13} genome copies per ml). AAVdj-

dCas9-VP64 and AAVdj-Prm-Mc4r-sgRNA were generated by the Stanford Gene Vector and Virus Core and details of vector constructs are as previously described (Matharu et al., 2019). Adult mice were secured in a Model 1900 stereotaxic frame (David Kopff Instruments) and 250–600 nl of virus was injected bilaterally at the following coordinates. For the VMHvl, anterior–posterior: bregma −1.48 mm, mediolateral: bregma ±0.85 mm, dorsoventral: skull −5.9 mm. For the ARC, anterior–posterior: bregma −1.58 mm, mediolateral: bregma ±0.25 mm, dorsoventral: skull −5.8 mm. For all surgeries regardless of viral vectors used, mice recovered for at least 2 weeks before any metabolic or behavioral assays.

Estrous cycle staging and oestradiol benzoate treatment. Reproductive stages in female mice were determined by comparing relative amounts of leukocytes, epithelial cells and cornified epithelial cells collected by vaginal lavage. Stage assessments were made daily between ZT3 and ZT5. Brains from oestrus or proestrus female mice were collected between ZT7 and ZT10 and processed for immunofluorescence, in situ hybridization (ISH) or qPCR.

Adult female mice (>8 weeks old) were ovariectomized. Oestradiol benzoate (Cayman Chemical, 10006487) was dissolved in DMSO and diluted in sesame oil (Sigma, S3547). Mice received a subcutaneous injection of either 1 ug oestradiol benzoate in 150 µl sesame oil or 150 µl of sesame oil with an equivalent amount of DMSO. Control mice received a subcutaneous injection of 150 µl of sesame oil with an equivalent amount of DMSO. To minimize changes in VMH gene expression or signal transduction associated with fear and/or anxiety, mice were handled daily in a manner that simulated injection for at least 5 days before oestradiol benzoate or vehicle treatment and tissue collection. For FOS analyses, mice were treated with 400 µg MT-II (Bachem) by intraperitoneal injection, and brains were collected 1–1.5 h later.

RNA sequencing and qPCR. Brains from OVX female mice treated with oestradiol benzoate ($n = 4$) or vehicle ($n = 3$) were rapidly dissected into ice-cold PBS with 0.1% DEPC. Coronal brain sections (250 µm thick) were cut on a vibratome and transferred to glass slides so that the VMH could be visualized and manually microdissected. Isolated tissue was flash-frozen and stored at −80 °C. RNA was prepared using the RNeasy Micro kit (Qiagen). Sequencing libraries were constructed using the TRIO RNA-seq Library Preparation kit (TECAN) using 15 ng of input RNA. Equal amounts of each sample library were multiplexed and sequenced (50-bp single-end reads) on a single flow cell lane HiSeq 4000 (Illumina). Demultiplexed reads were aligned to the mouse genome (mm10) using HISAT2 (D. Kim et al., 2015) v.2.1.0 and counted using HTSeq (Anders et al., 2015) v.0.9.1. Finally, differential gene expression testing was performed using DESeq2 (Love et al., 2014) v.1.14.1.

Isolated RNA, prepared as described above, was converted to cDNA using the SuperScript III reverse transcriptase (Invitrogen). qPCR was performed using a BioRad CFX instrument with Maestro software v.4.1.2433.1219. Target genes were amplified using specific primers (*Mc4r* forward, 5′-GCCAGGGTACCAACATGAAG-3′ and reverse, 5′-ATGAAGCACACGCAGTATGG-3′; *Nmur2* forward, 5′-CCTCCTTCCTCTTCTACATCCT-3′ and reverse, 5′-AGTCACTTTGTCTGCCTCAA-3′; *Esr1* forward, 5′-GAACGAGCCCAGCGCCTACG-3′ and reverse, 5′-TCTCGGCCATTCTGGCGTCG-3′; and *Ucp1* forward, 5′-CACGGGGACCTACAATGCTT-3′ and reverse, 5′-TAGGGGTCGTCCCTTTCCAA-3′). *C*t values were normalized to cyclophilin B (*Ppib*; forward primer, 5′-TGGAGAGCACCAAGACAGACA-3′ and reverse primer, 5′-TGCCGGAGTCGACAATGAT-3′) and relative expression levels were quantified using the comparative *C*t method. Individual values, representing the VMHvl or iBAT from one mouse are the average of two technical replicates.

CUT&RUN. ERα CUT&RUN was performed on 400,000 nuclei isolated from BNSTp, POA and MeA tissue using density gradient centrifugation (Mo et al., 2015). In brief, tissue was homogenized 15×

with a loose pestle in a glass homogenizer containing homogenization medium (250 mM sucrose, 25 mM KCl, 5 mM MgCl2, 20 mM Tricine KOH, 1 mM DTT, 0.15 mM spermine, 0.5 mM spermidine, 1× Roche EDTA-free protease inhibitor cocktail, pH 7.8). Then, 0.3% IGEPAL CA-630 was added, and the tissue was further dounced 5× with a tight pestle. After douncing, the homogenate was filtered through a 40-µm strainer and mixed 1:1 with 50% OptiPrep solution (Millipore Sigma) prepared in dilution buffer (150 mM KCl, 30 mM MgCl2, 120 mM Tricine KOH, pH 7.8). The homogenate was underlaid with 5 ml each of 30% and 40% OptiPrep solution, and centrifuged at 10,000*g* for 18 min at 4 °C in an ultracentrifuge. Then, 2 ml of nucleus-containing solution was removed from the 30–40% OptiPrep interface by direct tube puncture. After the isolation of the nuclei, 0.4% IGEPAL CA-630 was added to improve binding to concanavalin A magnetic beads (Bangs Laboratories BP531). CUT&RUN was performed on brain nuclei, according to the standard protocol (Skene et al., 2018). Nuclei were washed twice in wash buffer (20 mM HEPES, pH 7.5, 150 mM NaCl, 0.1% BSA, 0.5 mM spermidine, 1× protease inhibitor cocktail) and incubated overnight on a nutator with ERα antibody (Millipore Sigma, 06-935), diluted 1:100 in antibody buffer (wash buffer containing 2 mM EDTA). Nuclei were washed twice in wash buffer, and around 700 ng ml^{-1} protein-A-MNase (pA-MNase) was added. After 1 h incubation on a nutator at 4 °C, the nuclei were washed twice in wash buffer and placed in a metal heat block on ice. pA-MNase digestion was initiated by 2 mM CaCl2. After 90 min, pA-MNase activity was stopped by mixing 1:1 with 2× stop buffer (340 mM NaCl, 20 mM EDTA, 4 mM EGTA, 50 µg ml−1 RNase A, 50 µg ml−1 glycogen). Digested fragments were released by incubating at 37 °C for 10 min, followed by centrifuging at 16,000*g* for 5 min at 4 °C. DNA was purified from the supernatant by phenol–chloroform extraction.

CUT&RUN library preparation. CUT&RUN libraries were prepared using the SMARTer ThruPLEX DNA-seq Kit (Takara Bio), with the following PCR conditions: 72 °C for 3 min, 85 °C for 2 min, 98 °C for 2 min, (98 °C for 20 s, 67 °C for 20 s, 72 °C for 30 s) for 4 cycles, (98 °C for 20 s, 72 °C

for 15 s) for 10 cycles. Samples were size-selected with AMPure XP beads (1.5× right-sided and 0.5× left-sided) to remove residual adapter dimers and large DNA fragments. Individually barcoded libraries were multiplexed and sequenced with paired-end 75-bp reads on an Illumina NextSeq, using the High Output Kit.

CUT&RUN data processing. Paired-end reads were trimmed with cutadapt (Martin, 2011) v.3.2.0 to remove low-quality base calls (-q 30) and adapters. Trimmed reads were aligned to mm10 using Bowtie2 (Langmead & Salzberg, 2012) v.2.4.2 with the following flags: -dovetail -very-sensitive-local -no-unal -no-mixed -no-discordant -phred33. After alignment, duplicate reads were removed using Picard v.2.21.6 (http://broadinstitute.github.io/picard/) MarkDuplicates (REMOVE_DUPLICATES = true). Deduplicated reads were filtered by mapping quality (MAPQ > 40) using samtools (H. Li et al., 2009) v.1.11.0 and fragment length (<120 bp) using deepTools v.3.5.0 alignmentSieve (Ramírez et al., 2016). After filtering, peaks were called using MACS2 v.2.2.7.1 callpeak (Zhang et al., 2008) with a *q*-value threshold of 0.01 and min-length set to 25. Peaks shown in Fig. 3.1 were called in 2 out of 3 replicates. Individual replicate BAM files were normalized by CPM and converted to bigwig tracks, using deepTools bamCoverage (-bs 1, -normalize using CPM). CPM-normalized bigwig tracks for individual oestradiol benzoate and vehicle samples ($n = 3$ per condition) were plotted using Gviz (Hahne & Ivanek, 2016).

ISH. For colorimetric ISH, antisense *Mc4r* probes were PCR-amplified (forward primer, 5′-ACTCTGGGTGTCATAAGCCTGT-3′ and reverse primer, 5′-TCTGTCCCCCACTTAATACCTG-3′) from hypothalamic cDNA libraries, and in vitro transcribed with incorporation of digoxigenin-UTP (Roche) using the T7 or SP6 Riboprobe kit (Promega). The 20-µm sections from fixed tissue were labeled and detected by chromogenic immunohistochemistry as previously described (Correa et al., 2015). Fluorescent ISH was performed using RNAScope (ACD, Multiplex Fluorescent v.2) according

to the manufacturer's protocol using the following probes: *Esr1* (478201), *Mc4r* (319181-C2) and *Rprm* (466071).

Immunofluorescence staining and histology. Fixed central nervous system tissue was cryosectioned (20 µm) and stained overnight with primary antibodies against: ERα (EMD Millipore, C1355, polyclonal rabbit, 1:750 dilution or Abcam, 93021, monoclonal mouse, 1:100 dilution), pS6(S244/S247) (RPS6) (Invitrogen, 44-923G, polyclonal rabbit, 1:500 dilution), FOS (Santa Cruz, SC-52, polyclonal rabbit, 1:500 dilution) or red fluorescent protein (RFP; Rockland, 600-401-379, polyclonal rabbit, 1:1,000 dilution). For detection, sections were labeled with species-appropriate secondary Alexa-Fluor-coupled antibodies (Invitrogen, A11029 and A11037, 1:1,000 dilution for both). Widefield images were acquired using a Nikon microscope and NIS-Elements v.3.22.15. Confocal images were acquired at the UCSF Nikon Imaging Center using a Nikon CSU-22 with EMCCD camera and MicroManager v.2.0gamma. Images were processed and quantified using ImageJ FIJI v.1.52i and the Cell Counter plugin v.2.

Fixed gonadal white adipose tissue was paraffin-embedded, sectioned (5 µm) and stained with haematoxylin and eosin by the Gladstone Histology and Light Microscopy core. Brightfield images were thresholded to define adipocyte borders, and the adipocyte area was quantified using ImageJ FIJI.

Micro-computed tomography. After perfusion fixation, femurs from *Mc4r*^{CRISPRa} and control female mice were isolated. Volumetric bone density and bone volume were measured by micro-computed tomography as previously described (Herber et al., 2019).

Chapter 4

Activation of sex differences in gene regulation in the BNST

Sections of this chapter were previously uploaded to bioRxiv (2020) under the title "Regulation of neural gene expression by estrogen receptor alpha" by Bruno Gegenhuber, Melody V. Wu, Robert Bronstein, and Jessica Tollkuhn. This manuscript is now in press at Nature (2022) under the title "Gene regulation by gonadal hormone receptors defines neuronal sex differences" by Bruno Gegenhuber, Melody V. Wu, Robert Bronstein, and Jessica Tollkuhn. I thank Melody V. Wu for helping collect the immunofluorescence data shown in this chapter.

4.1 Abstract

Across many mammalian species, females and males differ in the display of mating, territoriality, and aggressive behaviors critical for reproduction and survival. While the neural circuits underlying these sex differences have been extensively characterized in the mouse brain, the genomic mechanisms by which gonadal hormone receptors act to specify sex differences within these circuits remain unknown. In this Chapter, I re-analyze a single-nucleus RNA-seq (snRNA-seq) dataset to characterize sex differences in cell type abundance and gene expression within a central node in the circuitry of sex-typical behaviors: the BNSTp. I find two neuron types, marked by TFs *Nfix* and *Esr2*, respectively, are more abundant in males than in females. I also detect extensive sex differences in gene expression across neuron types, with nearly all sex differences restricted to hormone receptor-expressing populations. Bulk ATAC-seq of BNSTp *Esr1*+ neurons from gonadally intact and GDX animals further revealed that sex differences in the underlying chromatin landscape largely depend on adult male gonadal hormones. Together, these data provide a molecular and cellular basis to the regulation of sex-

typical behaviors by the BNSTp and establish that the activational effects of gonadal hormones are required for adult sex differences in gene regulation.

4.2 Introduction

How do gonadal hormones activate sex differences in the brain? As discussed in Chapter 1, gonadal hormones bind gonadal hormone receptors, which principally act as TFs. During puberty, females and males initiate patterns of gonadal hormone production that persist until old age. The female ovaries secrete progesterone and estradiol throughout the estrous cycle, while the male testes secrete testosterone. The activational effects of these adult hormones is required for the full display of sex-typical reproductive, territorial, and aggressive behaviors (McEwen, 1981). Therefore, understanding how sex differences in adult hormone production give rise to sex differences in gene expression is essential to understanding how hormones influence circuit function and behavior.

Surprisingly, until recently, few studies have used bulk or single-cell RNA sequencing to identify transcriptomic sex differences within the social behavior network. The classic example of defining sex-biased genes in the brain was carried out using cDNA microarrays (X. Xu et al., 2012). By employing a well-designed experimental strategy and rigorous histological validation, the authors discovered 15 sex-biased genes, the majority of which were more abundant in males than in females. Moreover, male GDX feminized the expression of nearly all of these sex-biased genes, whereas female GDX largely did not influence their expression. This finding suggests that adult sex differences are driven primarily by male gonadal hormones, which can both activate and repress transcription.

More recently, the Correa, Anderson, Stuber, and Dulac labs have employed scRNA-seq and/or spatial transcriptomics to identify sex differences in gene expression in the VMH (D.-W. Kim et al., 2019; van Veen et al., 2020) and MPOA (Hashikawa et al., 2021; Moffitt et al., 2018a). These studies revealed additional sex-biased genes throughout the social behavior network; however, most lacked sufficient numbers of cells, transcript counts per cell, and/or biological replicates to systematically

detect sex differences in gene expression within individual neuron types. For instance, Hashikawa *et al.* 2021 sequenced ~59,000 MPOA cells across 8 different experimental conditions and captured ~1000 genes per cell; perhaps due to low statistical power, the authors aggregate all *Vgat*+ inhibitory neurons for differential expression testing rather than examine sex differences within individual types. Likewise, van Veen et al. 2020 sequenced only 530 cells from the VMH, then functionally characterized an individual sex-biased gene, *Rprm*. Interestingly, despite sequencing a high number of cells (~150,000), Kim et al. 2019 exclusively discovered sex-specific neuron types in the VMHvl, as opposed to sex differences in gene expression within sex-shared types; however, the authors did not examine how such sex-specific neuron types arise. Collectively, these prior studies also did not profile the chromatin and examine mechanisms by which gonadal hormones specify transcriptomic sex differences.

In this Chapter, I re-analyze a snRNA-seq dataset of the adult BNSTp, consisting of 76,693 neurons from 7 female and 8 male mice (Welch et al., 2019), to characterize sex differences in neuron type abundance and gene expression. I further examine the identity of BNSTp *Esr1*+ neuron types by comparing to GABAergic subclasses in the Allen Institute cortical and hippocampal scRNA-seq atlas (Z. Yao et al., 2021) with MetaNeighbor (Crow et al., 2018). To identify regulatory mechanisms giving rise to sex differences in gene expression, I perform ATAC-seq on BNSTp *Esr1*+ neurons from gonadally intact females and males and integrate these data with ATAC-seq data collected from gonadectomized animals in Chapter 2. Together, these experiments define cellular and molecular sex differences within the BNSTp and demonstrate that male gonadal hormones drive adult sex differences in the chromatin landscape.

4.3 Results

4.3.1 Identification of male-biased inhibitory neuron types in the BNSTp

Across rodents and humans, males have a ~1.5-2x larger BNSTp than females (Allen & Gorski, 1990b; Hines et al., 1992b). In mice, this structural dimorphism arises from male-specific neonatal ERα

activation, which promotes neuron survival (Tsukahara et al., 2011; M. V. Wu et al., 2009b). While BNSTp *Esr1*+ neurons are known to be GABAergic (M. V. Wu & Tollkuhn, 2017b), the identity of male-biased GABAergic neuron types remains unclear. To characterize the identity of male-biased cell populations, and assess whether they comprise sex-shared or sex-specific types, I re-analyzed a snRNA-seq dataset collected from the BNST of adult, gonadally intact females and males (Welch et al., 2019). Seven BNSTp *Esr1*+ transcriptomic neuron types emerged from this analysis, and two of these marked by *Nfix* (i1:Nfix) and *Esr2* (i3:Esr2) are more abundant in males than in females (Fig. 4.1a-b, Fig. 4.2a-b). While a male-bias in BNSTp *Esr2*/ERβ-labeled cells is known (Zuloaga et al., 2014a), Nfix expression has not been previously described in the BNSTp. Immunofluorescent staining confirmed that males have twice the number of ERα+/Nfix+ neurons than females (Fig. 4.1c).

To interpret the functional relevance of BNSTp *Esr1*+ neuron types, I compared their gene expression profiles to the Allen Institute mouse cortical and hippocampal scRNA-seq atlas using MetaNeighbor (Crow et al., 2018; Z. Yao et al., 2021). i1:Nfix neurons uniquely matched the identity of *Lamp5*+ neurogliaform interneurons, particularly the *Lamp5*+/*Lhx6*+ neuron subtype (Fig. 4.1d, Fig. 4.2d-e) (Fishell & Kepecs, 2020; Paul et al., 2017b) and also shared markers (*Moxd1* and *Cplx3*, Fig. 4.2f) with a male-biased neuron type (i20:Gal/Moxd1) in the SDN-POA that was previously found to activate during male-typical mating, inter-male aggression, and parenting behaviors, as determined by MERFISH (Moffitt et al., 2018b). Beyond these two marker genes, i1:Nfix and i20:Gal/Moxd1 neuron types share a transcriptomic identity, consistent with observed Nfix immunofluorescence and *in situ* hybridization signal spanning the BNSTp and SDN-POA (Fig. 4.1e, Fig. 4.2c, e-h). Together, these results define male-biased neurons in the BNSTp and reveal a common *Lamp5*+ neurogliaform identity between the BNSTp and SDN-POA.

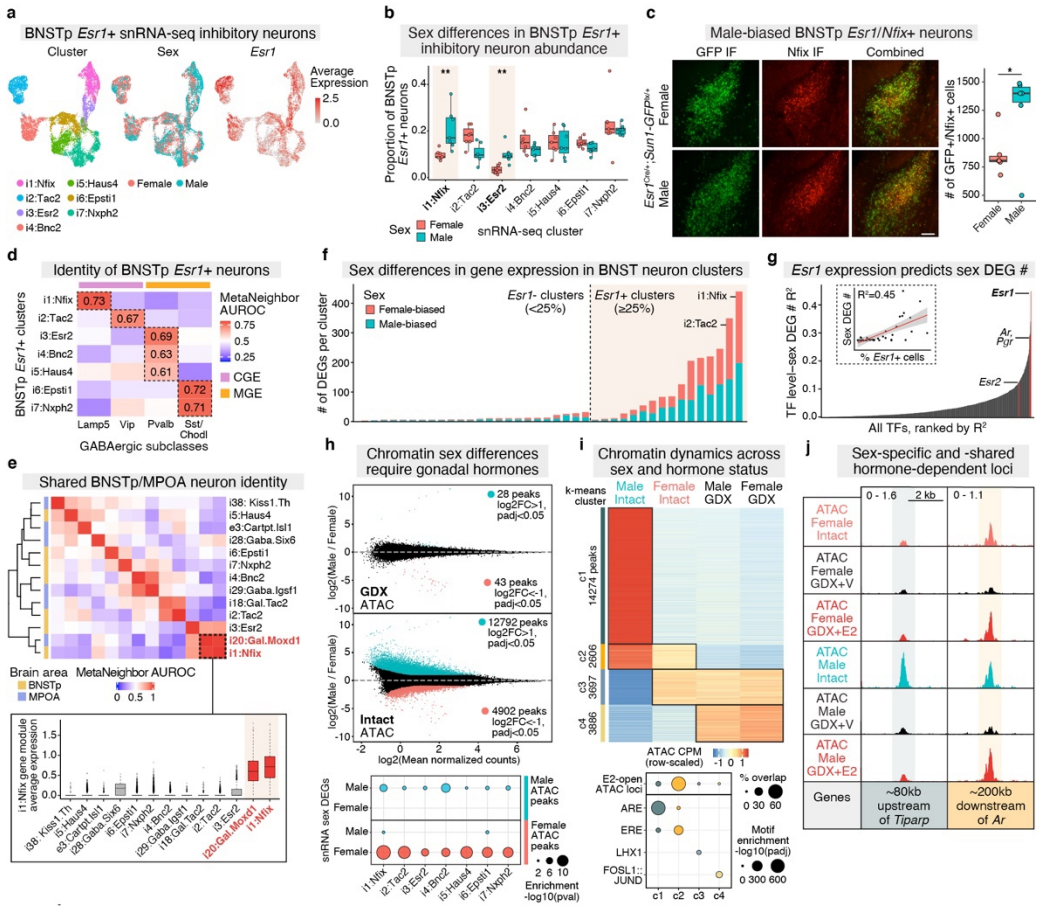

Figure 4.1. Sex differences in cell type abundance and gene regulation in BNSTp *Esr1*+ cells. (**a**) UMAP visualization of BNSTp *Esr1*+ snRNA-seq clusters, colored by identity (left), sex (middle), and *Esr1* expression (right). (**b**) Proportion of BNSTp *Esr1*+ nuclei in each BNSTp *Esr1*+ cluster per sex. Higher proportions of i1:Nfix (padj=0.002) and i3:Esr2 (padj=0.002) neurons are in males than females. Boxplot center=median, box boundaries=1st and 3rd quartile, whiskers=1.5*IQR from boundaries, n=7, **padj<0.01, one-sided, Wilcoxon rank-sum test, adjusted with Benjamini-Hochberg procedure. (**c**) BNSTp IF staining for GFP (left) and Nfix (middle) in P14 female and male *Esr1*$^{Cre/+}$;*Sun1-GFP*$^{lx/+}$ animals (scale=100um). Boxplot center=median, box boundaries=1st and 3rd quartile, whiskers=1.5*IQR from boundaries, n=6, p=0.0422, *p<0.05, two-sided, unpaired *t*-test. (**d**) Heatmap of median MetaNeighbor AUROC values for BNSTp *Esr1*+ clusters and cortical/hippocampal GABAergic neuron subclasses. Colorbar indicates developmental origin of GABAergic subclasses. (**e**) (top) Heatmap of MetaNeighbor AUROC values for BNSTp and MPOA *Esr1* + clusters. (bottom) Average expression of i1:Nfix marker genes across BNSTp and MPOA *Esr1*+ clusters. n=297 i20:Gal.Moxd1 cells, 2459 i1:Nfix cells. Boxplot center=median, box boundaries=1st and 3rd quartile, whiskers=1.5*IQR from boundaries. (**f**) Number of differentially expressed genes (DEGs) between females and males (DESeq2, padj<0.1) per snRNA-seq cluster. (**g**) R^2 between % TF expression and number of sex DEGs per cluster across snRNA-seq clusters. Inset shows correlation for top-ranked TF, *Esr1*. Error band represents 95% confidence interval. (**h**) Differential ATAC sites between (top) gonadectomized, Veh-treated females and males and (middle) gonadally intact females and males. blue dots (edgeR, log2FC>1, padj<0.05), red dots (edgeR, log2FC<-1, padj<0.05). (bottom) Enrichment analysis of sex-biased ATAC peaks at sex DEGs. (**i**) (top) k-means clustering of differentially-accessible ATAC peaks across four conditions (edgeR, padj<0.01). (bottom) Dotplot showing % of sites per cluster overlapping E2-open ATAC loci and motif enrichment analysis of peaks in each cluster (AME). (**j**) Example ATAC peaks in k-means clusters 1 and 2.

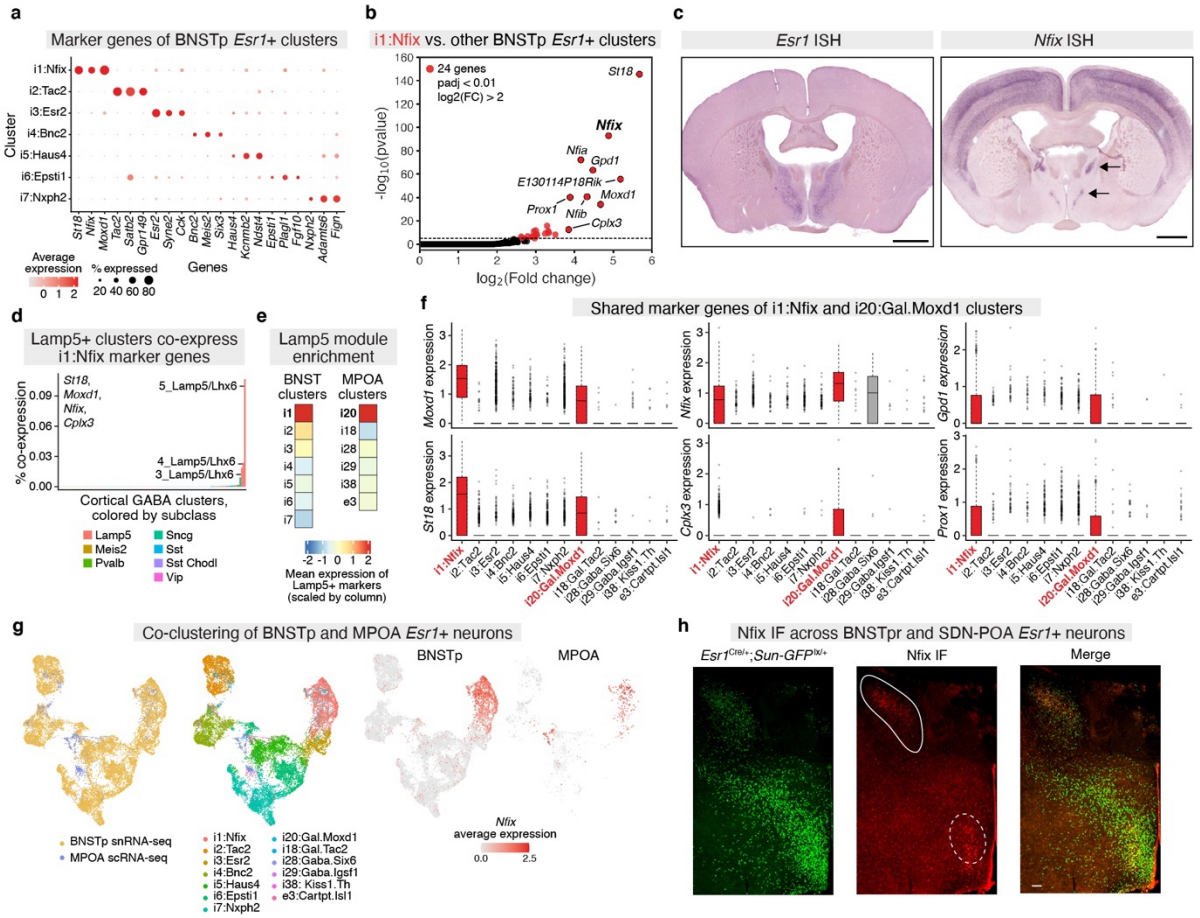

Figure 4.2. Characterization of a shared BNSTp/MPOA transcriptomic cluster. (**a**) Dotplot of top marker genes for each adult BNSTp *Esr1*+ GABAergic cluster (Wilcoxon rank-sum test, padj<0.05). (**b**) Differentially-expressed genes between the i1:Nfix cluster and the other six BNSTp *Esr1*+ inhibitory neuron clusters (DESeq2, log2FC>2, padj<0.01). (**c**) ISH of adult gonadectomized, Veh-treated male (*Esr1*) and adult male (*Nfix*) mouse. Arrows denote *Nfix* ISH staining in BNSTp (dorsal) and POA (ventral). Scale=1mm. (**d**) Co-expression of top i1:Nfix marker genes (*St18*, *Moxd1*, *Nfix*, *Cplx3*) in individual BICCN cortical and hippocampal scRNA-seq GABAergic clusters, colored by subclass. Co-expression defined as % of cells per cluster with non-zero counts for all 4 marker genes. (**e**) Mean expression of *Lamp5*+ subclass marker genes (Wilcoxon rank-sum test, avg_log_FC>0.75, <40% expression in non-*Lamp5*+ neurons, padj<0.05) in BNSTp (left) and MPOA (right) *Esr1*+ clusters, scaled across clusters within each brain region. (**f**) Normalized expression of top marker genes (*Moxd1*, *St18*, *Nfix*, *Cplx3*, *Gpd1*, *Prox1*) shared between i1:Nfix and i20:Gal.Moxd1 (labeled in red). Boxplot center=median, box boundaries=1st and 3rd quartile, whiskers=1.5*IQR from boundaries. n=297 i20:Gal.Moxd1 cells, 2459 i1:Nfix cells. (**g**) UMAP visualization of integrated BNSTp and MPOA *Esr1*+ clusters, demonstrating shared *Nfix* expression across datasets (see also Fig. 2e). (**h**) GFP (left) and Nfix (middle) immunofluorescence staining in an adult male *Esr1*$^{Cre/+}$;*Sun-GFP*$^{lx/+}$ mouse. Solid white circle indicates BNSTp; dotted white circle indicates SDN-POA. Scale=100um.

4.3.2 Sex differences in gene expression across BNSTp neuron types

Given that snRNA-seq was performed on both sexes with independent, biological replicates, I next tested for sex differences in gene expression and found extensive and robust (FDR<0.1) sex-biased expression across BNST neuron types (Fig. 4.1f, Fig. 4.3a). Hierarchical clustering of sex-biased genes across these types revealed a striking degree of heterogeneity (Fig. 4.3b): most sex differences were specific to individual types (e.g., *Dlg2*/PSD-93 and *Kctd16* in i1:Nfix neurons), whereas certain differences were detected across multiple populations (e.g., *Tiparp*, *Socs2*) (Fig. 4.3c). In line with prior circuit-mapping experiments, neuron types annotated to the BNSTp contained a higher number of sex-biased genes per type than those annotated to the anterior BNST (BNSTa) (Fig. 4.3e). To predict which TFs may regulate sex-biased gene expression in the BNSTp, I performed a linear regression analysis between % cells per type expressing a given TF of interest (for each mouse TF) and the number of sex-biased genes per type. Relative to all other TFs in the mouse genome, *Esr1*, along with co-expressed gonadal hormone receptors, *Ar* and *Pgr*, correlated best with the degree of sex-biased genes (Fig. 4.1g, Fig. 4.3f). This finding reveals that gonadal hormone receptors define neuron types with prominent sex-biased gene expression. Of note, *Esr2* did not correlate well with sex-biased expression, as it is largely restricted to a single BNSTp type (i3:Esr2), indicating it is not a key regulator of sex differences within this region – a concept I return to in Chapters 5 and 6.

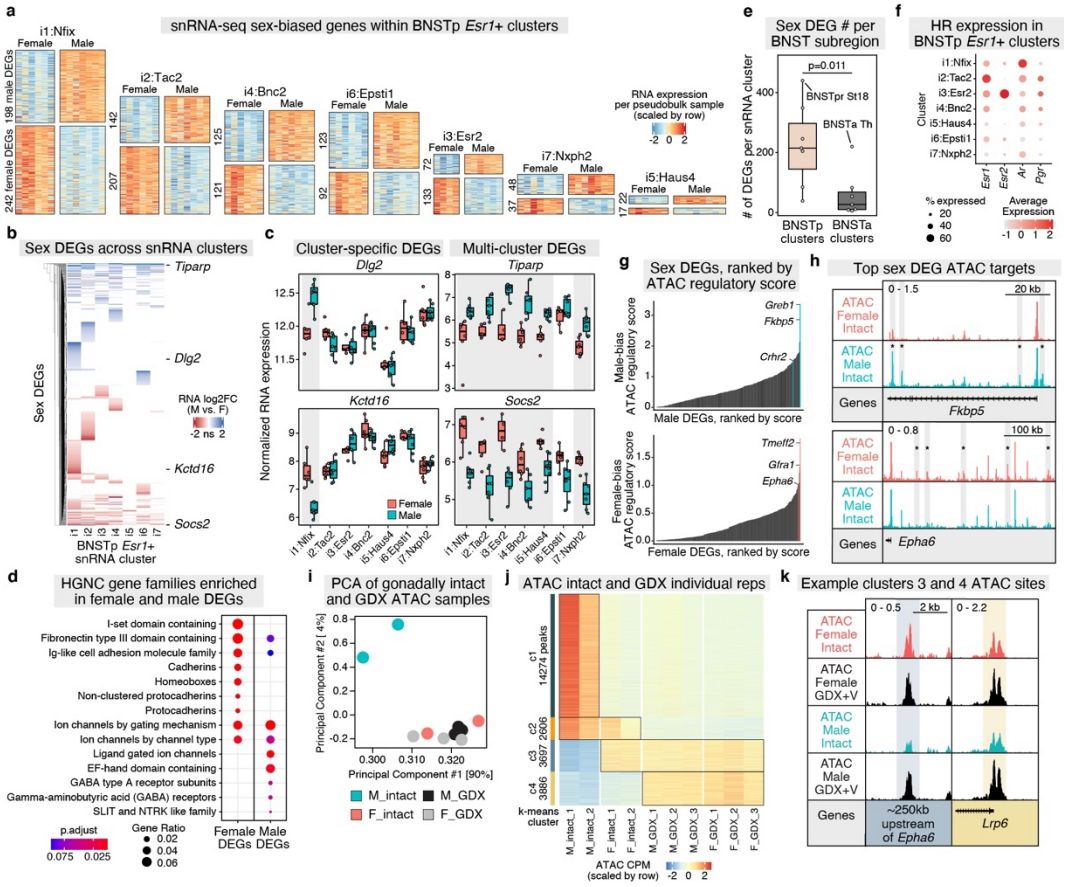

Figure 4.3. Additional analysis of BNSTp sex DEGs and gonadally intact ATAC-seq. (**a**) Normalized, pseudo-bulk expression of sex DEGs identified within each BNSTp *Esr1*+ cluster (DESeq2, padj<0.1). Each heatmap column corresponds to a pseudo-bulk sample (gene counts aggregated across cells in sample). (**b**) Hierarchical clustering of log2FC values for sex DEGs called as significant in at least one BNSTp *Esr1*+ cluster. Sex DEGs with non-significant differential expression colored in white. (**c**) Example sex DEGs with significant differential expression in a single *Esr1*+ cluster (*Dlg2*, *Kctd16*) and in multiple *Esr1*+ clusters (*Tiparp*, *Socs2*). Boxplot center=median, box boundaries=1st and 3rd quartile, whiskers=1.5*IQR from boundaries. n=4-7 female pseudo-bulk replicates, 6-8 male pseudo-bulk replicates. (**d**) Top HGNC gene families (clusterProfiler, padj<0.1) enriched within female-biased DEGs (left) and male-biased DEGs (right) relative to non-differential, expressed genes. (**e**) Number of sex DEGs per cluster in *Esr1*+ clusters annotated to the BNST posterior (BNSTp, n=7 clusters) and anterior (BNSTa, n=7 clusters) subregions. Boxplot center=median, box boundaries=1st and 3rd quartile, whiskers=1.5*IQR from boundaries. p-value from two-sided, Wilcoxon rank-sum test. (**f**) Dotplot of sex hormone receptor (HR) expression across BNSTp *Esr1*+ clusters. (**g**) Barplots of (top) male-biased DEGs ranked by male-biased ATAC peak regulatory potential score and (bottom) female-biased DEGs ranked by female-biased ATAC peak regulatory potential score. Higher score indicates higher density of sex-biased ATAC peaks around the TSS of sex DEGs. (**h**) Example sex DEGs (*Fkbp5*, *Epha6*) with high density of sex-biased ATAC peaks. *sex-biased ATAC peak. (**i**) PCA of gonadally intact and gonadectomized (GDX), Veh-treated ATAC CPM values within the consensus peak matrix. (**j**) Heatmap of ATAC CPM for gonadally intact (n=2 per condition) and GDX, Veh-treated ATAC samples (n=3 per condition) at differential peaks (edgeR, glmQLFTest, padj<0.01), grouped by k-means clustering. (**k**) Example differential ATAC peaks in k-means clusters c3 (left) and c4 (right).

4.3.3 Male gonadal hormones control adult sex differences in gene regulation

To identify chromatin regions and TFs controlling sex differences in BNSTp gene expression, I performed ATAC-seq on BNSTp *Esr1*+ cells harvested from adult, gonadally intact *Esr1*$^{Cre/+}$; *Sun1-GFP*$^{lx/+}$ mice. ~18,000 regions differed in accessibility between females and males (Fig. 4.1h). Importantly, these regions were statistically enriched at sex-biased genes detected across *Esr1*+ neuron types (Fig. 4.1h), suggesting they act as putative enhancers of these genes. Ranking sex-biased genes by their regulatory potential score, a metric that takes into account the number and distance of sex-biased peaks to the gene transcription start site (TSS), highlighted genes with multiple associated peaks, such as *Epha6* (female-biased, with 5 female-biased peaks) and *Fkbp5* (male-biased, with 4 male-biased peaks) (Fig. 4.3h).

In contrast, adult gonadectomy reduced the number of sex-biased regions to 71 (Fig. 4.1h), which were mostly located on sex chromosomes. To characterize patterns of chromatin accessibility across sexes and gonadal hormone status, I performed k-means clustering on ATAC peaks that differed significantly across the 4 groups (female GDX, female intact, male GDX, male intact). This approach revealed male-specific, but not female-specific, responses to gonadectomy (Fig. 4.1i, 4.3j) – namely, k-means cluster 1, which closes in males upon GDX, and cluster 3, which opens in males upon GDX. Clusters 2 and 4 changed across both sexes in response to GDX, either closing (cluster 2), such as at the *Ar* locus (Fig. 4.1j), or opening (cluster 4), respectively. Moreover, principal component analysis (PCA) of the 4 experimental groups revealed the chromatin accessibility of males undergoes global feminization upon GDX (Fig. 4.3i); one example of this is an aforementioned peak at the *Epha6* locus (Fig. 4.3k). Notably, cluster 1 regions primarily contained the ARE, while cluster 2 regions were enriched for the ERE and strongly overlapped E2-open regions defined in Chapter 2 (Fig. 4.1i). Therefore, in the BNSTp, estradiol maintains chromatin in an active state across both sexes, whereas testosterone, produced from the male testes, facilitates both chromatin activation and repression. Interestingly, testosterone-dependent repression does not involve direct AR binding to the genome, as

the GDX-opened regions did not contain the ARE. Rather, a male-biased AR target gene may control genomic repression. Collectively, these data indicate that gonadal hormone receptors activate sex differences in gene expression in the adult brain, largely as a consequence of acute hormonal state.

4.4 Discussion

In this Chapter, I describe cellular and gene regulatory sex differences in the adult BNSTp. Males contain ~1.5-2x more neurons in the BNSTp than females; however, the identity of these neurons, and whether they represent sex-shared or sex-specific types, was unclear. Here I find that the male-bias in BNSTp cell number is attributed to two male-biased inhibitory neuron types defined by the marker genes, *Nfix* (i1:Nfix) and *Esr2* (i3:Esr2) (Fig. 4.1b-c). Comparison to GABAergic subclasses from the Allen Institute cortical and hippocampal scRNA-seq atlas revealed that these two types resemble *Lamp5*+ and *Pv*+ neurons, respectively (Fig. 4.1d, 4.2d-e). Consistent with this prediction, ERβ has been shown to co-localize with PV throughout other regions of the brain (Blurton-Jones & Tuszynski, 2002). In layer V of the somatosensory cortex, estrogens modulate the excitability of ERβ+/PV+ inhibitory neurons throughout the estrous cycle (Clemens et al., 2019). Given this similarity in gene co-expression across brain regions, it is possible that these neurons share a developmental origin, with *Esr2* and *Pv* regulated by a common upstream identity regulator TF.

i1:Nfix neurons resemble the *Lamp5*+ GABAergic subclass, which is composed of neurogliaform and Ivy cells (Z. Yao et al., 2021). Further analysis revealed that the top marker genes of the i1:Nfix type (*Cplx3*, *Moxd1*, *Nfix*, *St18*) are co-expressed exclusively within *Lamp5*+/*Lhx6*+ neurons (Fig. 4.2d). Similar to i1:Nfix neurons, the *Lamp5*+/*Lhx6*+ type co-expresses TFs that distinguish MGE (*Lhx6*) from CGE (*Nfix*, *Prox1*) lineages (Z. Yao et al., 2021); hence, the developmental origin of this type, and its relationship to prior genetic targeting strategies, remains unclear (Z. J. Huang & Paul, 2019). In cortical circuits, neurogliaform interneurons provide regional inhibition through synaptic and ambient release of GABA (Overstreet-Wadiche & McBain, 2015). As

all BNSTp neurons are GABAergic, it is possible that elevated numbers of i1:Nfix inhibitory neurons in males generates stronger disinhibition of downstream projection sites, such as the VMHvl and MeA, resulting in increased responses to olfactory information and male-typical levels of mounting or attacking (Karigo et al., 2020). Of note, the i1:Nfix type matches a neuron type in the SDN-POA (Fig. 4.1e, 4.2f-h) that was previously found to be active during male-typical mating, aggression, and parenting behavior (Moffitt et al., 2018a). A shared identity between the BNSTp and SDN-POA may indicate a common developmental origin and raises the intriguing possibility that this population may have evolved to regulate a highly-conserved repertoire of innate, reproductive behaviors. In the future, it will be critical to develop genetic strategies to selectively label these neurons for circuit-tracing, electrophysiology, and fate-mapping experiments.

Consistent with a prior study utilizing cDNA microarrays (X. Xu et al., 2012), I find that male gonadal hormones, rather than female gonadal hormones, largely control sex differences in gene regulation in the adult BNSTp (Fig. 4.1i, 4.3i). ATAC-seq, followed by motif analysis, further revealed that male hormones, presumably testosterone, both activate and repress chromatin, with activation driven by direct binding of AR to the genome and repression not involving recruitment of a gonadal hormone receptor (Fig. 4.1i). How does testosterone facilitate genomic repression? One possible mechanism is that AR drives the male-biased expression of a transcriptional repressor, or an upstream molecule that signals to a transcriptional repressor, consequently resulting in male-specific chromatin closing that is reversed upon male GDX. Moreover, it is possible, if not likely, that repressive mechanisms vary across *Esr1*+ neuron types, given that they represent distinct populations with varying numbers of female-biased genes (Fig. 4.1f, 4.3a-c). To address this question, I have recently performed single-nucleus ATAC-sequencing (snATAC-seq) on ~40,000 adult female and male BNST inhibitory neurons ($Vgat^{Cre/+}$; $Sun1\text{-}GFP^{lx/+}$) and am currently analyzing this dataset.

The lack of a genomic response to female GDX raises the question: Do hormones released throughout estrous influence gene expression and chromatin state in the BNSTp? Female mice spend

longer in diestrus (low estradiol and progesterone) than in proestrus (high estradiol and progesterone); hence, it is possible that the females used in the snRNA-seq and ATAC-seq experiments were mostly in a low-hormone state. However, given that these datasets were collected by independent laboratories from 15 female mice (7 biological replicates with 1 animal per replicate for snRNA-seq, and 2 biological replicates with 4 animals per replicate for ATAC-seq), it seems unlikely that no animals were in proestrus at the time of collection. Another explanation is that estradiol and progesterone may transiently regulate gene expression on a timescale of minutes or hours, making their response difficult to detect. In line with this explanation, I performed one preliminary ATAC-seq experiment on BNSTp *Esr1*+ neurons from diestrus and proestrus females staged using vaginal smears, but did not detect any significant differences (data not shown). It is possible that applying a more sensitive approach to staging animals (*i.e.*, measuring serum hormone levels) and accounting for inter-individual variability by using 1 female per biological replicate may reveal modulation of gene regulation throughout the cycle, as observed for *Mc4r* in the VMHvl (Chapter 3).

Collectively, these data reveal that adult sex differences in gene regulation are largely driven by the activational effects of gonadal hormones. In the next Chapter, I examine the second component of the "Organizational-Activational" hypothesis: how do neonatal hormones organize sex differences in the brain?

4.5 Materials and Methods

Animals. All animals were maintained on a 12:12 light cycle and provided food and water *ad libitum*. All mouse experiments were performed under strict guidelines set forth by the CSHL Institutional Animal Care and Use Committee (IACUC). *Esr1*^{Cre} (H. Lee et al., 2014), *ROSA26*$^{CAG\text{-}Sun1\text{-}sfGFP\text{-}Myc}$ (Mo et al., 2015) (abbreviated as *Sun1-GFP*), and C57Bl6/J wildtype mice were obtained from Jackson labs. Adult male and female mice were used between 8-12 weeks of age.

Adult ATAC-seq nuclei isolation. GFP+ nuclei were isolated from adult *Esr1*$^{Cre/+}$; *Sun1-GFP*$^{lx/+}$ mice (4 pooled per condition), as previously described in Chapter 2 Materials and Methods.

ATAC-seq library preparation. ATAC-seq libraries were prepared, as previously described in Chapter 2 Materials and Methods.

Nfix immunofluorescence staining. Brains were dissected from perfused P14 *Esr1*$^{Cre/+}$; *Sun1-GFP*$^{lx/+}$ animals and cryosectioned at 40µm prior to immunostaining with primary antibodies against GFP (1:1000, Aves GFP-1020) and Nfix (1:1000, Thermo Fisher PA5-30897), and secondary antibodies against chicken (1:300, Jackson Immuno 703-545-155) and rabbit (1:800, Jackson Immuno 711-165-152), as previously described (M. V. Wu & Tollkuhn, 2017b). 20x widefield image stacks spanning the BNSTp (5 sections, both sides) were taken on a Zeiss Axioimager M2 System equipped with MBF Neurolucida Software. The number of Nfix+, GFP+, and Nfix+/GFP+ cells was quantified using Fiji/ImageJ from the center 3 optical slices by an investigator blinded to condition.

ATAC-seq data processing. ATAC-seq data were processed using the ENCODE ATAC-seq pipeline (https://github.com/ENCODE-DCC/atac-seq-pipeline) with default parameters, as previously described in Chapter 2 Materials and Methods.

Adult gonadally intact ATAC-seq analysis. ATAC-seq differential peak calling and comparison between gonadally intact (abbreviated as intact) and GDX ATAC samples were performed with DiffBind and edgeR. A DiffBind dba object was created from individual replicate BAM and MACS2 narrowPeak files for the four groups: female intact (n=2), male intact (n=2), female GDX Veh-treated (n=3), male GDX Veh-treated (n=3). A count matrix was created with dba.count, and consensus peaks were re-centered to ±250 bp around the point of highest read density (summits=250). The consensus

peakset count matrix was subsequently used as input to edgeR. Differential peaks (abs(log2FC)>1, padj<0.05) were calculated between female intact and male intact and between female GDX Veh-treated and male GDX Veh-treated groups using glmQLFTest. BETA was used to assess statistical association between gonadally intact, sex-biased ATAC peaks and sex DEGs called in BNSTp *Esr1*+ snRNA-seq clusters (top 500 genes per cluster, ranked by padj). Sex DEGs ranked by ATAC regulatory potential score (S. Wang et al., 2013), a metric that reflects the number of sex-biased peaks and distance of sex-biased peaks to the TSS, are shown in Fig. 4.3g. HGNC gene family enrichment was performed on sex DEGs, using a background of expressed genes in any of the 7 BNSTp *Esr1*+ clusters.

To identify differential peaks across the four conditions, an ANOVA-like design was created in edgeR by specifying multiple coefficients in glmQLFTest (coef=2:4). A matrix of normalized counts within these differential peaks (padj<0.01) was clustered using k-means clustering (kmeans function in R), with k=4 and iter.max=50. For each k-means cluster, the cluster centroid was computed, and outlier peaks within each cluster were excluded on the basis of having low Pearson's correlation with the cluster centroid (R<0.8). Depth-normalized ATAC CPM values within these peak clusters are shown in Fig. 4.1i (mean across biological replicates per group) and Fig. 4.3j (individual biological replicates). Peak cluster overlap with E2-open ATAC loci (abs(log2FC)>0, padj<0.05) was computed with bedtools intersect (-wa). For each peak cluster, motif enrichment analysis was performed by first generating a background peak list (matching in GC-content and accessibility) from the consensus ATAC peak matrix using chromVAR (addGCBias, getBackgroundPeaks) (Schep et al., 2017), then calculating enrichment with AME using the background peak list as the control (--control background peaks). In Fig. 4.1i, the JASPAR 2020 AR motif (MA0007.3) is labeled as 'ARE', and the ESR2 motif (MA0258.2) is labeled as 'ERE'.

Adult single-nucleus and -cell RNA-seq analysis. Mouse BNST snRNA-seq data containing 76,693 neurons across 7 adult female and 8 adult male biological replicates (Welch et al., 2019) were accessed

from GEO:GSE126836 and loaded into a Seurat object (Stuart et al., 2019). Mouse MPOA scRNA-seq data containing 31,299 cells across 3 adult female and 3 adult male biological replicates (Moffitt et al., 2018b) were accessed from GEO:GSE113576 and loaded into a Seurat object. Cluster identity, replicate, and sex were added as metadata features to each Seurat object. Pseudo-bulk RNA-seq analysis was performed to identify sex differences in gene expression in the BNST snRNA-seq dataset. Briefly, the Seurat object was converted to a SingleCellExperiment object (as.SingleCellExperiment). Genes were filtered by expression (genes with >1 count in ≥5 nuclei). NCBI predicted genes were removed. For each cluster, nuclei annotated to the cluster were subsetted from the main Seurat object. Biological replicates containing ≤20 nuclei in the subsetted cluster were excluded. Gene counts were summed for each biological replicate within each cluster. Differential gene expression analysis across sex within each cluster was performed on the filtered, aggregated count matrix using DESeq2 (design = ~ sex) with alpha = 0.1. The BNSTp_Cplx3 cluster was excluded, as none of the replicates in this cluster contained over 20 nuclei. Clusters containing ≥25% nuclei with ≥1 *Esr1* counts in the main Seurat object were classified as *Esr1*+ (i1:Nfix, i2:Tac2, i3:Esr2, i4:Bnc2, i5:Haus4, i6:Epsti1, i7:Nxph2, i8:Zeb2, i9:Th, i10:Synpo2, i11:C1ql3, i12:Esr1, i13:Avp, i14:Gli3). To identify TFs that correlate with sex DEG number per cluster (Fig. 4.1g), a linear regression model with % TF expression as the predictor variable and sex DEG number per cluster as the response variable was generated using the lm function in R stats (formula = % TF expression ~ DEG number). This model was tested for all TFs in the SCENIC (Aibar et al., 2017) mm10 database. All TFs were then ranked by R^2 to identify those most predictive of sex DEG number, and the ranked R^2 values are shown in Fig. 4.1g.

To visualize BNSTp *Esr1*+ snRNA-seq data (Fig. 4.1a), BNSTp *Esr1*+ clusters were subsetted from the main Seurat object. Gene counts were normalized and log-transformed ('LogNormalize'), and the top 2000 variable features were identified using FindVariableFeatures (selection.method='vst'). Gene counts were scaled, and linear dimensionality reduction was performed by principal component analysis (runPCA, npcs=10). BNSTp *Esr1*+ clusters were visualized with UMAP (runUMAP, dims =

10). To generate the heatmaps in Fig. 4.2a, pseudo-bulk counts for each biological replicate included in the analysis were normalized and transformed with variance stabilizing transformation (DESeq2 'vst'), subsetted for sex-biased genes in each cluster, and z-scaled across pseudo-bulk replicates.

To examine differential abundance of BNSTp *Esr1*+ clusters between sexes (Fig. 4.1b), the proportion of total nuclei in each BNSTp *Esr1*+ cluster was calculated for each biological replicate. After calculating the nuclei proportions, sample 'MALE6' was excluded as an outlier for having no detection (0 nuclei) of i1:Nfix and i2:Tac2 clusters and over-representation of the i5:Haus4 cluster. The one-sided Wilcoxon rank-sum test (wilcox.test in R stats) was used to test for male-biased nuclei abundance across biological replicates in each cluster. P-values were adjusted for multiple hypothesis testing using the Benjamini-Hochberg procedure (method='fdr').

To identify marker genes enriched in the i1:Nfix cluster relative to the remaining 6 BNSTp *Esr1*+ clusters (Fig. 4.2b), differential gene expression analysis was performed using DESeq2 with design = ~ cluster_id (betaPrior=TRUE), alpha=0.01, lfcThreshold=2, altHypothesis = "greater".

To identify the enrichment of *Lamp5*+ subclass markers in BNSTp and MPOA *Esr1*+ clusters (Fig. 4.2e), a Seurat object was created from the Allen Brain Atlas Cell Types dataset. Gene counts per cell were normalized and log-transformed ('LogNormalize'), and subclass-level marker genes were calculated with the Wilcoxon rank-sum test (FindAllMarkers, test.use='wilcox', min.diff.pct=0.2). The mean expression of *Lamp5*+ subclass markers (avg_logFC>0.75, padj<0.05, <40% in non-*Lamp5*+ subclasses) was calculated in BNSTp and MPOA *Esr1*+ clusters and visualized using pheatmap.

To generate the UMAP plots shown in Fig 4.1g, BNSTp *Esr1*+ clusters were integrated with MPOA/BNST *Esr1*-expressing clusters (e3: Cartpt_Isl1, i18: Gal_Tac2, i20: Gal_Moxd1, i28: Gaba_Six6, i29: Gaba_Igsf1, i38: Kiss1_Th) using Seurat. 'Anchors' were identified between cells from the two datasets, using FindIntegrationAnchors. An integrated expression matrix was generated using IntegrateData (dims=1:10). The resulting integrated matrix was used for downstream PCA and UMAP visualization (dims=1:10).

MetaNeighbor analysis. MetaNeighbor (Crow et al., 2018) was used to quantify the degree of similarity between BNSTp *Esr1*+ clusters and MPOA *Esr1*+ clusters and between BNSTp *Esr1*+ clusters and cortical/hippocampal GABAergic neuron subclasses from the Allen Brain Atlas Cell Types database (Z. Yao et al., 2020). Briefly, the BNST and MPOA Seurat objects were subsetted for *Esr1*+ clusters, then transformed and merged into one SingleCellExperiment object. For the BNSTp and cortex comparison, BNSTp *Esr1*+ clusters were merged into a SingleCellExperiment with cortical/hippocampal GABAergic cortical clusters. Unsupervised MetaNeighbor analysis was performed between BNST and MPOA clusters, and between BNST and cortical/hippocampal clusters, using highly variable genes identified across datasets (called with 'variableGenes' function). The median AUROC value per cortical/hippocampal GABAergic subclass across Allen Brain Atlas datasets for each BNSTp *Esr1*+ cluster is shown in Fig. 4.1d.

Chapter 5

Organization of sex differences in gene regulation in the BNST

Sections of this chapter were previously uploaded to bioRxiv (2020) under the title "Regulation of neural gene expression by estrogen receptor alpha" by Bruno Gegenhuber, Melody V. Wu, Robert Bronstein, and Jessica Tollkuhn. This manuscript is now in press at Nature (2022) under the title "Gene regulation by gonadal hormone receptors defines neuronal sex differences" by Bruno Gegenhuber, Melody V. Wu, Robert Bronstein, and Jessica Tollkuhn.

5.1 Abstract

In mice, a neonatal surge of testosterone permanently masculinizes the brain via conversion to estradiol and activation of ERα. A central hypothesis in the field of neuroendocrinology states that sex differences in the brain arise from an epigenetic program driven by the male-specific neonatal hormone surge (McCarthy, Auger, et al., 2009; McCarthy et al., 2017). However, the nature of this program remains unclear, given that the genomic targets of ERα, and their longevity across brain development, have not been identified. Here I utilize low-input and single-cell transcriptomic and chromatin profiling approaches to discover the genomic response to the neonatal hormone surge and define its longevity in the BNSTp. I find that the neonatal surge rapidly recruits ERα to the genome to drive a male-typical gene regulatory program concomitant with the onset of male-biased BNSTp cell survival and axonogenesis. A subset of these ERα target genes persists as sex-biased throughout the neonatal critical period, including those involved in axon outgrowth (*Nell2*, *Pak7*/PAK5), synapse formation (*Col25a1*, *Il1rap*), and transcriptional regulation (*Greb1*, *Plagl1*, *Sox5*). Selective deletion of ERα within inhibitory neurons both feminized the expression of these genes as well as the abundance of previously

identified male-biased neuron types (i1:Nfix and i3:Esr2). Together, these experiments reveal two mechanisms by which ERα organizes brain sexual differentiation: 1) inducing a persistent gene regulatory program and 2) promoting the abundance of two BNSTp inhibitory neuron types. The genes and enhancers identified here provide the first step toward defining molecular mechanisms underlying sex differences in the connectivity and survival of BNSTp neuron types.

5.2 Introduction

In mammals, gonadal steroid hormones regulate sex differences in neural activity and behavior. These hormones establish sex-typical neural circuitry during critical periods of development and activate the display of innate social behaviors in adulthood. Among these hormones, estradiol is the principal regulator of mouse brain sexual differentiation. In males, the testes briefly activate at birth, generating a sharp rise in testosterone that subsides within hours (Clarkson & Herbison, 2016). Neural aromatase converts circulating testosterone to 17β-estradiol, which acts through ERα in discrete neuronal populations in the brain to specify sex differences in cell number and connectivity (MacLusky & Naftolin, 1981; McCarthy, 2008b; Simerly, 2002b).

Prior studies have proposed that sex differences in the brain arise from permanent epigenomic modifications imparted by the neonatal hormone surge, constituting an epigenetic process (McCarthy, Auger, et al., 2009; McCarthy et al., 2017). While these studies implicated DNA methylation and histone acetylation in the development of brain sex differences using non-specific, pharmacological inhibitors and/or constitutive genetic deletion approaches, they did not identify genomic loci with persistent sex differences in epigenomic modifications arising from the neonatal surge (Auger et al., 2000; Matsuda et al., 2011; Murray et al., 2009; Nugent et al., 2015). Therefore, it remains unclear whether sex differences in chromatin both exist and persist in the developing brain. Likewise, aside from *Ar* (Juntti et al., 2010), the identity and function of sex-biased genes targeted by neonatal ERα activation remain unclear.

Until recently, studying the chromatin of small cell populations, such as genetically defined neuron types in the developing brain, has been limited by technology. With the advent of CUT&RUN (discussed in Chapter 2) and CUT&Tag (Kaya-Okur et al., 2019; Skene & Henikoff, 2017), it is now possible to profile TF binding and histone modifications in small cell populations or at single-cell resolution. TF binding and histone modifications associated with active enhancers, histone H3 lysine 4 monomethylation (H3K4me1) and lysine 27 acetylation (H3K27ac), modify the accessibility of chromatin by driving nucleosome turnover and/or decompacting chromatin fibers (Klemm et al., 2019). Hence measuring chromatin accessibility with ATAC-seq (as done in Chapter 2) captures the activation status of enhancers genome-wide without *a priori* knowledge of the specific epigenomic modifications involved. In addition, recent advancements in single-cell sequencing have led to the simultaneous detection of chromatin accessibility and RNA from the same cell at high-throughput (Ma et al., 2020; C. Zhu et al., 2021), enabling accurate reconstruction of gene regulatory networks. In the context of brain sexual differentiation, such advancements offer the ability to link ERα binding to sex differences in gene expression within specific neuron types and predict how such genes give rise to sexual differentiation of neural circuits.

In this Chapter, I use ATAC-seq to investigate sex differences in chromatin state in BNSTp *Esr1*+ cells four days after the neonatal surge. To determine the extent to which these differences depend on the neonatal surge, I additionally profile the chromatin of females treated at birth with estradiol. These experiments revealed that the neonatal surge drives extensive and sustained sex difference in chromatin accessibility. I further demonstrate that male-biased, and neonatal estradiol-dependent, chromatin loci are primarily driven by neonatal ERα recruitment by performing CUT&RUN on females treated acutely with vehicle or estradiol at birth. To link these sex-biased genomic regions to gene expression, I perform single-nucleus multiome sequencing (RNA & ATAC) on females and males at P4 and P14. I identify transcriptomic sex differences linked to sex-biased enhancers within individual BNSTp neuron types. Furthermore, a subset of sex-biased genes and enhancers remained

sex-biased throughout the neonatal critical period, indicative of a *bona fide* epigenetic program of brain sexual differentiation. Many of these genes regulate key neurodevelopmental processes, such as axon outgrowth, synapse formation, and neurotransmission. Future investigation will reveal how these genes coordinate sex differences in BNSTp circuitry.

5.3 Results

5.3.1 ERα drives sex differences in chromatin state following birth

Sexual dimorphism in BNSTp neuron survival and wiring emerges throughout a two-week window following birth, well after neural estradiol has subsided in males. To determine the genomic targets of the neonatal surge, I performed ATAC-seq on BNSTp *Esr1*+ cells harvested from females and males treated on the day of birth with vehicle (NV), and females treated with E2 (NE), at postnatal day 4 (P4), which corresponds to the onset of male-biased BNSTp cell survival and axonogenesis (Cooke & Simerly, 2005; Gotsiridze et al., 2007). I detected ~2000 sex-biased chromatin regions at this time, the majority of which were male-biased (Fig. 5.1a). Hierarchical clustering of differential ATAC peaks across the three groups (NV female, NV male, NE female) demonstrated, in an unbiased fashion, that nearly all sex differences depended on neonatal estradiol (clusters 3 and 5, Fig. 5.2a). The 54 sex-biased regions uninfluenced by estradiol treatment, corresponding to clusters 1 and 4, were primarily on sex chromosomes. The majority (~85%) of the neonatal estradiol (NE)-open regions were also induced by estradiol in the adult E2 ATAC-seq dataset (Fig. 5.2c-d), suggesting that developmental state largely does not restrict the genomic response to estradiol. Likewise, the overwhelming majority of NE-open regions contained the ERE (Fig. 5.2b). To determine whether activation of ERα by the neonatal surge drives male-typical chromatin opening, I performed ERα CUT&RUN on FACS-isolated *Esr1*+ cells from female *Esr1*$^{Cre/+}$; *Sun1-GFP*$^{lx/+}$ pups treated acutely with vehicle or estradiol on the day of birth. Estradiol rapidly recruited ERα to NE-open regions (Fig. 5.1a, Fig. 5.2e-h). Interestingly, while neonatal estradiol also led to chromatin repression, ERα did not bind these sites (Fig. 5.1a, Fig. 5.2h),

indicating a downstream target of ERα may close chromatin, similar to how a downstream target of AR closes chromatin in adulthood. Together, these results demonstrate neonatal ERα activation drives early-life sex differences in the chromatin landscape.

Collecting chromatin accessibility data at different life stages (P4, adulthood) allowed me to test whether chromatin regions regulated by the neonatal surge maintain sex-biased activation status into adulthood. A small, but detectable, number of NE-regulated regions (152 regions) were maintained as sex-biased in gonadally intact adults (Fig. 5.3a). Of note, while most NE-open loci did not maintain male-biased accessibility after puberty, they still localized at adult male-biased genes and clustered around adult male-biased ATAC peaks (Fig. 5.3b-d). Therefore, certain male-biased genes undergo sequential regulation by ERα and AR in early life and adulthood, respectively, such as *Prlr*, *Cckar*, *Pdzrn4*, and *Tiparp*. Moreover, the close proximity of NE-open and adult male-biased sites along the linear genome suggests early-life organization of 3D chromatin structure may influence the response to gonadal hormones released at puberty.

5.3.2 ERα organizes a sustained gene expression program during brain sexual differentiation

The identification of ~2000 chromatin regions controlled by the neonatal surge suggests ERα drives extensive sex differences in the expression of genes that control brain sexual differentiation. To identify these genes, and assess the longevity of their expression, I optimized and performed single-nucleus multiome (RNA & ATAC) sequencing on female and male BNST *Esr1*+ cells collected at P4 and P14, after the closure of the neonatal critical period (McCarthy, 2008b). I profiled 14,836 cells across groups (Fig. 5.1b, Fig. 5.4a-b). By comparing *de novo* clustering of the neonatal BNST to the previously published adult snRNA-seq dataset (Fig. 5.1b, Fig. 5.4c-e), I found that *Esr1*+ neuron identity is largely

the same across P4, P14, and adulthood, consistent with recent neurodevelopmental studies examining the timing of cortical neuron fate specification (Allaway et al., 2021; Di Bella et al., 2021).

Differential gene expression analysis between females and males across *Esr1*+ neuron types on P4 revealed >400 sex-biased genes (Fig. 5.1c). Similar to the adult BNST, these genes are heterogeneously sex-biased across neuron types (Fig. 5.5a), with the majority of sex differences arising in 5 populations (i1:Nfix, i2:Tac2, i3:Esr2, i4:Bnc2, i12:Esr1). Performing bulk RNA-seq on BNSTp *Esr1*+ cells harvested from females treated at birth with vehicle or estradiol validated that many of these sex-biased genes are regulated by neonatal estradiol (Fig. 5.5b). Notably, estradiol-dependent, sex differences in gene expression and chromatin state occurred in neuron types lacking *Cyp19a1*/aromatase expression (Fig. 5.1c-e), indicative of non-cell-autonomous estradiol signaling.

To link the chromatin and gene expression data, I constructed a gene regulatory map across *Esr1*+ neuron types consisting of sex-biased genes and NE-regulated ATAC peaks with correlated accessibility (Fig. 5.1c, Fig. 5.5e-h). This map demonstrates both divergent responses across neuron types, such as *Htr4* and *Csgalnact1* upregulation and enhancer activation in i1:Nfix and i3:Esr2 cells (Fig. 5.5f), respectively, as well as neuron type-specific enhancers for common sex-biased targets, such as *Arid1b* (Fig. 5.5g), a known autism spectrum disorder (ASD) candidate. Further examination showed that ~40% of high-confidence (FWER≤0.05) ASD candidate genes (Satterstrom et al., 2020), including *Grin2b, Scn2a1*, and *Slc6a1*, contained NE-open chromatin regions and ERα occupancy (Fig. 5.2j).

Lastly, I examined whether sex-biased genes, and their corresponding sex-biased enhancers, are sustained across the neonatal critical period by comparing *Esr1*+ neurons between P4 and P14. While the total number of sex-biased genes declined between P4 and P14, a subset persisted as sex-biased throughout the neonatal critical window (Fig. 5.1f-g, Fig. 5.5h); in i1:Nfix neurons, ~20% of differentially-expressed genes on P4 persisted as sex-biased on P14. These genes regulate distinct components of neural circuit development, including neurite extension (*Klhl1*, *Pak7*), axon pathfinding

(*Epha3*, *Nell2*), neurotransmission (*Kcnab1*, *Scg2*), and synapse formation (*Il1rap*, *Tenm2*) (Fig. 5.1g, Fig. 5.5h). Together, these results demonstrate that neonatal ERα activation drives the epigenetic maintenance of a gene expression program that facilitates sexual differentiation of BNSTp circuitry.

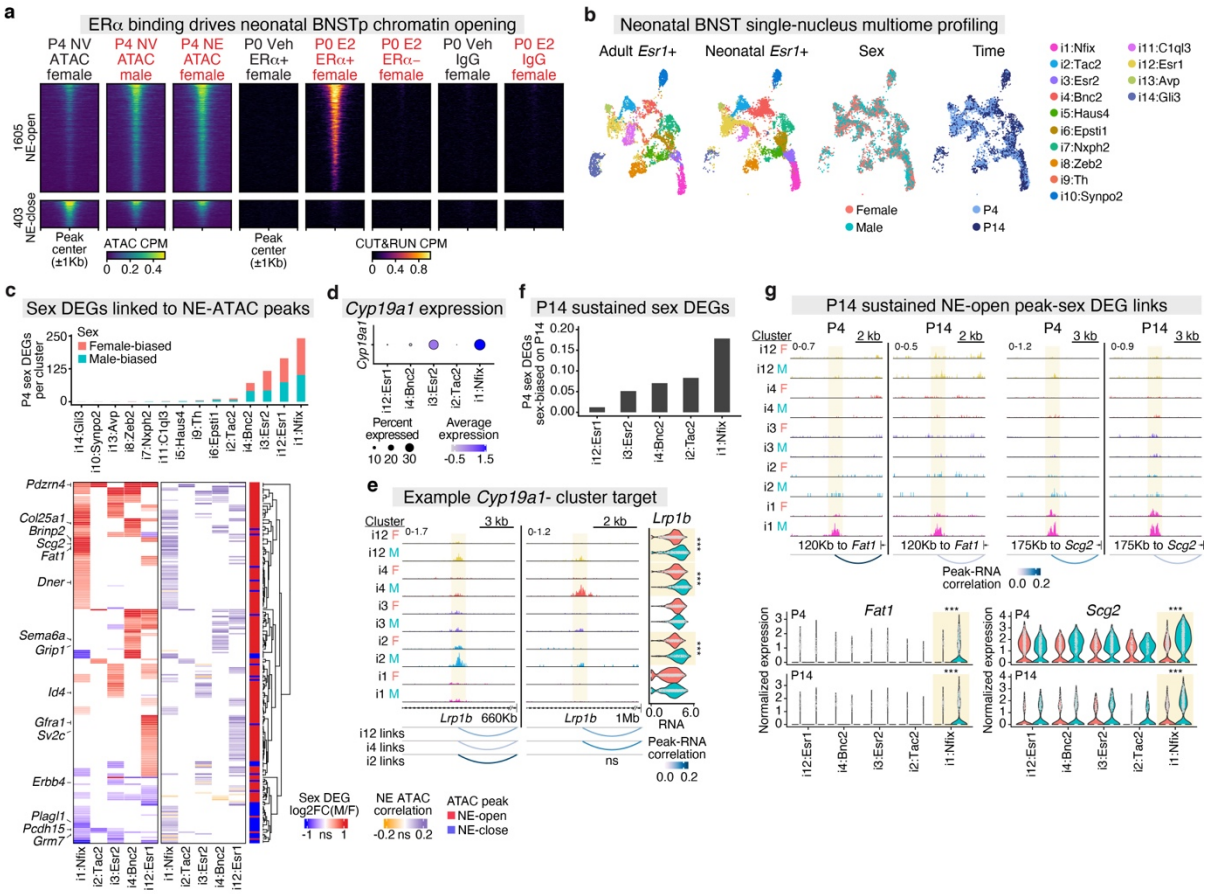

Figure 5.1. Neonatal ERα binding drives a sustained male-biased gene expression program. (**a**) Heatmap of P4 BNST *Esr1*+ ATAC, P0 IgG CUT&RUN, and P0 ERα CUT&RUN CPM ±1Kb around 1605 neonatal estradiol (NE)-open and 403 NE-close ATAC peaks (edgeR, padj<0.1). ERα+=Sun1-GFP+ nuclei, ERα-=Sun1-GFP- nuclei. (**b**) UMAPs of adult (left) and neonatal (middle left) BNST *Esr1*+ snRNA-seq clusters; neonatal snRNA-seq clusters colored by sex (middle right) and timepoint (right). (**c**) (top) Number of sex DEGs (MAST, padj<0.05) in P4 multiome clusters. (bottom) Heatmaps indicating (left) RNA log2FC of P4 sex DEGs and (right) Pearson's correlation coefficient of NE-open (red) and -close (blue) peaks linked to sex DEGs in each cluster. Non-significant genes and coefficients colored in white. (**d**) *Cyp19a1* expression on P4. (**e**) (left) NE-open peaks correlating with *Lrp1b* expression in *Cyp19a1*- clusters, i2:Tac2 and i12:Esr1. (right) Sex difference in *Lrp1b* expression in i2:Tac2 (n=260 female, 153 male, padj=2.13×10^{-8}), i4:Bnc2 (n=437 female, 373 male, padj=5.62×10^{-37}), i12:Esr1 cells (n=803 female, 507 male, padj=1.09×10^{-12}). ***padj<0.001, MAST. (**f**) Proportion of P4 sex DEGs detected as sex-biased on P14. (**g**) (top) i1:Nfix-specific, NE-open ATAC peaks at *Fat1* and *Scg2* loci on P4 and P14. (bottom) Sex difference in i1:Nfix *Fat1* and *Scg2* expression on P4 (*Fat1*, padj=1.28×10^{-37}; *Scg2*, padj=1.54×10^{-46}; n=887 female, 676 male) and P14 (*Fat1*, padj=1.13×10^{-11}; *Scg2*, padj=1.52×10^{-5}; n=554 female, 829 male). ***padj<0.001, MAST.

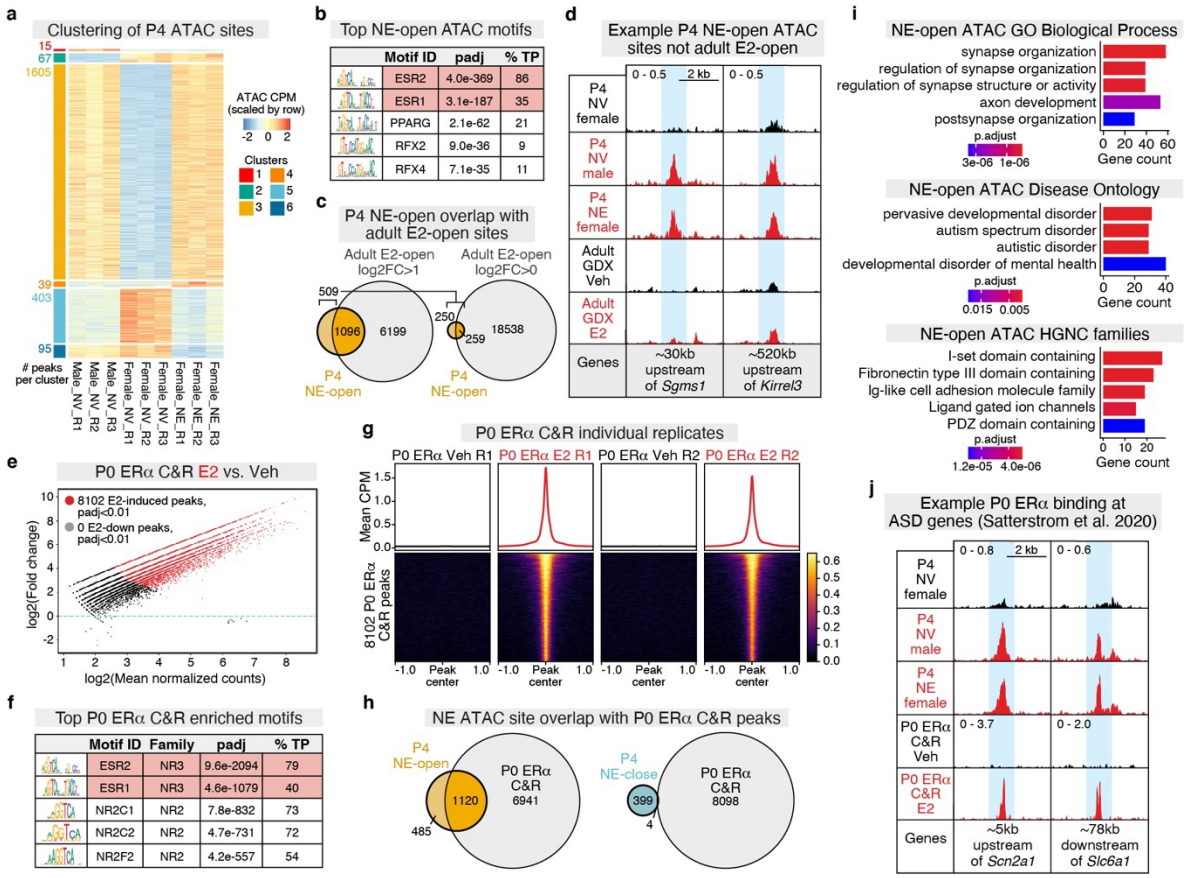

Figure 5.2. P4 ATAC-seq and P0 ERα CUT&RUN analysis. (**a**) Heatmap of mean ATAC CPM for P4 NV male, NV female, and NE female individual replicates (n=3 per condition) at differential peaks (edgeR, glmQLFTest, padj<0.1), grouped by hierarchical clustering (cutree, k=6). Clusters c3 and c5 correspond to NE-open and NE-close sites, respectively, shown in Fig. 5.1a. (**b**) Top enriched motifs (AME) in NE-open ATAC peaks. (**c**) (left) Overlap between P4 NE-open ATAC peaks and adult E2-open ATAC peaks (log2FC>1). (right) Overlap between remaining 509 P4 NE-open ATAC peaks and log2FC>0 E2-open ATAC peaks. (**d**) Example P4 NE-open ATAC peaks not detected as E2-induced in adult E2-open ATAC peakset. (**e**) MA plot of P0 female E2 vs. female Veh ERα CUT&RUN peaks (DiffBind, DESeq2, padj<0.01); red dots=E2-induced peaks, grey dots=E2-down peaks. (**f**) Top enriched motifs (AME) in P0 E2-induced ERα peaks. (**g**) Heatmap of mean P0 ERα CUT&RUN CPM ±1Kb around 8102 E2-induced ERα peaks for individual replicates (n=2 per condition). (**h**) (left) Overlap between P4 NE-open ATAC peaks and P0 E2-induced ERα peaks. (right) Overlap between P4 NE-close ATAC peaks and P0 E2-induced ERα peaks. (**i**) (top) Top GO Biological Process terms (clusterProfiler, padj<0.1), (middle) DO terms (clusterProfiler, padj<0.1), and (bottom) HGNC gene families (clusterProfiler, padj<0.1) enriched within P4 NE-open peak-associated genes. (**j**) Example P0 ERα peaks overlapping P4 NE-open peaks at high-confidence ASD candidate genes, *Scn2a1* and *Slc6a1*.

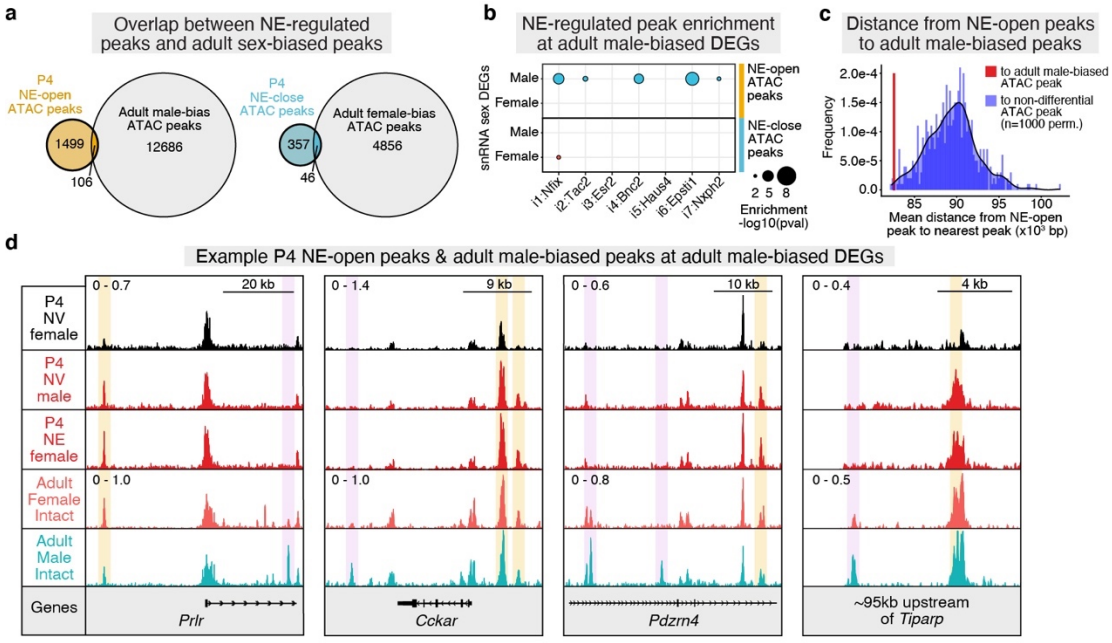

Figure 5.3. Comparison of P4 and adult *Esr1*+ ATAC-seq. (**a**) (left) Overlap between P4 NE-open ATAC peaks and gonadally intact adult male-biased ATAC peaks. (right) Overlap between P4 NE-close ATAC peaks and gonadally intact adult female-biased ATAC peaks. (**b**) Dotplot of BETA enrichment p-values for P4 NE-open ATAC peaks (top) and NE-close ATAC peaks (bottom) at adult BNST snRNA-seq sex DEGs relative to a background of non-differential, expressed genes. (**c**) Histogram of mean distance between P4 NE-open peaks and nearest gonadally intact adult male-biased ATAC peak (red line) vs. nearest chromosome-matched, non-differential adult ATAC peak (n=1000 permutations) (blue histogram). Mean distance between P4 NE-open peaks and adult male-biased peaks is significantly smaller than the expected distribution (Permutation test, p=0.007). (**d**) Example adult male-biased DEGs (*Prlr*, *Cckar*, *Pdzrn4*, *Tiparp*) with neighboring P4 NE-open (highlighted in yellow) and adult male-biased ATAC peaks (highlighted in purple).

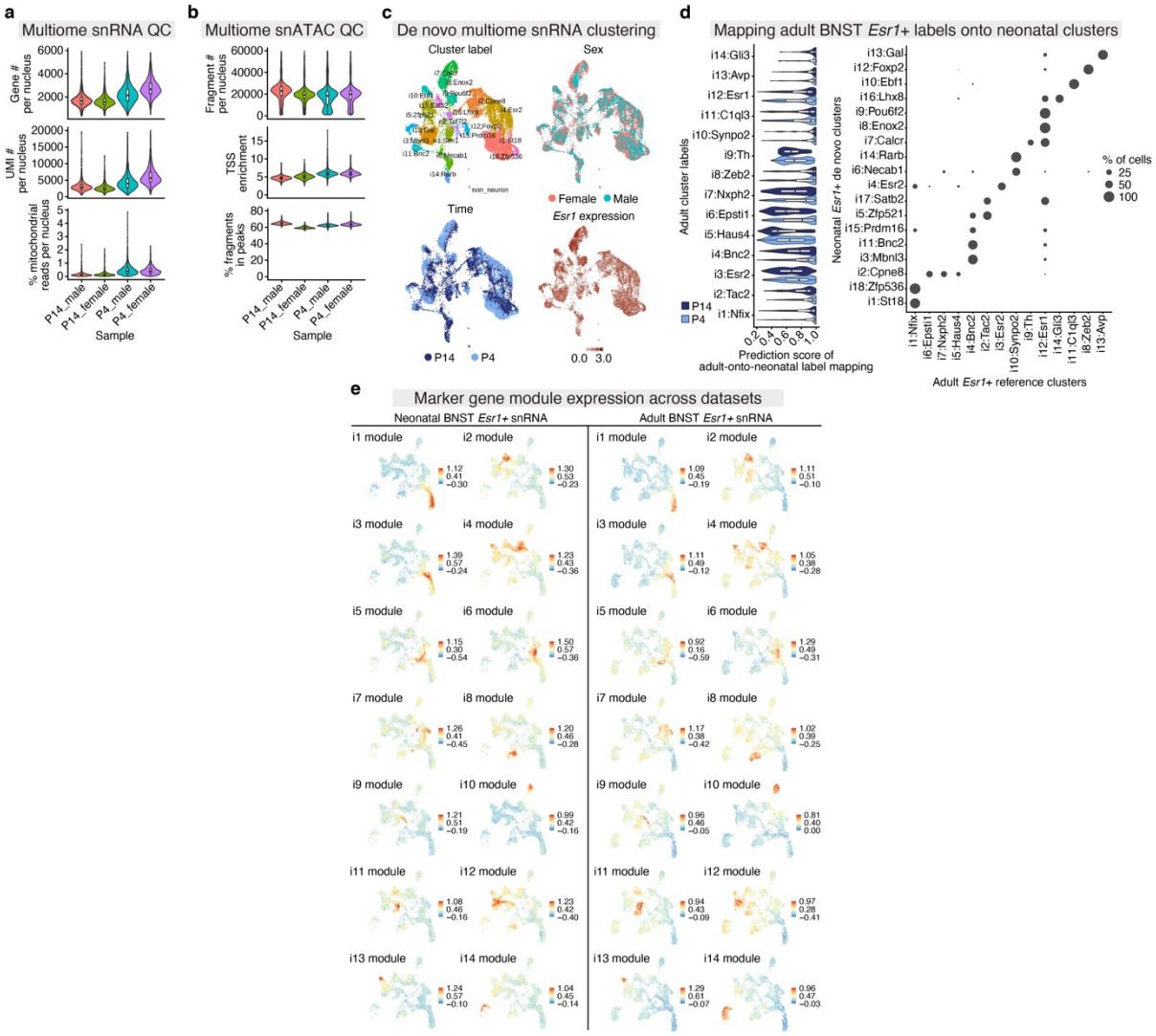

Figure 5.4. Additional analysis of neonatal BNST *Esr1*+ single-nucleus multiome dataset. (**a-b**), RNA (**a**) and ATAC (**b**) quality control (QC) metrics for neonatal (P4, P14) single-nucleus multiome experiments, split by timepoint and sex. Boxplot center=median, box boundaries=1st and 3rd quartile, whiskers=minimum and maximum values. n=4265 P14_male, 3148 P14_female, 3128 P4_male, 4295 P4_female. (**c**) UMAPs of *de novo* clustering of neonatal multiome snRNA data, colored by cluster identity (top left), sex (top right), timepoint (bottom left), and *Esr1* expression (bottom right). (**d**) (left) Prediction scores of adult-to-neonatal label transfer for each adult BNST *Esr1*+ reference cluster, split by timepoint. Boxplot center=median, box boundaries=1st and 3rd quartile, whiskers=minimum and maximum values. n=14836 cells. (right) % of nuclei in each neonatal *de novo* cluster that mapped to each adult BNST *Esr1*+ cluster. (**e**) UMAPs of neonatal marker gene module expression in neonatal dataset (left) and adult dataset (right).

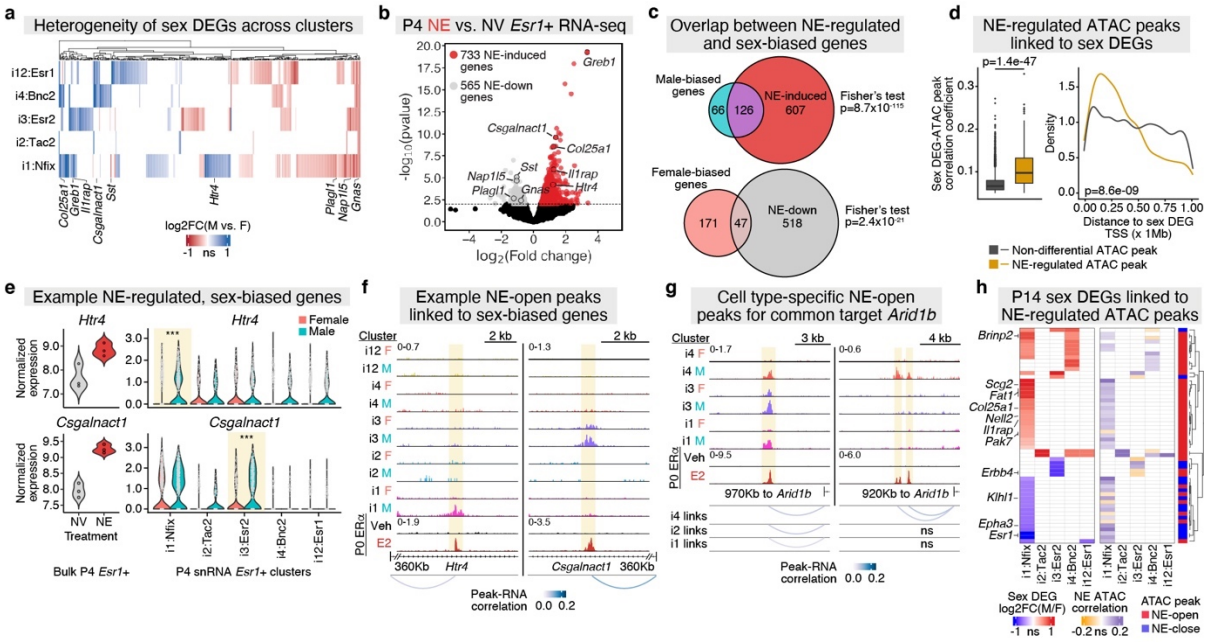

Figure 5.5. Sex differences in single-nucleus multiome dataset. (**a**) Hierarchical clustering of log2FC values for P4 sex DEGs detected in *Esr1*+ inhibitory neuron clusters (see also Fig. 3e). Sex DEGs with non-significant differential expression colored in white. (**b**) Neonatal E2 (NE) vs. neonatal vehicle (NV) female nuclear RNA-seq on P4 BNST *Esr1*+ cells; grey, red dots (DESeq2, padj<0.1). (**c**) (top) Overlap between NE-induced genes and P4 multiome male-biased genes. (bottom) Overlap between NE-downregulated genes and female-biased genes. (**d**) (left) Pearson's correlation coefficient values for non-differential (grey) and NE-regulated (gold) ATAC peaks that correlate with P4 sex DEG expression. Boxplot center=median, box boundaries=1st and 3rd quartile, whiskers=1.5*IQR from boundaries. n=5169 non-differential, 244 NE-regulated. p-value from two-sided, Wilcoxon rank-sum test. (right) Distance between non-differential (grey) and NE-regulated (gold) ATAC peaks to P4 sex DEG transcription start sites (TSS). p-value from Kolmogorov-Smirnov test. (**e**) Example P4 sex-biased genes that are also NE-regulated, *Htr4* (top) and *Csgalnact1* (bottom). (left) n=3, (right) n=887 i1:Nfix female cells, 676 i1:Nfix male cells, 404 i3:Esr2 female cells, 550 i3:Esr2 male cells. (**f**) Tracks for NE-open ATAC peaks that correlate with NE-regulated, sex-biased targets, *Htr4* and *Csgalnact1*. (**g**) Different NE-open ATAC peaks across i1:Nfix, i3:Esr2, and i4:Bnc2 neurons correlated with a common male-biased target, *Arid1b*. (**h**) Heatmaps indicating (left) RNA log2FC of P14 sex DEGs and (right) Pearson's correlation coefficient of NE-open (red) and -close (blue) ATAC peaks linked to sex DEGs within each cluster. Non-significant genes and correlation values colored in white.

5.3.3 ERα is required for sustained sex differences in gene expression

The adult display of male mating and territoriality behaviors requires ERα expression in GABAergic neurons (M. V. Wu & Tollkuhn, 2017b). To determine whether ERα is also required for the sustained sex differences in gene expression observed in the multiome dataset, I performed snRNA-seq on 38,962 BNST GABAergic neurons isolated from P14 conditional mutant males lacking ERα (*Vgat*^{Cre}; *Esr1*$^{lx/lx}$; *Sun1-GFP*^{lx}), and littermate control females and males (*Vgat*^{Cre}; *Esr1*$^{+/+}$; *Sun1-GFP*^{lx}) (Fig. 5.6a). The identity of GABAergic neurons in ERα mutant males did not deviate from P14 control or adult BNST neuron types (Fig. 5.6a). However, the abundance of male-biased i1:Nfix and i3:Esr2 neurons reduced to female levels in *Vgat*^{Cre}; *Esr1*$^{lx/lx}$ males (Fig. 5.7a), indicating neonatal ERα activation is required for their male-typical abundance.

Differential expression analysis between control females and control or ERα KO males within each neuron type revealed that ERα is required for nearly all sexually dimorphic gene expression, with the exception of genes located on the Y chromosome or escaping X inactivation (Fig. 5.6b). As expected, the loss of sex differences was largely driven by feminized expression in ERα KO males (Fig. 5.6b). Of note, despite expressing ERβ, which binds estradiol, i3:Esr2 neurons still lost sex-biased gene expression in the absence of ERα (Fig. 5.6b). Together, these findings demonstrate the neonatal hormone surge drives a sustained male-typical gene expression program via activation of a master regulator TF, ERα (Fig. 5.6c).

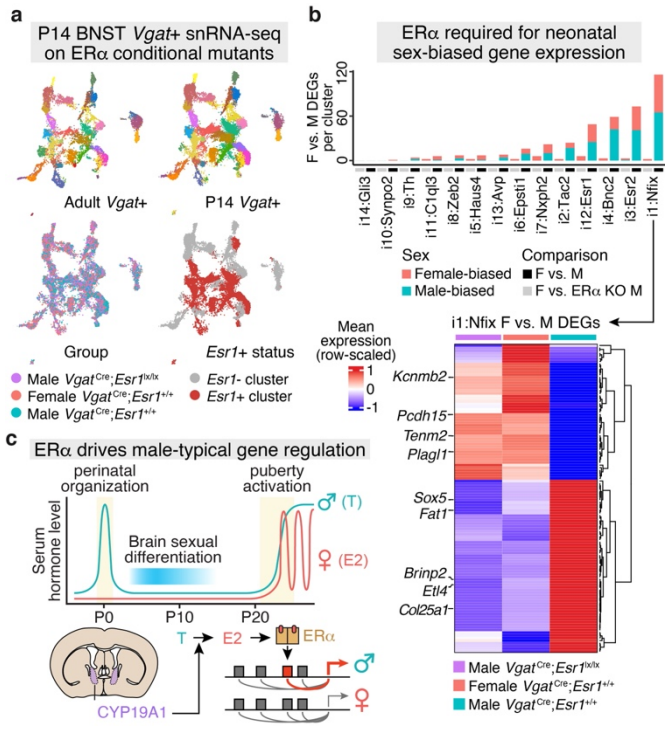

Figure 5.6. ERα is required for sustained sex differences in gene expression. (**a**) UMAPs of adult (top left) and P14 (top right) BNST *Vgat*+ snRNA-seq clusters, P14 *Vgat*+ snRNA clusters colored by group (bottom left) and *Esr1*+ status (bottom right). (**b**) (top) Number of F vs. M sex DEGs (MAST, padj<0.05) in P14 snRNA clusters (black bar). Number of F vs. M sex DEGs detected in KO vs. M comparison (grey bar). (bottom) Heatmap of mean expression of i1:Nfix sex DEGs, scaled across control males, control females, and conditional ERα KO males. (**c**) Neonatal ERα activation drives a sustained male-typical gene expression program.

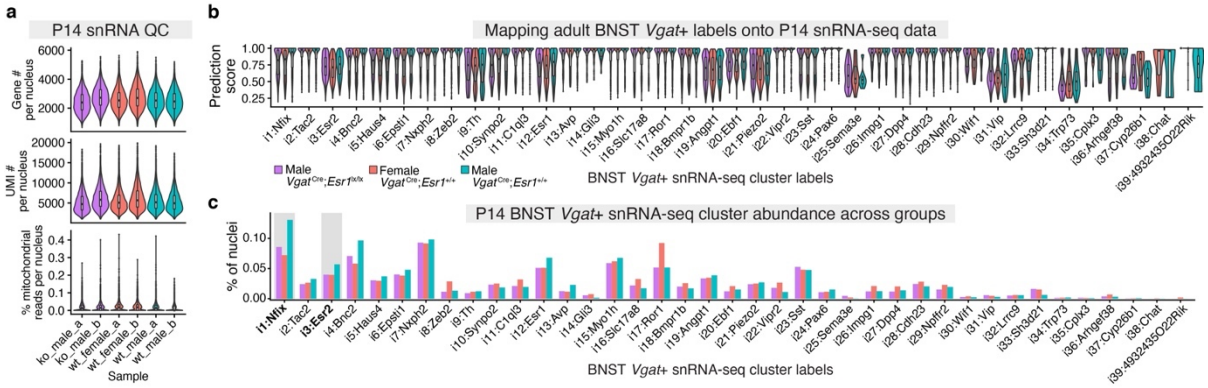

Figure 5.7. Additional analysis of P14 BNST *Vgat*+ snRNA-seq dataset. (**a**) RNA QC metrics for P14 BNST *Vgat*+ snRNA-seq experiment, split by sample. a and b refer to technical replicates. Boxplot center=median, box boundaries=1st and 3rd quartile, whiskers=minimum and maximum values. n=6355 ko_male_a, 5614 ko_male_b, 7184 wt_female_a, 6367 wt_female_b, 6881 wt_male_a, 6561 wt_male_b. (**b**) Prediction scores of adult-to-P14 label transfer for each adult BNST *Vgat*+ reference cluster, split by group. Boxplot center=median, box boundaries=1st and 3rd quartile, whiskers=minimum and maximum values. n=38962 cells. (**c**) Proportion of total P14 *Vgat*+ nuclei in each *Vgat*+ cluster, separated by group. Adult male-biased *Esr1*+ clusters i1:Nfix and i3:Esr2 are indicated in grey.

5.4 Discussion

In this Chapter, I report that activation of ERα by the neonatal hormone surge drives a sustained male-typical gene regulatory program in the developing BNSTp. This program is required for brain sexual differentiation, as selective deletion of ERα in *Vgat*+ inhibitory neurons eliminates neonatal sex differences in gene expression and, as previously shown, impairs male-typical mating and territoriality behavior (M. V. Wu & Tollkuhn, 2017a). Interestingly, ERβ does not appear to compensate for the removal of ERα, as the i3:Esr2 neuron type loses sex-biased gene expression upon ERα deletion (Fig. 5.6b). This finding implies that ERβ does not respond to neonatal estradiol, consistent with a non-sex-specific function in brain development (L. Wang et al., 2003). In Chapter 6, I discuss the possibility that ERβ may instead act as a regulator of inhibitory neuron identity, although more experiments are required to determine whether it acts in a ligand-independent manner or utilizes alternative estrogen ligands.

How does ERα regulate chromatin accessibility following the neonatal surge? Prior studies in MCF-7 cells have revealed ERα binds an extensive repertoire of cofactors in response to ligand (Mohammed et al., 2013; Papachristou et al., 2018), including histone acetyltransferases CREBBP, EP300, NCOA3, and CARM1. In addition, ERα binds the dioxygenase TET2, which catalyzes the conversion of 5mC to 5hmC, resulting in enhancer activation via demethylation (Broome et al., 2021; L. Wang et al., 2018). To date, protein binding partners of ERα have not been identified in the brain. Identifying these partners, either via co-immunoprecipitation mass spectrometry of a peptide-tagged ERα or biotinylation-based proximity labeling, is essential to determining the epigenomic mechanisms underlying sex differences in chromatin accessibility. Moreover, it remains unclear how neonatal ERα binding leads to persistent regulation of chromatin accessibility, as the duration of neural estradiol production following the testosterone surge has not been directly measured. For instance, ERα may remain bound to the chromatin for several days, or it may transiently open the chromatin as a PF, allowing other TFs to then bind and maintain accessibility. While hormone receptors have not

historically been considered PFs, recent *in vitro* data from Gordon Hager's lab suggests that ERα, AR, and GR can open closed chromatin and facilitate recruitment of other TFs (Paakinaho et al., 2019; Swinstead et al., 2016). Consistent with this model, in Chapter 6, I describe how estradiol drives the genomic binding of a neuron type-specific TF.

How does the neonatal hormone surge gise rise to sex differences in BNSTp circuitry? The identity of sex-biased genes provides some clues as to how neonatal estradiol specifies these sex differences. For instance, *Pak7*/*Pak5*, male-biased specifically within the i1:Nfix neuron type, encodes a mitochondrial kinase that facilitates axonogenesis by boosting axonal ATP production (N. Huang et al., 2021). *Crim1* has also been recently found to drive extension of corticospinal neuron axons, around the same time window (~P4) that BNSTp neurons initiate this process (G. Gu et al., 2003; Sahni et al., 2021). A number of genes involved in neurotransmission and physiology, such as *Asic2*, *Htr4*, *Kcnab1*, and *Scg2*, were also sex-biased in expression, suggesting sex differences in the regulation of neuronal activity. Surprisingly, despite the well-characterized male-biased neuron survival phenotype during BNSTp development (Forger et al., 2004b; Gotsiridze et al., 2007; M. V. Wu et al., 2009a), none of the canonical pro-survival/pro-apoptotic genes (*Bax*, *Bak*, *Bcl-2*, *Puma*, or caspases) had sex-biased expression, suggesting that neuron survival may be a downstream consequence of neuronal wiring and/or integration into functional circuits (Pfisterer & Khodosevich, 2017). Future investigation into neonatal sex differences in the connectivity and activity of specific BNSTp neuron types will link neuron type-specific sex-biased genes to sexually differentiated circuitry.

Lastly, it will be important to examine whether epigenomic modifications imparted by the neonatal surge persist throughout life, and whether such persistence is required for sex differences in behavior. Despite a major increase in sex differences following puberty (Fig. 4.1h), a low number of sites (~150) maintain sex-biased accessibility between P4 and adulthood (Fig. 5.3). It is possible that additional sites are maintained but cannot be detected as sex-biased in the bulk *Esr1*+ population. Preliminary analysis of a snATAC-seq dataset collected from adult conditional mutant females and

males lacking ERα (*Vgat*^{Cre}; *Esr1*$^{lx/lx}$; *Sun1-GFP*^{lx}), and littermate controls (*Vgat*^{Cre}; *Esr1*$^{lx/+}$; *Sun1-GFP*^{lx}) has revealed additional persistent sites, which are also feminized upon ERα deletion (data not shown). However, it remains unclear whether the persistency of sex differences at these loci is, in itself, required for behavioral sex differences. For instance, it is possible that they only need to differ between sexes during circuit formation, after which time they simply represent a scar or shadow of a prior developmental event. Silencing persistent, sex-biased loci in adulthood by injecting a Cre-dependent CRISPR interference (CRISPRi) virus and sgRNA library targeting these sites into the BNSTp may identify whether their persistency is required for the display of sex-typical behaviors.

5.5 Materials and Methods

Animals. All animals were maintained on a 12:12 light cycle and provided food and water *ad libitum*. All mouse experiments were performed under strict guidelines set forth by the CSHL Institutional Animal Care and Use Committee (IACUC). *Esr1*^{Cre} (H. Lee et al., 2014), *ROSA26*$^{CAG\text{-}Sun1\text{-}sfGFP\text{-}Myc}$ (Mo et al., 2015) (abbreviated as *Sun1-GFP*), and *Vgat*^{Cre} (Vong et al., 2011) mice were obtained from Jackson labs. *Esr1*^{lx} mice were received from Sohaib A. Khan (Feng et al., 2007). For neonatal CUT&RUN, ATAC-seq, and RNA-seq experiments, animals were treated with 5µg E2 or vehicle on P0 and harvested four hours later (ERα CUT&RUN) or four days later (ATAC-seq, nuclear RNA-seq). For neonatal multiome, snRNA-seq, and IF quantification, animals were harvested on P4 (multiome) or P14 (multiome, snRNA-seq, IF staining).

Neonatal bulk ATAC-seq. Female and male *Esr1*$^{Cre/+}$; *Sun1-GFP*$^{lx/+}$ mice were injected subcutaneously with 5µg E2 or vehicle on P0 and harvested on P4 (4-5 animals pooled per condition and per replicate). The BNSTp was microdissected and collected in 300 µl of cold, supplemented homogenization buffer. Nuclei were extracted as described in Chapter 2 Materials and Methods. After filtering through a 40-µm strainer, the nuclei were diluted 3:1 with 600 µl of cold, supplemented

homogenization buffer and immediately used for sorting. 30,000 GFP+ nuclei were collected into cold ATAC-RSB buffer using the Sony SH800S Cell Sorter (purity mode) with a 100-µm sorting chip. After sorting, nuclei transposition and library preparation were performed, as previously described.

P0 ERα CUT&RUN. Female *Esr1*$^{Cre/+}$; *Sun1-GFP*$^{lx/+}$ mice were injected subcutaneously with 5µg E2 or vehicle on P0 and harvested four hours later (5 animals pooled per condition and per replicate). The BNSTp, MPOA, and MeA were microdissected, and nuclei were extracted, as described for the neonatal bulk ATAC-seq experiment. After filtering through a 40-µm strainer, the nuclei were diluted 3:1 with 600 µl of cold, supplemented homogenization buffer. 2 mM EDTA was added, and the sample was immediately used for sorting. 150,000 GFP+ nuclei were collected into cold CUT&RUN wash buffer using the Sony SH800S Cell Sorter (purity mode) with a 100-µm sorting chip. GFP- events were collected into cold CUT&RUN wash buffer, and 150,000 nuclei were subsequently counted on the Countess II FL Automated Cell Counter for ERα- and IgG negative control CUT&RUN. All subsequent steps were performed as described for the adult brain CUT&RUN experiments in Chapter 2 Materials and Methods. P0 CUT&RUN libraries were prepared with 10 PCR cycles.

Neonatal single-nucleus multiome-seq. The BNST was microdissected fresh from P4 and P14 female and male *Esr1*$^{Cre/+}$; *Sun1-GFP*$^{lx/+}$ mice, as described above (4-5 animals pooled per condition). Nuclei were extracted and prepared for sorting, as done for the neonatal bulk ATAC-seq experiment, with the inclusion of 1 U/µl Protector RNase inhibitor (Sigma) in the homogenization buffer. 40,000-50,000 GFP+ nuclei were collected into 1 ml of cold ATAC-RSB buffer, supplemented with 0.1% Tween-20, 0.01% digitonin, 2% sterile-filtered BSA (Sigma A9576), and 1 U/µl Protector RNase inhibitor. The nuclei were centrifuged in a swinging-bucket rotor at 500 x g for 10 minutes at 4ºC. ~950 µl of supernatant was carefully removed, and 200 µl 10X Genomics Dilute Nuclei Buffer was added to the side of the tube without disturbing the pellet. The nuclei were centrifuged again at 500 x g for 10

minutes at 4ºC. ~240 µl of supernatant was carefully removed, and the nuclei were resuspended in the remaining volume (~7 µl). Samples were immediately used for the 10X Genomics Single Cell Multiome ATAC + Gene Expression kit (1000285), following manufacturer's instructions. snRNA-seq and snATAC-seq libraries were sequenced on an Illumina NextSeq, using the High Output kit. Each sample was sequenced to a depth of ~40,000-80,000 mean reads per cell for the snATAC library and ~40,000-50,000 mean reads per cell for the snRNA library.

P14 snRNA-seq. The BNSTp was microdissected from P14 female and male *Vgat*^{Cre};*Esr1*$^{+/+}$;*Sun1-GFP*^{lx} and male *Vgat*^{Cre};*Esr1*$^{lx/lx}$;*Sun1-GFP*^{lx} mice. Tissue samples from individual animals were immediately flash-frozen in an ethanol dry ice bath and stored at -80ºC until n=3 animals were collected per group. On the day of the experiment, tissue samples were removed from -80ºC and maintained on dry ice. With the tissue still frozen, cold, supplemented homogenization buffer was added to the tube, and the tissue was immediately transferred to a glass homogenizer and mechanically dounced and filtered, as described for our other neonatal experiments. 80,000-90,000 GFP+ nuclei were collected into 100 µl of cold ATAC-RSB buffer, supplemented with 1% sterile-filtered BSA (Sigma A9576), and 1 U/µl Protector RNase inhibitor, in a 0.5 ml DNA lo-bind tube (Eppendorf) pre-coated with 30% BSA. After collection, nuclei were pelleted with 2 rounds of gentle centrifugation (200 x g for 1 min) in a swinging-bucket centrifuge at 4ºC. After the second round, the supernatant was carefully removed, leaving ~40 µl in the tube. The nuclei were gently resuspended in this remaining volume and immediately used for the 10X Genomics Single Cell 3' Gene Expression kit v3 (1000424), following the manufacturer's instructions. Each biological sample was split into two 10X lanes, producing 6 libraries that were pooled and sequenced on an Illumina NextSeq 2000 to a depth of ~45,000-60,000 mean reads per cell.

Neonatal nuclear RNA-seq. Female *Esr1*$^{Cre/+}$; *Sun1-GFP*$^{lx/+}$ mice were injected subcutaneously with 5µg E2 or vehicle on P0. Four days later, animals were rapidly decapitated, and 400-µm sections were collected in cold homogenization buffer using a microtome (Thermo Scientific Microm HM 650V). The BNST was microdissected (4 animals pooled per condition) and collected in 1 ml of cold, supplemented homogenization buffer containing 0.4 U/ml RNAseOUT (Thermo Fisher, 10777019). Nuclei isolation was performed as described for neonatal bulk ATAC-seq. 12,000 GFP+ nuclei were collected into cold Buffer RLT Plus supplemented 1:100 with β-mercaptoethanol (Qiagen, 74034) using the Sony SH800S Cell Sorter (purity mode) with a 100-µm sorting chip. Nuclei lysates were stored at -80°C until all replicates were collected. Nuclei samples for all replicates were thawed on ice, and RNA was isolated using the Qiagen RNeasy Plus Micro Kit (74034). Strand-specific RNA-seq libraries were prepared using the Ovation SoLo RNA-seq system (Tecan Genomics, 0501-32), following manufacturer's guidelines. Individually barcoded libraries were multiplexed and sequenced with single-end 76bp reads on an Illumina NextSeq, using the Mid Output Kit.

CUT&RUN data processing. P0 ERα CUT&RUN data were processed, as described previously in Chapter 2 Materials and Methods.

CUT&RUN data analysis. CUT&RUN differential peak calling was performed with DiffBind (Stark et al., 2011). A count matrix was created from individual replicate BAM and MACS2 narrowpeak files (n=2 per condition). Consensus peaks were re-centered to ±100 bp around the point of highest read density (summits=100). Differential peak-calling for the P0 ERα CUT&RUN experiment was performed with DESeq2 (padj<0.01) within DiffBind.

Neonatal bulk ATAC-seq analysis. Differential peak calling on the neonatal bulk ATAC-seq experiment was performed with DiffBind and edgeR. A count matrix was created from individual

replicate BAM and MACS2 narrowpeak files (n=3 per condition). Consensus peaks were re-centered to ±250 bp around the point of highest read density (summits=250), and the consensus peakset count matrix was subsequently used as input to edgeR. Differential peaks across the three treatment groups (NV female, NV male, NE female) were calculated by specifying multiple coefficients in glmQLFTest (coef=4:5). To identify accessibility patterns across differential peaks (padj<0.05), a matrix of normalized counts within differential peaks was hierarchically clustered using pheatmap, and the resulting dendrogram tree was cut with k=6 to achieve 6 peak clusters. The two largest clusters were identified as having higher accessibility in NV males and NE females compared to NV females (cluster 3, labeled as NE-open), or lower accessibility in NV male and NE female compared to NV females (cluster 5, labeled as NE-close). Motif enrichment analysis of NE-open peaks was performed with AME using the 2020 JASPAR core non-redundant vertebrate database. GO Biological Process, DO, and HGNC gene family enrichment analyses were performed, as described above for adult GDX treatment ATAC-seq data analysis.

Neonatal single-nucleus multiome data processing and analysis. Raw sequencing data were processed using the Cell Ranger ARC pipeline (v2.0.0) with the cellranger-arc mm10 reference. Default parameters were used to align reads, count unique fragments or transcripts, and filter high-quality nuclei. Individual HDF5 files for each sample containing RNA counts and ATAC fragments per cell barcode were loaded into Seurat (Read10X_h5). Nuclei with lower-end ATAC and RNA QC metrics (<1000 ATAC fragments, <500 counts, nucleosomal signal>3, TSS enrichment<2) were removed. DoubletFinder (McGinnis et al., 2019) was then used to remove predicted doublets from each sample (nExp=9% of nuclei per sample). Following doublet removal, nuclei surpassing upper-end ATAC and RNA QC metrics (>60000 ATAC fragments, >20000 RNA counts, >6000 genes detected) were removed. After filtering, Seurat objects for each sample were subsetted for the RNA assay and merged. Gene counts were normalized and log-transformed ('LogNormalize'), and the top 2000 variable

features were identified using FindVariableFeatures (selection.method='vst'). Gene counts were scaled, regressing out the following variables: number of RNA counts, number of RNA genes, percentage of mitochondrial counts, and biological sex. Linear dimensionality reduction was performed by principal component analysis (runPCA, npcs=25). A *k*-nearest-neighbors graph was constructed on the basis of Euclidean distance in PCA space and refined (FindNeighbors, npcs=25), then the nuclei were clustered using the Louvain algorithm (FindClusters, resolution=0.8). snRNA clusters were visualized with UMAP (runUMAP, dims=25). To reduce the granularity of clustering, a phylogenetic tree of cluster identities was generated from a distance matrix constructed in PCA space (BuildClusterTree) and visualized as a dendrogram (PlotClusterTree). DEGs between clusters within terminal nodes of the phylogenetic tree were calculated (FindMarkers, test.use='wilcox', padj<0.05), and clusters were merged if they had fewer than 10 DEGs with the following parameters: >0.5 avg_logFC, <10% expression in negative nuclei, and >25% expression in positive nuclei. The final *de novo* snRNA-seq clusters are shown in Fig. 5.4c.

Inhibitory neuron clusters (*Slc32a1*/*Gad2*+) from the neonatal multiome dataset were subsequently assigned to adult BNST *Esr1*+ cluster labels using Seurat. Adult BNST *Esr1*+ clusters (as defined above) were subsetted from the adult snRNA-seq object and randomly downsampled to 5000 nuclei. Normalization, data-scaling, and linear dimensionality reduction were performed with the same parameters for neonatal and adult *Esr1*+ inhibitory neuron clusters. Anchor cells between adult (reference) and neonatal (query) datasets were first identified using FindTransferAnchors. Reference cluster labels, as well as the corresponding UMAP structure, were subsequently transferred to the neonatal dataset using MapQuery. Prediction scores, which measure anchor consistency across the neighborhood structure of reference and query datasets as previously described (Stuart et al., 2019), were used to quantify the confidence of label transfer from adult to neonatal nuclei. Fig. 5.4d shows the prediction scores per reference cluster and timepoint of nuclei mapped onto adult reference cluster labels as well as the % of nuclei from each de novo cluster mapped onto each adult reference cluster

(prediction score>0.5). To further validate the quality of label-transfer between adult and neonatal datasets, we computed DEGs between neonatal clusters post-label-transfer (FindMarkers, test.use='wilcox', padj<0.05, min.diff.pct=0.1, avg_logFC>0.5) and calculated their background-subtracted, average expression (AddModuleScore) in neonatal and adult BNST *Esr1*+ nuclei (shown in Fig. 5.4e).

To generate pseudo-bulk, normalized ATAC bigwig tracks for each snATAC cluster, I first re-processed the cellranger ARC output BAM file for each sample using SAMtools (-q 30 -f 2) and removed duplicate reads per cell barcode using picard MarkDuplicates (BARCODE_TAG=CB REMOVE_DUPLICATES=true). Sinto (https://timoast.github.io/sinto/) was used to split ATAC alignments for each cluster into individual BAM files using cell barcodes extracted from the Seurat object. CPM-normalized bigwig files were computed for each pseudo-bulk BAM file using DeepTools bamCoverage (--binSize 1 --normalizeUsing CPM).

To analyze the neonatal multiome snATAC data, I used ArchR (Granja et al., 2021). Separate Arrow files were created for each multiome sample, then merged into a single ArchR project. Gene activity scores per nucleus were calculated at the time of Arrow file creation (addGeneScoreMat=TRUE). Metadata (cluster label, sex, time, QC metrics) were transferred from the previously generated Seurat object to the ArchR project by cell barcode-matching. Dimensionality reduction was performed on the snATAC data using ArchR's iterative Latent Semantic Indexing approach (addIterativeLSI). Per-nucleus imputation weights were added using MAGIC (van Dijk et al., 2018) within ArchR (addImputeWeights) to denoise sparse ATAC data for UMAP visualization. Cluster-aware ATAC peak-calling was performed using ArchR's iterative overlap peak merging approach (addReproduciblePeaks, groupBy='cluster').

To call sex DEGs (padj<0.05) within each cluster and timepoint, I used MAST (Finak et al., 2015) within Seurat (FindMarkers, test.use='MAST', min.pct=0.05, logfc.threshold=0.2, latent.vars= 'nFeature_RNA', 'nCount_RNA'). To link NE-regulated loci to sex DEGs at P4 and P14 (Fig. 5.1c,

Fig. 5.5h), I computed the Pearson correlation coefficient between sex DEG expression and NE-regulated peak accessibility for each cluster (LinkPeaks, min.distance=2000, distance=1000000, min.cells=2% of cluster size). Sex DEG log2FC values and NE-regulated ATAC site correlation coefficients were hiearchically clustered and visualized using ComplexHeatmap (Z. Gu et al., 2016).

P14 snRNA-seq data processing and analysis. Raw sequencing data were processed using the Cell Ranger pipeline (v6.0.0) with the refdata-gex-mm10-2020-A reference. Default parameters were used to align reads, count unique transcripts, and filter high-quality nuclei. Individual HDF5 files for each sample were loaded into Seurat. Nuclei with lower-end RNA QC metrics (<1000 counts) were removed. DoubletFinder (McGinnis et al., 2019) was then used to remove predicted doublets from each sample (nExp=9% of nuclei per sample). Following doublet removal, nuclei surpassing upper-end RNA QC metrics (>20000 counts, >6000 genes detected) were removed. After filtering, Seurat objects were merged. Gene counts were normalized and scaled, as described for the single-nucleus multiome data processing.

The P14 snRNA-seq dataset was assigned to adult BNST inhibitory cluster labels using Seurat. Adult BNST inhibitory clusters were subsetted from the adult snRNA-seq object and randomly downsampled to 10000 nuclei. Normalization, data-scaling, and linear dimensionality reduction were performed with the same parameters for P14 and adult inhibitory neuron clusters. Label transfer was then performed, as described for the single-nucleus multiome data processing. To validate the quality of label-transfer between adult and P14 datasets, I computed DEGs between P14 clusters post-label-transfer, as described above, and calculated their background-subtracted, average expression (AddModuleScore) in P14 and adult BNST inhibitory clusters. Sex DEGs between control female and control male, and between control female conditional ERα KO, conditions were calculated for each P14 cluster, as described above for the multiome analysis. Cluster abundance for each group was computed and plotted in Fig. 5.7.

Neonatal bulk nuclear RNA-seq data processing and analysis. Reads were trimmed to remove Illumina adapters and low-quality basecalls (cutadapt -q 30), then mapped to the mm10 reference genome using STAR. Technical duplicate reads (identical start and end positions with same strand orientation and identical molecular identifier) were removed using the nudup.py python package (https://github.com/tecangenomics/nudup). The number of reads mapping to each gene (including introns) on each strand (-s 1) was calculated with featureCounts (Liao et al., 2014), using the mm10.refGene.gtf file. Differential gene expression analysis was performed using DESeq2 (design = ~ treatment) after prefiltering genes by expression (rowMeans>=5).

Chapter 6

Characterization of neuron identity regulators in the BNST

Sections of this chapter were previously uploaded to bioRxiv (2020) under the title "Regulation of neural gene expression by estrogen receptor alpha" by Bruno Gegenhuber, Melody V. Wu, Robert Bronstein, and Jessica Tollkuhn. This manuscript is now in press at Nature (2022) under the title "Gene regulation by gonadal hormone receptors defines neuronal sex differences" by Bruno Gegenhuber, Melody V. Wu, Robert Bronstein, and Jessica Tollkuhn.

6.1 Abstract

Advancements in single-cell sequencing methods have revealed extensive transcriptomic and epigenomic diversity among neuron types in the mammalian brain. A central goal of neuroscience is to understand how combinatorial codes of TFs specify and maintain the identity of these neuron types, enabling their unique morphological, connectivity, and electrophysiological properties. Here I adopt a computational strategy to predict neuron identity regulator TFs from single-cell multiome (RNA + ATAC-seq) data. I demonstrate that this strategy highlights canonical GABAergic identity TFs *a priori* and reveals novel, putative regulator TFs. To experimentally validate this approach, I profile the genomic binding of a predicted identity regulator, Nfix, in the adult BNSTp and find that Nfix binding events are primarily accessible in the i1:Nfix neuron type. Moreover, sex-biased enhancers in the i1:Nfix type are also preferentially bound by Nfix, suggesting neuron identity TFs may enable diverse cellular responses to common signaling events. Together these results demonstrate an improved computational approach for predicting neuron identity regulators and provide a template for defining such regulators across brain regions and developmental stages.

6.2 Introduction

How can one understand the operation of a complex machine without a detailed understanding of its parts? This analogy describes a major goal of modern neuroscience: to characterize and catalog neuron types in the brain by integrating phenotypic measurements (*i.e.*, transcriptome, epigenome, electrophysiology, morphology, connectivity, *etc.*). Since the advent of high-throughput, single-cell RNA-sequencing, considerable progress has been made toward unsupervised, transcriptomic-based neuron classification (Tasic et al., 2016, 2018; Z. Yao et al., 2021), as well as establishing criteria for defining neuron types from transcriptomic data (Paul et al., 2017a). More recently, the Brain Initiative Cell Census Network, (BICCN), has generated a complete cell atlas of the mammalian primary motor cortex (MOp) by integrating multiple data modalities, including scRNA/snRNA-seq, single-nucleus ATAC-seq (snATAC-seq), single-nucleus DNA methylation-seq (snmC-seq2), MERFISH, Patch-seq, and anterograde/retrograde tracing (BRAIN Initiative Cell Census Network (BICCN), 2021). While this effort succeeded in matching transcriptome to phenotype for each neuron type, it still remains unclear how neuron identity is established and maintained, and whether mechanisms governing neuron identity are conserved across species.

At a fundamental level, cell identity is specified and maintained by TFs. In *C. elegans*, a model system with considerable genetic access, combinatorial codes of TFs, particularly of the homeobox family, specify neuron types by coordinating the expression of neuron type-specific gene batteries that impart unique physiological features (Hobert, 2021; Hobert et al., 2010; Serrano-Saiz et al., 2013; Taylor et al., 2021). Although such a model has been proposed to describe the regulation of neuron types in the mammalian brain (Fishell & Kepecs, 2020; Z. J. Huang & Paul, 2019; Paul et al., 2017a), aside from a few important examples (e.g., *Lhx6*, *Nkx2-1*, *Prox1*), a systematic understanding of this process, incorporating developmental single-cell transcriptomics and epigenomics, TF genomic binding, and loss-/gain-of-function experiments, has not yet been achieved.

To date, understanding of gene regulatory mechanisms in the mammalian brain has been limited both by technology and complexity of the system. As discussed in Chapter 2, prior to CUT&RUN, it has been nearly impossible to measure TF genomic binding in low numbers of cells, such as individual neuron types in the mammalian brain. Moreover, identifying identity regulator TFs requires detection of both TF expression *and* chromatin state at single-cell resolution, which has only recently been made possible with single-cell multiome sequencing approaches (Ma et al., 2020). To date, there is still no approach to measure RNA and TF binding in the same cell at high-throughput, although future advancements in CUT&RUN or CUT&Tag may enable this (see Discussion). In addition to technology, the complexity of the mammalian brain, and of neuron fate specification generally, has precluded systematic characterization of identity regulators. Aspects of this complexity are summarized in the following points:

1) Thus far, there is no clear relationship between mechanisms of gene regulation and the neuron types identified via integration of datasets generated by the BICCN. In other words, do all 56 MOp types have a distinct combinatorial code of TFs specifying their identity, or do these codes specify types at a certain level of granularity, such as distinguishing *Pv*+ from *Sst*+ interneurons, after which other non-genomic factors (*e.g.*, mRNA translation, alternative splicing, network connectivity) influence identity?

2) TFs act dynamically and combinatorially throughout development to control cell fate (Mayer et al., 2018), meaning TFs specifying identity may differ from those maintaining identity. Therefore, building a model of neuron type regulation requires integration of datasets across multiple developmental timepoints.

3) TFs specify fate via chromatin activation or repression (Greig et al., 2013). While it is feasible to predict transcriptional activators from accessible chromatin data, it is much more challenging to identify repressors without measuring gene expression, as closed chromatin may reflect the action of a repressive TF or lack of expression of an activator TF. Thus, identifying

modes of TF action in fate specification requires simultaneous measurement of transcriptomic and epigenomic data.

4) TFs can differentially influence fate specification between neuron types without themselves being differentially expressed (Allaway et al., 2021). Therefore, the function of a TF is better represented in its genomic binding profile rather than its expression pattern.

5) TFs within the same protein family often have a near-identical recognition motif, such as the homeodomain family or NFI family, making it difficult to predict specific identity regulators from chromatin profiling data alone.

In this Chapter, I describe a computational strategy to predict neuron identity regulator TFs from the single-cell multiome data collected in Chapter 5, taking into account several of the above points. I demonstrate that this approach improves predictions relative to the standard approach of performing motif enrichment analysis within cell type-specific marker peaks, in that it reveals canonical GABAergic identity TFs *a priori*. Moreover, several novel regulator TFs are predicted, which may be important for specifying identity in the BNST. To experimentally validate this computational strategy, I profile the genomic binding of a predicted identity regulator, Nfix, in the adult BNSTp and demonstrate that Nfix binding events are primarily accessible within the i1:Nfix neuron type. Future loss-/gain-of-function experiments will reveal whether Nfix is necessary and/or sufficient for specifying this identity. Collectively, these results establish an approach for high-throughput prediction of neuron identity regulator TFs across brain regions.

6.3 Results

6.3.1 Predicting neuron identity regulator TFs from single-nucleus multiome data

To identify TFs regulating BNSTp *Esr1*+ neuron identity, I first used a standard approach (Y. E. Li et al., 2021; Stuart et al., 2021) of calculating marker peaks for each neuron type in the dataset and subsequently finding the top enriched TF motifs within these marker peaks (Fig. 6.1a). This approach

highlighted specific families of TFs with restricted motif enrichment across *Esr1+* types; namely, POU TFs in i2:Tac2 neurons, NFI TFs in i1:Nfix neurons, and MEF2 TFs in i9:Th neurons (Fig. 6.1a). Many TFs known to specify the identity of inhibitory neurons throughout other regions of the brain, including *Prox1*, *Lhx6*, *Meis2*, *Pax6*, and *Nkx2-1 (Fishell & Kepecs, 2020; Lim et al., 2018)*, did not emerge from this analysis, suggesting either that they do not regulate BNST neuron identity or that intrinsic limitations to this approach prevent their detection.

After closer examination, I detected a number of computational assumptions that may prohibit accurate detection of TF regulators, which I outline here. First, cell type-specific peak-calling is, in itself, problematic, as there are several user-defined parameters that strongly bias results, including fold-change and significance thresholds, as well as a bias toward calling more marker peaks in larger clusters due to increased statistical power. Second, restricting motif analysis to cell type-specific peaks masks TFs that specify the identity of multiple neuron types within the dataset; rather than such a 1:1 relationship, it is possible, if not likely, that combinatorial binding of TFs drives neuron identity. Third, as stated above, motif enrichment analysis often fails to distinguish between TFs with highly-similar motifs, which is often the case for those belonging to the same TF family. Fourth, this approach assumes all TFs specify identity by opening chromatin, whereas it is likely that certain TFs specify identity by repressing chromatin.

To bypass these limitations, I adopted an approach incorporating the ChromVAR algorithm developed in the Greenleaf lab (Schep et al., 2017) and the simultaneous detection of TF RNA expression from the multiome dataset. First, I identified TFs whose expression predicts motif accessibility by calculating the correlation between genome-wide motif accessibility and RNA expression for each TF across all cells in the dataset (Fig. 6.1b). To then identify TFs with restricted expression across neuron types, I plotted this correlation coefficient against the maximum TF expression fold-change value between each cluster. I identified TFs scoring highly on both parameters; here I used a coefficient > 0.5 and max log2FC value in the top 50% of all TFs, although these

parameters can be adjusted for stringency. Together, this approach ranked TFs both on their potential to regulate chromatin accessibility and their expression specificity across neuron types (Ma et al., 2020).

Remarkably, this method revealed several experimentally-validated GABAergic identity TFs *a priori*, including *Lhx6*, *Meis2*, *Nkx2-1, Pax6*, and *Prox1* (Frazer et al., 2017; Miyoshi et al., 2015; Paul et al., 2017a; Quinn et al., 2007; Sandberg et al., 2016), as well as novel regulators *Egr4*, *Zfhx3*, and *Nr4a2*. While most TFs putatively maintained open chromatin (positive correlation), several TFs were predicted to act as repressors (negative correlation), including *Prox1* and *Zfhx3*, which have previously been shown to act as transcriptional repressors *in vitro* (Q. Hu et al., 2019; Takeda & Jetten, 2013). Interestingly, despite having similar motifs, ERβ/*Esr2*, but not ERα/*Esr1*, emerged as a regulator of neuron identity, specifically of the i3:Esr2 and i13:Avp neuron types (Fig. 6.1c). This finding is consistent with a non-sex-specific role for ERβ in early cortical development and neural differentiation (Fan et al., 2006; Varshney et al., 2017; L. Wang et al., 2001, 2003).

6.3.2 Nfix defines the chromatin landscape of a male-biased BNSTp neuron type

In addition to those listed above, *Nfix* was predicted to regulate the identity of the male-biased i1:Nfix neuron type (Fig. 6.1b, Fig. 6.2e). To validate this prediction, we performed Nfix CUT&RUN on the adult BNSTp, following the treatment paradigm described in Chapter 2. This experiment revealed ~33,000 consensus Nfix binding sites, which primarily contained the NFI motif, as well as a small number of E2-regulated Nfix sites (Fig. 6.2a-d). Of note, the CTCF motif was enriched within consensus Nfix sites, and a high proportion of these sites (~30%) overlapped CTCF-bound loci previously detected in the mouse forebrain using ChIP-seq (Fig. 6.2c).

By comparing Nfix CUT&RUN and multiome data, I found that Nfix-bound sites, including at the *Nfix* locus itself, are maintained in an active state primarily in i1:Nfix neurons (Fig. 6.2e). Nfix-

bound sites at the *Nfix* locus correlate in accessibility with *Nfix* expression, indicating they act as a positive feedback loop (Fig. 6.2e).

Are neuron identity regulators also important for brain masculinization? Additional examination of neonatal estradiol-responsive chromatin regions (defined in Chapter 5) showed that NE-open regions vary strongly as a function of neuron identity (see also Fig. 5.1c, 5.1e, Fig. 5.5e-f), with NE-open regions in i1:Nfix neurons preferentially containing both consensus and E2-induced Nfix binding events (Fig. 6.2f). Together, these data suggest that in addition to specifying the chromatin landscape, neuron identity TFs may diversify the response to neonatal estradiol by influencing ERα recruitment.

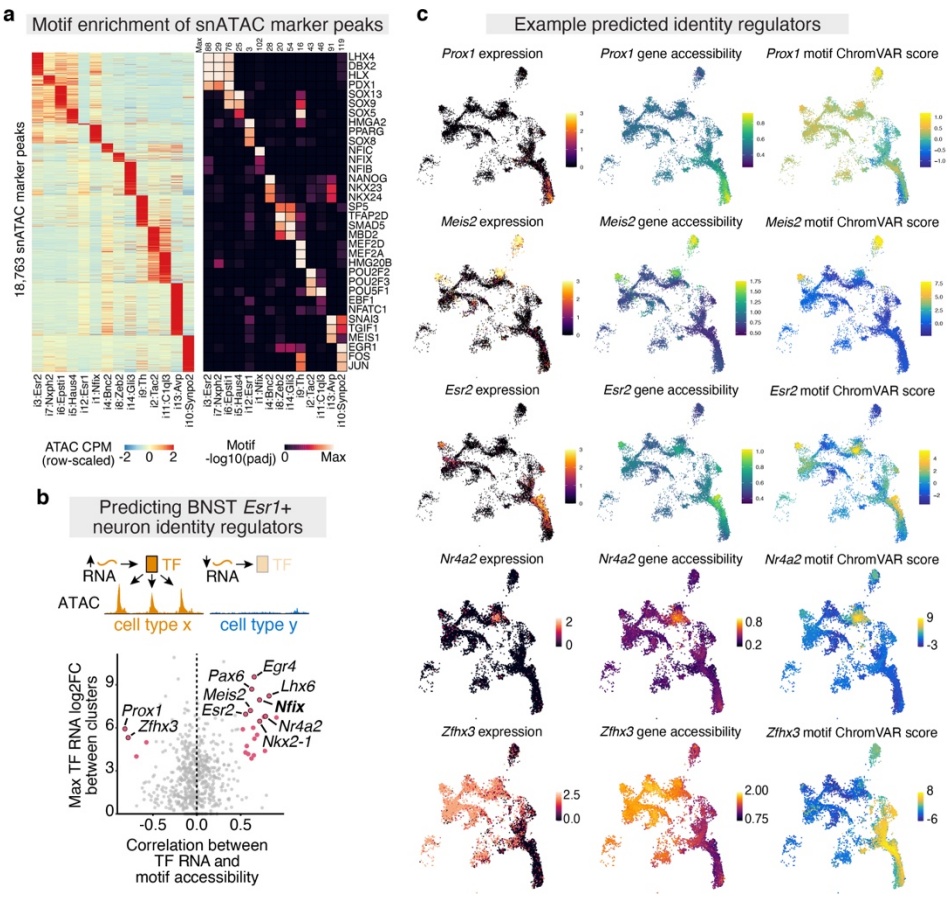

Figure. 6.1. Predicting neuron identity regulator TFs from single-nucleus multiome data. (**a**) (left) Heatmap of pseudo-bulk ATAC CPM at 18783 marker peaks for neonatal multiome clusters. (right) Top three motifs enriched in marker peaks for each multiome cluster. (**b**) (top) Schematic of a putative identity regulator TF, based on two criteria: 1) robust correlation between TF expression and genome-wide motif accessibility and 2) differential TF expression between cell type x and cell type y. (bottom) Correlation of TF expression and genome-wide motif accessibility vs. maximum inter-cluster log2FC TF expression. TFs with a correlation coefficient greater than 0.5, and in the top 50% of of log2FC expression colored in pink and classified as putative regulator TFs. (**c**) UMAPs of TF RNA expression (left), gene activity score (middle), and ChromVAR motif deviation score (right) of example putative BNST *Esr1*+ neuron identity regulators: *Prox1*, *Meis2*, *Esr2*, *Nr4a2,* and *Zfhx3*.

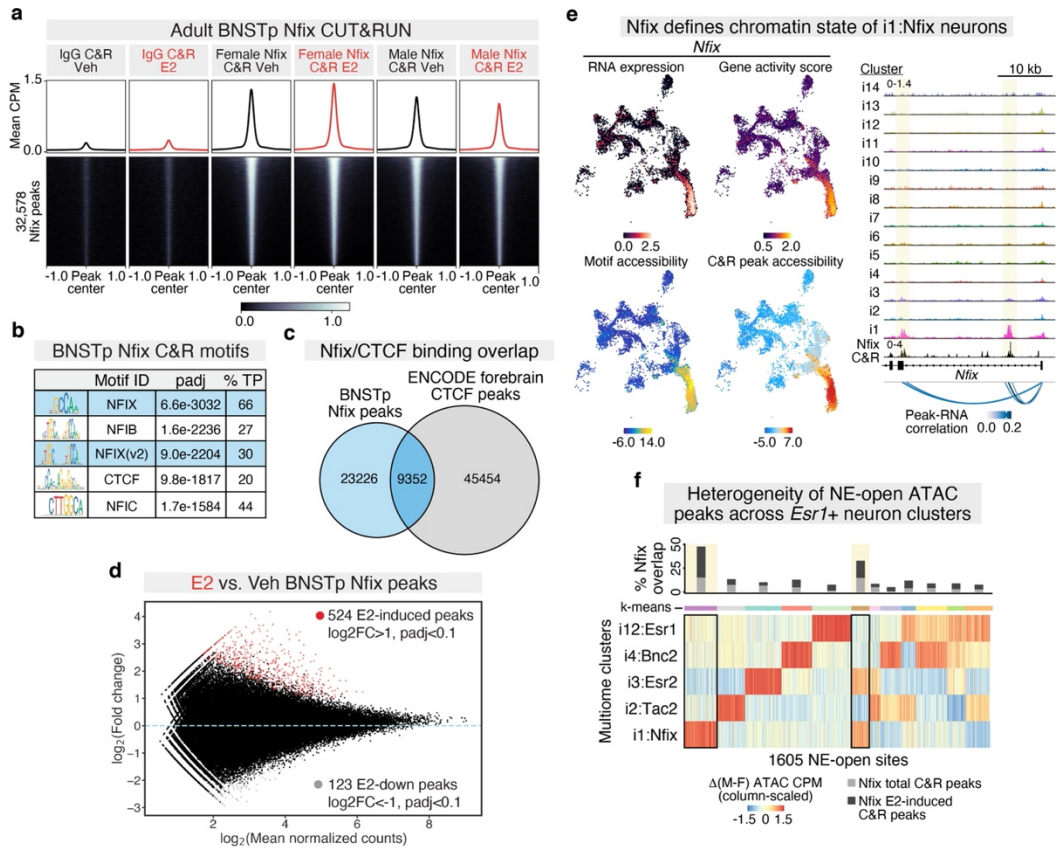

Figure 6.2. Nfix specifies the chromatin landscape of i1:Nfix neurons. (**a**) Heatmap of mean cortical IgG and BNSTp Nfix CUT&RUN CPM ±1Kb around 32,578 consensus Nfix peaks. (**b**) Top motifs enriched (AME) in 32,578 consensus BNSTp Nfix CUT&RUN peaks (MACS2, q<0.01; peaks intersected across treatment and sex). %TP=% of peaks called as positive for the indicated motif. (**c**) Overlap between consensus BNSTp Nfix CUT&RUN peaks and mouse forebrain CTCF ChIP-seq peaks, accessed from the ENCODE database. (**d**) MA plot of E2 vs. vehicle BNSTp Nfix CUT&RUN peaks (DiffBind, edgeR, padj<0.1); red dots=E2-induced peaks, grey dots=E2-down peaks. (**e**) (left) UMAPs of Nfix expression (top left), gene activity score (top right), motif chromVAR deviation score (bottom left), and CUT&RUN chromVAR deviation score (bottom right). (right) Neonatal snATAC and adult BNSTp Nfix CUT&RUN tracks at *Nfix* locus. Peak-RNA correlation indicates correlation coefficient for snATAC peaks correlated with *Nfix* expression. (**f**) Heatmap of differential snATAC CPM between males and females at 1605 NE-open sites, scaled across snRNA-seq clusters and grouped via k-means clustering. Barplot indicates % overlap for each k-means cluster with consensus and E2-induced BNSTp Nfix CUT&RUN peaks.

6.4 Discussion

Here I describe a computational strategy for identifying neuron identity regulator TFs from single-cell multiome sequencing data. I use this strategy to predict regulators of BNSTp *Esr1*+ neuron types. The BNSTp contains inhibitory neurons from the medial and caudal ganglionic eminences (MGE and CGE), as well as the embryonic POA (García-López et al., 2008; Nery et al., 2002). Consistent with this, several canonical regulators of inhibitory neurons deriving from the CGE (*Nfix*, *Prox1*) and MGE (*Lhx6*, *Nkx2-1*) were identified by this analysis (Fig. 6.1b-c). Of note, not every regulator acts as a transcriptional activator. In particular, *Prox1* expression and gene body accessibility is largely restricted to the i1:Nfix type, as opposed to the accessibility of its binding motif (Fig. 6.1c). Conversely, *Zfhx3* expression and gene accessibility occurs across most BNST *Esr1*+ neuron types, except for i1:Nfix and i3:Esr2, whereas its motif is primarily accessible within those two types (Fig. 6.1c). While not studied in cortical development, *Nr4a2*/*Nurr1*, a nuclear receptor TF, has previously been shown to be required for the formation of midbrain dopaminergic neurons; *Nr4a2*$^{-/-}$ mice lack dopaminergic neurons throughout the mesolimbic pathway and consequently lack expression of tyrosine hydroxylase (*Th*), a key enzyme in dopamine synthesis (J.-Y. Kim et al., 2003; Simeone, 2005). Consistent with these prior experiments, *Nr4a2* predicts the identity of the BNST *Esr1*+ i9:Th neuron type, which selectively expresses *Th*. This finding suggests mechanisms of neuron fate specification may be shared across brain regions and further validates the accuracy of this computational strategy.

Surprisingly, rather than influencing sex differences in gene expression within the BNSTp (see Chapter 5 Discussion), ERβ/*Esr2* is predicted to regulate the identity of i3:Esr2 and i13:Avp neuron types independently of biological sex (Fig. 6.1c). A non-sex-specific role of *Esr2* in regulating brain development is supported by findings from constitutive *Esr2*$^{-/-}$ mice, which have gross cortical morphological abnormalities prior to the neonatal surge on E18.5 (L. Wang et al., 2001) resulting from impairments in cortical neuron migration (L. Wang et al., 2003). Moreover, *Esr2* is required for neural differentiation of mouse embryonic stem cells (Varshney et al., 2017). How does *Esr2* control these

aspects of brain development? Given that aromatase KO animals do not display deficits in brain morphology (Fan et al., 2010), ERβ must act either in a ligand-independent manner or bind a ligand other than estradiol that is present across both sexes *in utero*. It has been proposed that a dihydrotestosterone metabolite, 5α-androstane-3β, 17β-diol (3β-Adiol), secreted by the immature ovaries and testes may bind ERβ in the embryonic brain, although additional data are required to validate this mechanism (Fan et al., 2010).

To further test the prediction of BNST *Esr1*+ identity regulators, we profiled Nfix genomic binding in the adult BNSTp by CUT&RUN (Fig. 6.2a-b). Because *Nfix* is restricted in expression to the i1:Nfix type among BNSTp *Esr1*+ neurons (Fig. 6.2e), I can conclude that the binding sites detected in the bulk BNSTp population derive from this neuron type. Consistent with this assumption, consensus BNSTp Nfix binding sites, including at the *Nfix* locus, are primarily accessible within i1:Nfix neurons. Nfix self-binding implies an positive feedback loop by a terminal selector TF (Ptashne, 2013); therefore, *Nfix* may maintain the i1:Nfix identity by regulating itself along with a gene battery that specifies intrinsic properties of i1:Nfix neurons.

Outside of the BNST, NFI TFs regulate cell fate specification across diverse biological contexts, including adipocytes (Waki et al., 2011), prostate epithelial cells (Grabowska et al., 2014), Müller glia and retinal bipolar interneurons (Clark et al., 2019), and hair follicle stem cells (Adam et al., 2020). NFIs are thought to control cell fate decisions by invading closed chromatin and acting as PFs (Adam et al., 2020). Given that a high proportion of Nfix peaks contain forebrain CTCF binding events (Fig. 6.2b-c), and also that NFI factors frequently occupy chromatin boundaries (Bashkirova et al., 2020; Pjanic et al., 2013), it is possible that Nfix-dependent chromatin opening regulates 3D chromatin organization, resulting in the formation of TADs specific to the i1:Nfix neuron type. Such differences in chromatin topology have recently been shown to deviate between neuron types across brain regions and may be an important mechanism for the long-term maintenance of neuron identity (Winick-Ng et al., 2021). Future gain-/loss-of-function experiments within the BNSTp will reveal how

Nfix controls the identity of the i1:Nfix type, which as shown in Chapters 4 and 5, is sexually dimorphic both in neuron number and gene expression.

6.5 Materials and Methods

Animals. All animals were maintained on a 12:12 light cycle and provided food and water *ad libitum*. All mouse experiments were performed under strict guidelines set forth by the CSHL Institutional Animal Care and Use Committee (IACUC). C57Bl6/J wildtype mice were obtained from Jackson labs. Adult male and female mice were used between 8-12 weeks of age. For adult hormone treatment experiments, animals were sacrificed for tissue collection four hours after subcutaneous administration of 5µg estradiol benzoate (E2) (Sigma E8515) suspended in corn oil (Sigma C8267) or vehicle three weeks post-gonadectomy.

Adult brain CUT&RUN. Nuclei were isolated from the BNSTp as described in Chapter 2 Materials and Methods. For Nfix CUT&RUN (1:100, Abcam ab101341), 200,000 nuclei were isolated from BNSTp of 5 animals. 400,000 cortical nuclei were used for the CUT&RUN IgG control (1:100, Antibodies-Online ABIN101961). All subsequent steps, including library preparation and sequencing, were performed as described in Chapter 2 Materials and Methods. 10 PCR cycles were used for Nfix CUT&RUN libraries.

CUT&RUN data processing. Processing of sequencing reads was performed as described for other CUT&RUN experiments in Chapter 2 Materials and Methods. To identify consensus Nfix peaks across samples, MACS2 callpeak was performed on BAM files merged across biological replicates (n=2) and subsequently intersected across treatment and sex. TF peaks that overlapped peaks called in the IgG control were removed using bedtools intersect (-v) (Quinlan & Hall, 2010) prior to downstream analysis.

CUT&RUN data analysis. Nfix CUT&RUN differential peak calling was performed with DiffBind (Stark et al., 2011). A count matrix was created from individual replicate BAM and MACS2 narrowpeak files (n=2 per condition). Consensus peaks were re-centered to ±100 bp around the point of highest read density (summits=100). edgeR (Robinson et al., 2010) was used for differential peak calling. For Nfix, differential peaks with an padj<0.1 and abs(log2FC)>1 were used for downstream analysis. Motif enrichment analysis of consensus Nfix peaks was performed with AME using the 2020 JASPAR core non-redundant vertebrate database. The overlap between consensus BNSTp Nfix peaks and conservative, IDR-thresholded ENCODE forebrain CTCF peaks (accession: ENCFF715PXI) was performed using bedtools intersect.

Single-nucleus multiome neuron identity regulator analysis.

Following peak-calling of single-nucleus multiome data (as described in Chapter 5 Materials and Methods), Signac (Stuart et al., 2021) was used to generate and store peak-by-cell count matrices for each sample. snATAC markers for each cluster were calculated (FindAllMarkers, test.use='LR', vars.to.regress= 'nCount_ATAC', min.pct=0.1, min.diff.pct=0.05, logfc.threshold=0.15). Pseudo-bulk snATAC cluster CPM was computed for each marker peak using DeepTools multiBigwigSummary and visualized with pheatmap (Fig. 6.1a). Motif enrichment analysis of snATAC marker peaks for each cluster was performed using FindMotifs. The top 3 enriched motifs per snATAC cluster are shown in Fig. 6.1a.

To perform neuron identity regulator analysis (Fig. 6.1b), TF motif annotations from the CISBP database were added to each peak (addPeakAnnotation) in ArchR, and ChromVAR deviation scores (addDeviationsMatrix), which indicate the background-subtracted, genome-wide motif accessibility signal per cell, were calculated for each cell and each motif. ChromVAR was also used to calculate per-cell deviation scores for the consensus BNSTp Nfix CUT&RUN peakset (Fig. 6.2e). Correlation

of TF RNA expression and motif deviation score was calculated for all TFs in the CISBP motif database (correlateMatrices). TFs with a correlation coefficient >0.5 and a maximum TF RNA log2FC value between each cluster in the top 50% of all TFs were classified as "neuron identity regulators" (colored pink in Fig. 6.1b).

To visualize gene activity and motif deviation scores (Fig. 6.1c), scores were imputed (imputeMatrix), transferred to the original Seurat object by cell barcode-matching, and visualized alongside TF expression using FeaturePlot.

To identify sex-biased enrichment of NE-open ATAC regions across P4 snATAC clusters (Fig. 6.2f), I first filtered out low-abundant P4 snATAC clusters (<400 nuclei), then computed the difference in ATAC CPM between males and females at NE-open loci within each cluster. Differential ATAC CPM values were scaled across clusters, and subsequently grouped via k-means clustering (k=12, iter.max=50) and visualized with pheatmap. The overlap between each k-means cluster and consensus and/or E2-induced Nfix peaks was calculated using bedtools intersect.

Chapter 7

Perspectives and Future Directions

In this thesis, I discovered mechanisms by which gonadal hormones organize and activate neural sex differences in gene regulation. I defined the first genomic targets of a gonadal hormone receptor (ERα) in the brain (Chapter 2) and revealed many targets unique to this tissue. I then described that one of these targets, *Mc4r*, drives female activity across the estrous cycle in response to estradiol (Chapter 3), providing a model for how ERα genomic regulation activates a sex-typical behavior. After establishing this methodology, I next identified gene regulatory principles underlying "activation" (Chapter 4) and "organization" (Chapter 5) of brain sex differences by gonadal hormones. I found that the neonatal hormone surge activates ERα to drive a sustained, male-typical gene regulatory program in the developing brain, comprised of both sex-biased enhancers and genes. A subset of this program persists as sex-biased into adulthood, yet many additional sex-biased loci and genes emerge following puberty as a result of gonadal hormone production. Lastly, I utilized the wealth of single-cell sequencing data generated in these experiments to predict TFs that regulate the identity of BNST *Esr1*+ neuron types and further validated identity regulator function of Nfix in the male-biased i1:Nfix neuron type (Chapter 6). In the Appendix, I describe a novel approach for high-throughput, single-cell RNA sequencing that offers improvements in cost and flexibility compared to existing commercial options. The findings described here provide the first step toward understanding how gonadal hormone receptors program the display of innate behaviors. In this section, I outline a number of open questions, which I hope inspire others to continue this research.

7.1 Hormones, gene regulation, and the brain

In this thesis, I have focused on how gonadal hormones, principally estradiol acting through ERα and testosterone acting through AR, generate neural sex differences under basal conditions. However, there are many other hormones secreted by peripheral organs, and many other critical windows of hormone signaling, which remain unexplored. For example, during proestrus and prior to parturition, the ovaries also secrete progesterone (Murr et al., 1974), which binds its cognate receptor PR. Similar to ERα, regulation of gene expression by PR within specific populations of neurons may coordinate the display of appropriate behaviors, such as lordosis and maternal aggression, with internal physiological state. Consistent with this hypothesis, different populations of VMHvl ERα+/PR+ neurons control the display of female lordosis and maternal aggression behaviors (Hashikawa et al., 2017; Liu et al., 2022). Of note, PR and AR share a near-identical DBD and recognition motif (Hill et al., 2012); therefore, one tempting speculation is that progesterone and testosterone drive aggression across females and males, respectively, by converging upon a similar set of genomic targets within the VMHvl. In addition to progesterone, the neural mode of action for other estrogens, such as estriol, which is produced in high quantities by the placenta and released into maternal circulation (Tulchinsky et al., 1972), remains poorly understood. Prolactin, another hormone released from the pituitary gland during proestrus, early pregnancy, and lactation, regulates maternal care, food intake, and stress response by acting on prolactin receptors expressed throughout the social behavior network (Grattan, 2002; Kokay et al., 2018; Phillipps et al., 2020). While prolactin receptor activation is known to stimulate STAT5 phosphorylation within this network (Furigo et al., 2016; Gustafson et al., 2020), the neural gene regulatory mechanisms of pSTAT5, and their contribution to maternal behaviors, also remain unexplored.

Another critical window of hormone signaling – menopause – was not directly examined in this thesis. The decline in ovarian estradiol production during this period can lead to several physiological changes, such as metabolic decline, sleep disturbance, mood swings, and memory lapses

(de Boer et al., 2018; Searles et al., 2018). Men also undergo a more gradual decline in androgen production during aging, termed “andropause”. Across both sexes, hormone replacement therapy has been and is still under investigation as an approach to improve metabolism, mood, and cognitive function during these life stages (Bassil et al., 2009; Sherwin & Gelfand, 1985). Understanding how gradual reduction in hormone levels influences hormone receptor-dependent gene regulation during aging, and how changes at the genomic level influence mood and cognition, will be critical for developing strategies to mitigate these negative effects.

One current limitation of hormone replacement therapy is the non-specific targeting of tissues throughout the body, resulting in potentially elevated risk of prostate and breast cancer (Beral & Million Women Study Collaborators, 2003). A similar phenomenon exists for the treatment of cancer itself: selective estrogen receptor modulators (SERMs), such as raloxifene and tamoxifen, and aromatase inhibitors, such as letrozole, are regularly used to treat and prevent relapse of ER positive breast cancer, yet these drugs also act on the brain (M. A. Arevalo et al., 2011; M.-A. Arevalo et al., 2015). While tamoxifen is associated with an increased risk of cognitive impairment and depression, raloxifene appears to have beneficial effects on brain function (de Boer et al., 2018) and is currently in a phase 3 clinical trial for the treatment of schizophrenia. To date, the mode of action for SERMs in the brain remains unclear. Understanding how these drugs influence ERα genomic binding and cofactor recruitment may explain their effects on cognition and mood and lead to the development of hormonal drugs with tissue-specific activity.

Lastly, comparing and contrasting gonadal hormone receptor gene regulatory mechanisms across species will be critical to understanding the immense variety of sex-typical behaviors displayed throughout the animal kingdom. In the era of single-cell sequencing, CRISPR-based genetic tools, and genome assembly with long-read sequencing, addressing this question is now more possible than ever before.

7.2 Epigenetics of brain sex differences

To study chromatin sex differences in the brain, I have extensively used ATAC-seq, as it provides a useful proxy for epigenomic modifications. Chromatin accessibility depends on nucleosome turnover, which is controlled by TF binding, histone PTMs and DNA methylation (Klemm et al., 2019). However, to better understand the epigenetic programming of brain sexual differentiation, it will be critical to study how hormone receptors impart specific epigenomic modifications in neurons. As discussed in Chapter 1, hormone receptors physically modify the genome upon binding by recruiting a suite of coactivator and corepressor proteins, which influence histone PTMs, DNA methylation, and nucleosome positioning (Perissi & Rosenfeld, 2005). To date, the protein binding partners of gonadal hormone receptors have not been identified in the brain, nor in most other tissues in which they are expressed. Developing genetic tools for co-immunoprecipitation mass spectrometry or proximity biotinylation labeling, such as FLAG-tagged ERα and/or APEX2-ERα knock-in mouse lines, would address this question and provide a critical resource for the field of neuroendocrinology.

Along these lines, it is not only important to understand how ERα imparts epigenomic modifications in the brain but to understand how such modifications persist throughout brain development, as it may reveal fundamental principles of how post-mitotic cells encode transient signals in the genome. In mice, testosterone falls to undetectable levels in serum within hours after birth (Clarkson & Herbison, 2016), suggesting ERα does not remain bound at loci that persist as male-biased throughout the neonatal critical window. Instead, it is possible that neonatal ERα activation initiates chromatin opening, which provides access for other, constitutively-expressed TFs to bind and recruit cofactors that maintain the epigenome. In line with this hypothesis, a high proportion of neonatal ERα-bound sites in the i1:Nfix neuron type were frequently co-bound by Nfix in response to estradiol (Chapter 6). This finding suggests that identity regulator TFs may play a role in coordinating sex differences in gene regulation that vary across neuron types. Future genetic manipulation experiments

will reveal whether Nfix, or other identity regulators, are required for maintaining sex-biased chromatin states in a neuron type-specific manner.

Another possibility is that neonatal ERα activation drives lasting changes in gene regulation at the level of 3D chromatin organization. Similar to other transcription factors, gonadal hormone receptors regulate gene expression by looping distal enhancer elements and associated regulatory machinery to target gene promoters, resulting in the formation of TADs (Fullwood et al., 2009; Hsieh et al., 2014; Hsu et al., 2010; Le Dily et al., 2019; Le Dily & Beato, 2018; Ross-Innes et al., 2011; Q. Wang et al., 2005). Such large-scale rearrangements of chromatin can coordinately activate or repress a high number of genes (Jubb et al., 2017; Rafique et al., 2015). E2 has previously been shown to rapidly affect nuclear organization, primarily reducing nuclear envelope invaginations and clumped heterochromatin speckles, in a pioneering electron microscopy study of the VMH (Jones et al., 1985). Measuring neonatal sex differences in chromatin structure using chromosome conformation capture methods, such as Hi-C, and comparison to sex-biased chromatin loci, will reveal whether neonatal ERα activation leads to long-lasting changes to the 3D genome.

Why do persistent sex differences in the epigenome occur? As discussed in Chapter 1, the male-bias in BNSTp neuron survival and axon outgrowth occurs between P4 and P7 (G. Gu et al., 2003; Hutton et al., 1998; Ibanez et al., 2001; M. V. Wu et al., 2009a), long after neonatal hormone subsides from the serum (Clarkson & Herbison, 2016). Throughout the brain, the first two weeks of postnatal life constitute a critical period of development, during which time neurons undergo apoptosis, axonogenesis, synaptogenesis, myelination, and synaptic pruning (Marín, 2016). Therefore, it is possible that maintaining sex-biased gene expression across this prolonged developmental window is critical for ensuring male-typical BNSTp circuit formation. Consistent with this hypothesis, the majority of genes persisting as sex-biased between P4 and P14 regulate axon outgrowth (*Nell2*, *Pak7*/PAK5), synapse formation (*Col25a1*, *Il1rap*), and neurotransmission (*Adra1a*, *Asic2*, *Kcnab1*, *Scg2*). A small number of transcriptional regulators (*Greb1*, *Klf12*, *Plagl1*, *Sox5*) also persist as sex-

biased, which may be important for maintaining sex differences in chromatin state. Another possibility is that a persistent sex difference in gene regulation is required to prime a genomic or upstream signaling response to gonadal hormones released at puberty. In line with this prediction, neonatal and adult sex-biased loci occur in close proximity along the linear genome at many of the same sex-biased genes (Fig. 5.3). Therefore, it is possible that neonatal ERα recruitment influences adult AR binding site selection by controlling accessible chromatin domains via loop extrusion (Le Dily et al., 2019).

7.3 Controlling epigenetic programs of brain development

The logical next step after observing an epigenetic program of brain sexual differentiation is to perturb and ask whether it is required for sexual differentiation of the BNSTp and, ultimately, the display of sex-typical behaviors. With CRISPRa and CRISPRi tools, it is now feasible to directly manipulate the activity of enhancers, either singly or combinatorially, in specific regions of the brain and assess their involvement in behavioral display (Matharu et al., 2019; Savell et al., 2020). For instance, as shown in Chapter 3 and previously (Matharu et al., 2019), hypothalamic co-delivery of AAVs expressing dCas9-VP64 and sgRNAs targeting either a *Sim1* enhancer or *Mc4r* promoter increases physical activity.

To faithfully manipulate the epigenetic program identified here, it will be necessary to target CRISPRa/i to the same neuron types in which this program occurs. Currently, Cre-dependent CRISPRa/i lentiviral constructs have been described (Carullo et al., 2021), although they have not yet been applied in the brain. Cre-dependent CRISPRa/i mouse lines have also recently been developed (Gemberling et al., 2021; H. Zhou et al., 2018); however, these prior studies delivered Cre via AAV injection, rather than targeting a particular cell type by crossing animals with a Cre-driver line. As a proof-of-principle experiment to test the feasibility of CRISPRa/i enhancer manipulation on the day of birth, I have developed sgRNAs targeting a neonatal sex-biased enhancer at the *Greb1* locus (Fig. 7.1a) and validated that they upregulate *Greb1* expression *in vitro* upon co-nucleofection with dCas9-VPR (Fig. 7.1b). Likewise, performing CRISPRa with a previously published sgRNA targeting the *Sim1*

promoter (Matharu et al., 2019) robustly increased *Sim1* expression (Fig. 7.1b). In parallel, I have generated lentiviruses expressing these sgRNAs, as well as a non-targeting control (*lacZ* promoter), and optimized stereotactic targeting of the BNST on P0 (Fig. 7.1c). Currently, I am testing whether this approach stably induces *Greb1* expression in the neonatal BNST by *in situ* hybridization.

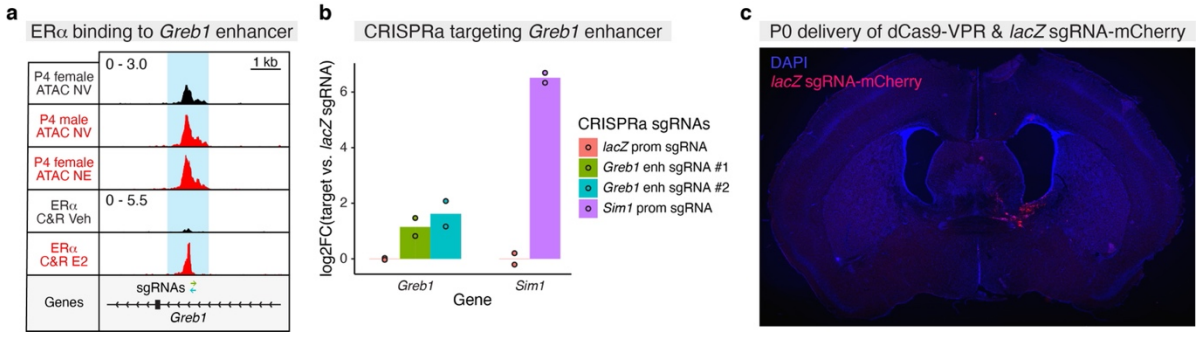

Figure 7.1. CRISPRa enhancer activation and stereotactic delivery into P0 BNST. (**a**) Neonatal ERα binding and male-biased chromatin accessibility at an intronic enhancer of the male-biased gene, *Greb1*. Green and blue arrows indicate position of CRISPRa sgRNAs on the antisense (*Greb1* enh sgRNA #1) and sense (*Greb1* enh sgRNA #2) strands, respectively. (**b**) qPCR analysis of *Greb1* expression following co-nucleofection of dCas9-VPR and sgRNAs targeting the *lacZ* promoter (non-targeting control), *Greb1* enhancer, or the *Sim1* promoter in mHypoA cells. (**c**) Coronal section of a P14 mouse co-injected with dCas9-VPR and *lacZ* sgRNA-mCherry lentiviruses on P0.

7.4 Future advancements in low-input genomics

The experiments described here utilized novel approaches for studying TF binding in low numbers of cells and chromatin accessibility in single cells. It is possible that future advancements in these technologies may reveal further insights into the heterogeneity of gene regulation by gonadal hormone receptors across neuron types and developmental time.

For instance, CUT&Tag has recently been optimized to profile histone PTMs within single cells at high-throughput (Bartosovic et al., 2021; S. J. Wu et al., 2021), to simultaneously measure histone PTMs and RNA (C. Zhu et al., 2021), and to detect multiple PTMs within the same cell (Bartosovic & Castelo-Branco, 2022; Janssens et al., 2021). These advancements raise the possibility of examining whether neonatal ERα activation imparts specific histone PTMs, such as H3K27ac and/or H3K4me1, and whether such modifications are involved in maintaining sex differences in gene regulation. However, CUT&Tag currently suffers from a few technical limitations, prohibiting its widespread use. For instance, unlike the pA-MNase used in CUT&RUN, pA-Tn5, like unconjugated Tn5, has preferential affinity for accessible chromatin, necessitating the inclusion of a high salt wash (300 mM) to mitigate non-specific binding and tagmentation (Kaya-Okur et al., 2019). Despite increasing the salt concentration, residual pA-Tn5 still tagments accessible regions of the genome independently of the targeted histone PTM, obscuring true from false positive peaks (Kaya-Okur et al., 2019; M. Wang & Zhang, 2021). Moreover, several of the early implementations of single-cell CUT&Tag (scCUT&Tag) using the 10X Genomics platform detected few unique fragments per cell relative to snATAC-seq (~400-500 vs. ~30,000) and failed to resolve individual neuron types (Bartosovic et al., 2021). Lastly, as scCUT&Tag exchanges wash buffers by centrifuging nuclei, rather than immobilizing on magnetic beads, there is major sample loss prior to single nucleus droplet capture. Therefore, a high amount of input material is required (50,000 to 1 million nuclei, depending on the protocol), which prohibits its application to small numbers of cells, such as genetically defined neuron populations.

If these technical obstacles are overcome, scCUT&Tag may be an invaluable tool for characterizing gene regulatory mechanisms in the brain. One possibility is that improving Tn5 transposase itself may reduce open chromatin bias and improve efficiency of fragment detection. For instance, protein engineering of Tn5 transposase, or another member of the transposase family, may result in an enzyme with reduced affinity for open chromatin but preserved tagmentation function when targeted to a histone PTM by protein A. This idea is supported by the observation that the version of the Tn5 enzyme used in CUT&Tag and ATAC-seq was originally engineered from the inactive, wildtype enzyme to have high activity on accessible chromatin, suggesting low-affinity enzymes may also be derived (Reznikoff, 2003). Moreover, implementing a deterministic tagmentation by loading pA-Tn5 with only the P5 adaptor and subsequently introducing the P7 adaptor with a second round of tagmentation post-library amplification has been shown to substantially improve fragment recovery in scCUT&Tag (Bartosovic & Castelo-Branco, 2022). A pA-Tn5 with minimal non-specific targeting may also enable high-throughput, single-cell profiling of TF binding, as the 300 mM salt wash used in CUT&Tag strips most TFs off of chromatin, preventing their detection. Such an improvement would undoubtedly revolutionize our understanding of how TFs specify neuron identity and coordinate stimulus-dependent transcription.

7.5 Conclusion

In this thesis, I used low-input and single-cell genomics to define the molecular principles of the "Organizational-Activational Hypothesis". I revealed that neonatal ERα activation initiates an epigenetic program of brain sexual differentiation, which undergoes heterogeneous regulation across BNSTp neuron types. While a small subset of genes and enhancers persists as sex-biased throughout life, gonadal hormones released following puberty drive the majority of adult sex differences in gene regulation. Gonadectomy in adulthood attenuated sex differences in chromatin state, and under these conditions, *Esr1*+ neurons of both sexes responded similarly to estradiol at the genomic level. Together,

these findings demonstrate that while early-life gonadal hormone signaling organizes irreversible sex differences in BNSTp circuitry, the genome remains responsive to the activational influence of the hormonal milieu, consistent with the observation that both sexes retain the potential to display behaviors typical of the opposite sex (Edwards & Burge, 1971; Hashikawa et al., 2017; Kimchi et al., 2007; Wei et al., 2018). Therefore, rather than restricting or canalizing phenotypic endpoints, gonadal hormones and their receptors may have evolved to enable dynamic, sexually differentiated states.

I hope this thesis inspires future scientists to study the nature of the epigenetic program described herein and its involvement in innate behaviors.

Bibliography

Adam, R. C., Yang, H., Ge, Y., Infarinato, N. R., Gur-Cohen, S., Miao, Y., Wang, P., Zhao, Y., Lu, C. P., Kim, J. E., Ko, J. Y., Paik, S. S., Gronostajski, R. M., Kim, J., Krueger, J. G., Zheng, D., & Fuchs, E. (2020). NFI transcription factors provide chromatin access to maintain stem cell identity while preventing unintended lineage fate choices. *Nature Cell Biology*, *22*(6), 640–650.

Ahrens, S., Wu, M. V., Furlan, A., Hwang, G.-R., Paik, R., Li, H., Penzo, M. A., Tollkuhn, J., & Li, B. (2018). A Central Extended Amygdala Circuit That Modulates Anxiety. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(24), 5567–5583.

Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.-C., Geurts, P., Aerts, J., van den Oord, J., Atak, Z. K., Wouters, J., & Aerts, S. (2017). SCENIC: single-cell regulatory network inference and clustering. *Nature Methods*, *14*(11), 1083–1086.

Allaway, K. C., Gabitto, M. I., Wapinski, O., Saldi, G., Wang, C.-Y., Bandler, R. C., Wu, S. J., Bonneau, R., & Fishell, G. (2021). Genetic and epigenetic coordination of cortical interneuron development. *Nature*, *597*(7878), 693–697.

Allen, L. S., & Gorski, R. A. (1990a). Sex difference in the bed nucleus of the stria terminalis of the human brain. *The Journal of Comparative Neurology*, *302*, 697–706.

Allen, L. S., & Gorski, R. A. (1990b). Sex difference in the bed nucleus of the stria terminalis of the human brain. *The Journal of Comparative Neurology*, *302*(4), 697–706.

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* , *31*(2), 166–169.

Apostolakis, E. M., Ramamurphy, M., Zhou, D., Oñate, S., & O'Malley, B. W. (2002). Acute disruption of select steroid receptor coactivators prevents reproductive behavior in rats and unmasks genetic adaptation in knockout mice. *Molecular Endocrinology* , *16*(7), 1511–1523.

Arevalo, M.-A., Azcoitia, I., & Garcia-Segura, L. M. (2015). The neuroprotective actions of oestradiol and oestrogen receptors. *Nature Reviews. Neuroscience*, *16*(1), 17–29.

Arevalo, M. A., Santos-Galindo, M., Lagunas, N., Azcoitia, I., & Garcia-Segura, L. M. (2011). Selective estrogen receptor modulators as brain therapeutic agents. *Journal of Molecular Endocrinology*, *46*(1), R1–R9.

Auger, A. P., Tetel, M. J., & McCarthy, M. M. (2000). Steroid receptor coactivator-1 (SRC-1) mediates the development of sex-specific brain morphology and behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(13), 7551–7555.

Azcoitia, I., Yague, J. G., & Garcia-Segura, L. M. (2011). Estradiol synthesis within the human brain. *Neuroscience*, *191*, 139–147.

Bailey, T. L. (2011). DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* , *27*(12), 1653–1659.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., & Noble, W. S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, *37*(Web Server issue), W202–W208.

Balthasar, N., Dalgaard, L. T., Lee, C. E., Yu, J., Funahashi, H., Williams, T., Ferreira, M., Tang, V., McGovern, R. A., Kenny, C. D., Christiansen, L. M., Edelstein, E., Choi, B., Boss, O., Aschkenasi, C., Zhang, C.-Y., Mountjoy, K., Kishi, T., Elmquist, J. K., & Lowell, B. B. (2005). Divergence of melanocortin pathways in the control of food intake and energy expenditure. *Cell*, *123*(3), 493–505.

Balthazart, J., & Ball, G. F. (1998). New insights into the regulation and function of brain estrogen synthase (aromatase). *Trends in Neurosciences*, *21*(6), 243–249.

Bangasser, D. A., & Wicks, B. (2017). Sex-specific mechanisms for responding to stress. *Journal of Neuroscience Research*, *95*(1-2), 75–82.

Bao, A.-M., & Swaab, D. F. (2011). Sexual differentiation of the human brain: relation to gender

identity, sexual orientation and neuropsychiatric disorders. *Frontiers in Neuroendocrinology*, *32*(2), 214–226.

Bartosovic, M., & Castelo-Branco, G. (2022). Multimodal chromatin profiling using nanobody-based single-cell CUT&Tag. In *bioRxiv* (p. 2022.03.08.483459). https://doi.org/10.1101/2022.03.08.483459

Bartosovic, M., Kabbe, M., & Castelo-Branco, G. (2021). Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nature Biotechnology*, *39*(7), 825–835.

Bashkirova, E., Monahan, K., Campbell, C. E., Osinski, J. M., Tan, L., Schieren, I., Barnea, G., Sunnie Xie, X., Gronostajski, R. M., & Lomvardas, S. (2020). Homeotic Regulation of Olfactory Receptor Choice via NFI-dependent Heterochromatic Silencing and Genomic Compartmentalization. In *Cold Spring Harbor Laboratory* (p. 2020.08.30.274035). https://doi.org/10.1101/2020.08.30.274035

Bassil, N., Alkaade, S., & Morley, J. E. (2009). The benefits and risks of testosterone replacement therapy: a review. *Therapeutics and Clinical Risk Management*, *5*(3), 427–448.

Beach, F. A. (1981). Historical origins of modern research on hormones and behavior. *Hormones and Behavior*, *15*(4), 325–376.

Becker, J. B. (2009). Sexual differentiation of motivation: a novel mechanism? *Hormones and Behavior*, *55*(5), 646–654.

Becker, J. B., & Chartoff, E. (2019). Sex differences in neural mechanisms mediating reward and addiction. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *44*(1), 166–183.

Bennesch, M. A., & Picard, D. (2015). Minireview: Tipping the balance: ligand-independent activation of steroid receptors. *Molecular Endocrinology* , *29*(3), 349–363.

Beral, V., & Million Women Study Collaborators. (2003). Breast cancer and hormone-replacement

therapy in the Million Women Study. *The Lancet*, *362*(9382), 419–427.

Bergan, J. F., Ben-Shaul, Y., & Dulac, C. (2014). Sex-specific processing of social cues in the medial amygdala. *eLife*, *3*, e02743.

Blurton-Jones, M., & Tuszynski, M. H. (2002). Estrogen receptor-beta colocalizes extensively with parvalbumin-labeled inhibitory neurons in the cortex, amygdala, basal forebrain, and hippocampal formation of intact and ovariectomized adult rats. *The Journal of Comparative Neurology*, *452*(3), 276–287.

Boulware, M. I., Weick, J. P., Becklund, B. R., Kuo, S. P., Groth, R. D., & Mermelstein, P. G. (2005). Estradiol activates group I and II metabotropic glutamate receptor signaling, leading to opposing influences on cAMP response element-binding protein. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *25*(20), 5066–5078.

Brainard, M. S., & Doupe, A. J. (2002). What songbirds teach us about learning. *Nature*, *417*(6886), 351–358.

BRAIN Initiative Cell Census Network (BICCN). (2021). A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature*, *598*(7879), 86–102.

Broome, R., Chernukhin, I., Jamieson, S., Kishore, K., Papachristou, E. K., Mao, S.-Q., Tejedo, C. G., Mahtey, A., Theodorou, V., Groen, A. J., D'Santos, C., Balasubramanian, S., Farcas, A. M., Siersbæk, R., & Carroll, J. S. (2021). TET2 is a component of the estrogen receptor complex and controls 5mC to 5hmC conversion at estrogen receptor cis-regulatory regions. *Cell Reports*, *34*(8), 108776.

Brzozowski, A. M., Pike, A. C., Dauter, Z., Hubbard, R. E., Bonn, T., Engström, O., Ohman, L., Greene, G. L., Gustafsson, J. A., & Carlquist, M. (1997). Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature*, *389*(6652), 753–758.

Cai, D., Cohen, K. B., Luo, T., Lichtman, J. W., & Sanes, J. R. (2013). Improved tools for the Brainbow toolbox. *Nature Methods*, *10*(6), 540–547.

Carroll, J. S. (2016). Mechanisms of oestrogen receptor (ER) gene regulation in breast cancer. *European Journal of Endocrinology / European Federation of Endocrine Societies*, *175*(1), R41–R49.

Carroll, J. S., Liu, X. S., Brodsky, A. S., Li, W., Meyer, C. A., Szary, A. J., Eeckhoute, J., Shao, W., Hestermann, E. V., Geistlinger, T. R., Fox, E. A., Silver, P. A., & Brown, M. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*, *122*(1), 33–43.

Carullo, N. V. N., Hinds, J. E., Revanna, J. S., Tuscher, J. J., Bauman, A. J., & Day, J. J. (2021). A Cre-Dependent CRISPR/dCas9 System for Gene Expression Regulation in Neurons. *eNeuro*, *8*(4). https://doi.org/10.1523/ENEURO.0188-21.2021

Clark, B. S., Stein-O'Brien, G. L., Shiau, F., Cannon, G. H., Davis-Marcisak, E., Sherman, T., Santiago, C. P., Hoang, T. V., Rajaii, F., James-Esposito, R. E., Gronostajski, R. M., Fertig, E. J., Goff, L. A., & Blackshaw, S. (2019). Single-Cell RNA-Seq Analysis of Retinal Development Identifies NFI Factors as Regulating Mitotic Exit and Late-Born Cell Specification. *Neuron*, *102*(6), 1111–1126.e5.

Clarkson, J., & Herbison, A. E. (2016). Hypothalamic control of the male neonatal testosterone surge. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1688), 20150115.

Clemens, A. M., Lenschow, C., Beed, P., Li, L., Sammons, R., Naumann, R. K., Wang, H., Schmitz, D., & Brecht, M. (2019). Estrus-Cycle Regulation of Cortical Inhibition. *Current Biology: CB*, *29*, 1–11.

Conradt, B., & Horvitz, H. R. (1999). The TRA-1A sex determination protein of C. elegans regulates sexually dimorphic cell deaths by repressing the egl-1 cell death activator gene. *Cell*, *98*(3), 317–327.

Cooke, B. M., & Simerly, R. B. (2005). Ontogeny of bidirectional connections between the medial

nucleus of the amygdala and the principal bed nucleus of the stria terminalis in the rat. *The Journal of Comparative Neurology*, *489*(1), 42–58.

Cooper, S. J., & Dourish, C. T. (1990). Multiple cholecystokinin (CCK) receptors and CCK-monoamine interactions are instrumental in the control of feeding. *Physiology & Behavior*, *48*(6), 849–857.

Corces, M. R., Trevino, A. E., Hamilton, E. G., Greenside, P. G., Sinnott-Armstrong, N. A., Vesuna, S., Satpathy, A. T., Rubin, A. J., Montine, K. S., Wu, B., Kathiria, A., Cho, S. W., Mumbach, M. R., Carter, A. C., Kasowski, M., Orloff, L. A., Risca, V. I., Kundaje, A., Khavari, P. A., … Chang, H. Y. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature Methods*, *14*(10), 959–962.

Correa, S. M., Newstrom, D. W., Warne, J. P., Flandin, P., Cheung, C. C., Lin-Moore, A. T., Pierce, A. A., Xu, A. W., Rubenstein, J. L., & Ingraham, H. A. (2015). An estrogen-responsive module in the ventromedial hypothalamus selectively drives sex-specific activity in females. *Cell Reports*, *10*(1), 62–74.

Cowley, M. A., Pronchuk, N., Fan, W., Dinulescu, D. M., Colmers, W. F., & Cone, R. D. (1999). Integration of NPY, AGRP, and melanocortin signals in the hypothalamic paraventricular nucleus: evidence of a cellular basis for the adipostat. *Neuron*, *24*(1), 155–163.

Cowley, M. A., Smart, J. L., Rubinstein, M., Cerdán, M. G., Diano, S., Horvath, T. L., Cone, R. D., & Low, M. J. (2001). Leptin activates anorexigenic POMC neurons through a neural network in the arcuate nucleus. *Nature*, *411*(6836), 480–484.

Crow, M., Paul, A., Ballouz, S., Huang, Z. J., & Gillis, J. (2018). Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nature Communications*, *9*(1), 884.

de Boer, J., Prikken, M., Lei, W. U., Begemann, M., & Sommer, I. (2018). The effect of raloxifene augmentation in men and women with a schizophrenia spectrum disorder: a systematic review

and meta-analysis. *NPJ Schizophrenia*, *4*(1), 1.

Delevich, K., Thomas, A. W., & Wilbrecht, L. (2019). Adolescence and “Late Blooming” Synapses of the Prefrontal Cortex. *Cold Spring Harbor Symposia on Quantitative Biology*. https://doi.org/10.1101/sqb.2018.83.037507

Della Torre, S., & Maggi, A. (2017). Sex Differences: A Resultant of an Evolutionary Pressure? *Cell Metabolism*, *25*(3), 499–505.

Dhillon, H., Zigman, J. M., Ye, C., Lee, C. E., McGovern, R. A., Tang, V., Kenny, C. D., Christiansen, L. M., White, R. D., Edelstein, E. A., Coppari, R., Balthasar, N., Cowley, M. A., Chua, S., Jr, Elmquist, J. K., & Lowell, B. B. (2006). Leptin directly activates SF1 neurons in the VMH, and this action by leptin is required for normal body-weight homeostasis. *Neuron*, *49*(2), 191–203.

Di Bella, D. J., Habibi, E., Stickels, R. R., Scalia, G., Brown, J., Yadollahpour, P., Yang, S. M., Abbate, C., Biancalani, T., Macosko, E. Z., Chen, F., Regev, A., & Arlotta, P. (2021). Molecular logic of cellular diversification in the mouse cerebral cortex. *Nature*, *595*(7868), 554–559.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* , *29*(1), 15–21.

DonCarlos, L. L., & Handa, R. J. (1994). Developmental profile of estrogen receptor mRNA in the preoptic area of male and female neonatal rats. *Brain Research. Developmental Brain Research*, *79*(2), 283–289.

DonCarlos, L. L., McAbee, M., Ramer-Quinn, D. S., & Stancik, D. M. (1995). Estrogen receptor mRNA levels in the preoptic area of neonatal rats are responsive to hormone manipulation. *Brain Research. Developmental Brain Research*, *84*(2), 253–260.

Droog, M., Mensink, M., & Zwart, W. (2016). The Estrogen Receptor α-Cistrome Beyond Breast Cancer. *Molecular Endocrinology* , *30*(10), 1046–1058.

Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, *4*(8), 1184–1191.

Edelmann, M. N., & Auger, A. P. (2011). Epigenetic impact of simulated maternal grooming on estrogen receptor alpha within the developing amygdala. *Brain, Behavior, and Immunity*, *25*(7), 1299–1304.

Edwards, D. A., & Burge, K. G. (1971). Early androgen treatment and male and female sexual behavior in mice. *Hormones and Behavior*, *2*(1), 49–58.

Evans, R. M. (1988). The steroid and thyroid hormone receptor superfamily. *Science*, *240*(4854), 889–895.

Fan, X., Warner, M., & Gustafsson, J.-A. (2006). Estrogen receptor beta expression in the embryonic brain regulates development of calretinin-immunoreactive GABAergic interneurons. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(51), 19338–19343.

Fan, X., Xu, H., Warner, M., & Gustafsson, J.-A. (2010). ERbeta in CNS: new roles in development and function. *Progress in Brain Research*, *181*, 233–250.

Farooqi, I. S., Yeo, G. S., Keogh, J. M., Aminian, S., Jebb, S. A., Butler, G., Cheetham, T., & O'Rahilly, S. (2000). Dominant and recessive inheritance of morbid obesity associated with melanocortin 4 receptor deficiency. *The Journal of Clinical Investigation*, *106*(2), 271–279.

Favuzzi, E., Deogracias, R., Marques-Smith, A., Maeso, P., Jezequel, J., Exposito-Alonso, D., Balia, M., Kroon, T., Hinojosa, A. J., F Maraver, E., & Rico, B. (2019). Distinct molecular programs regulate synapse specificity in cortical inhibitory circuits. *Science*, *363*(6425), 413–417.

Feng, Y., Manka, D., Wagner, K.-U., & Khan, S. A. (2007). Estrogen receptor-alpha expression in the mammary epithelium is required for ductal and alveolar morphogenesis in mice. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(37), 14718–14723.

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A. K., Slichter, C. K., Miller, H. W., McElrath, M. J., Prlic, M., Linsley, P. S., & Gottardo, R. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biology*, *16*, 278.

Fishell, G., & Kepecs, A. (2020). Interneuron Types as Attractors and Controllers. *Annual Review of Neuroscience*, *43*, 1–30.

Fix, C., Jordan, C., Cano, P., & Walker, W. H. (2004). Testosterone activates mitogen-activated protein kinase and the cAMP response element binding protein transcription factor in Sertoli cells. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(30), 10919–10924.

Forger, N. G., Hodges, L. L., Roberts, S. L., & Breedlove, S. M. (1992). Regulation of motoneuron death in the spinal nucleus of the bulbocavernosus. *Journal of Neurobiology*, *23*(9), 1192–1203.

Forger, N. G., Rosen, G. J., Waters, E. M., Jacob, D., Simerly, R. B., & de Vries, G. J. (2004a). Deletion of Bax eliminates sex differences in the mouse forebrain. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(37), 13666–13671.

Forger, N. G., Rosen, G. J., Waters, E. M., Jacob, D., Simerly, R. B., & de Vries, G. J. (2004b). Deletion of Bax eliminates sex differences in the mouse forebrain. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(37), 13666–13671.

Forger, N. G., Strahan, J. A., & Castillo-Ruiz, A. (2016). Cellular and molecular mechanisms of sexual differentiation in the mammalian nervous system. *Frontiers in Neuroendocrinology*, *40*, 67–86.

Foulds, C. E., Feng, Q., Ding, C., Bailey, S., Hunsaker, T. L., Malovannaya, A., Hamilton, R. A., Gates, L. A., Zhang, Z., Li, C., Chan, D., Bajaj, A., Callaway, C. G., Edwards, D. P., Lonard, D. M., Tsai, S. Y., Tsai, M.-J., Qin, J., & O'Malley, B. W. (2013). Proteomic analysis of coregulators bound to ERα on DNA and nucleosomes reveals coregulator dynamics. *Molecular*

Cell, *51*(2), 185–199.

Franco, H. L., Nagari, A., & Kraus, W. L. (2015). TNFα signaling exposes latent estrogen receptor binding sites to alter the breast cancer cell transcriptome. *Molecular Cell*, *58*(1), 21–34.

Frazer, S., Prados, J., Niquille, M., Cadilhac, C., Markopoulos, F., Gomez, L., Tomasello, U., Telley, L., Holtmaat, A., Jabaudon, D., & Dayer, A. (2017). Transcriptomic and anatomic parcellation of 5-HT3AR expressing cortical interneuron subtypes revealed by single-cell RNA sequencing. *Nature Communications*, *8*, 14219.

Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., Orlov, Y. L., Velkov, S., Ho, A., Mei, P. H., Chew, E. G. Y., Huang, P. Y. H., Welboren, W.-J., Han, Y., Ooi, H. S., Ariyaratne, P. N., Vega, V. B., Luo, Y., Tan, P. Y., … Ruan, Y. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, *462*(7269), 58–64.

Furigo, I. C., Ramos-Lobo, A. M., Frazão, R., & Donato, J., Jr. (2016). Brain STAT5 signaling and behavioral control. *Molecular and Cellular Endocrinology*, *438*, 70–76.

Gagnidze, K., Weil, Z. M., & Pfaff, D. W. (2010). Histone modifications proposed to regulate sexual differentiation of brain and behavior. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *32*(11), 932–939.

Gangolli, E. A., Conneely, O. M., & O'Malley, B. W. (1997). Neurotransmitters activate the human estrogen receptor in a neuroblastoma cell line. *The Journal of Steroid Biochemistry and Molecular Biology*, *61*(1-2), 1–9.

Gantz, I., Miwa, H., Konda, Y., Shimoto, Y., Tashiro, T., Watson, S. J., DelValle, J., & Yamada, T. (1993). Molecular cloning, expression, and gene localization of a fourth melanocortin receptor. *The Journal of Biological Chemistry*, *268*(20), 15174–15179.

García-López, M., Abellán, A., Legaz, I., Rubenstein, J. L. R., Puelles, L., & Medina, L. (2008). Histogenetic compartments of the mouse centromedial and extended amygdala based on gene expression patterns during development. *The Journal of Comparative Neurology*, *506*(1), 46–74.

Garfield, A. S., Li, C., Madara, J. C., Shah, B. P., Webber, E., Steger, J. S., Campbell, J. N., Gavrilova, O., Lee, C. E., Olson, D. P., Elmquist, J. K., Tannous, B. A., Krashes, M. J., & Lowell, B. B. (2015). A neural basis for melanocortin-4 receptor-regulated appetite. *Nature Neuroscience*, *18*(6), 863–871.

Gee, D. G., Bath, K. G., Johnson, C. M., Meyer, H. C., Murty, V. P., van den Bos, W., & Hartley, C. A. (2018). Neurocognitive Development of Motivated Behavior: Dynamic Changes across Childhood and Adolescence. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *38*(44), 9433–9445.

Gemberling, M. P., Siklenka, K., Rodriguez, E., Tonn-Eisinger, K. R., Barrera, A., Liu, F., Kantor, A., Li, L., Cigliola, V., Hazlett, M. F., Williams, C. A., Bartelt, L. C., Madigan, V. J., Bodle, J. C., Daniels, H., Rouse, D. C., Hilton, I. B., Asokan, A., Ciofani, M., … Gersbach, C. A. (2021). Transgenic mice for in vivo epigenome editing with CRISPR-based systems. *Nature Methods*, *18*(8), 965–974.

Gertz, J., Savic, D., Varley, K. E., Partridge, E. C., Safi, A., Jain, P., Cooper, G. M., Reddy, T. E., Crawford, G. E., & Myers, R. M. (2013). Distinct properties of cell-type-specific and shared transcription factor binding sites. *Molecular Cell*, *52*(1), 25–36.

Ghahramani, N. M., Ngun, T. C., Chen, P.-Y., Tian, Y., Krishnan, S., Muir, S., Rubbi, L., Arnold, A. P., de Vries, G. J., Forger, N. G., Pellegrini, M., & Vilain, E. (2014). The effects of perinatal testosterone exposure on the DNA methylome of the mouse brain are late-emerging. *Biology of Sex Differences*, *5*, 8.

Gibcus, J. H., & Dekker, J. (2013). The hierarchy of the 3D genome. *Molecular Cell*, *49*(5), 773–782.

González, M., Cabrera-Socorro, A., Pérez-García, C. G., Fraser, J. D., López, F. J., Alonso, R., & Meyer, G. (2007). Distribution patterns of estrogen receptor alpha and beta in the human cortex and hippocampus during development and adulthood. *The Journal of Comparative Neurology*, *503*(6), 790–802.

Gordon, F. K., Vallaster, C. S., Westerling, T., Iyer, L. K., Brown, M., & Schnitzler, G. R. (2014). Research resource: Aorta- and liver-specific ERα-binding patterns and gene regulation by estrogen. *Molecular Endocrinology* , *28*(8), 1337–1351.

Gorski, R. A., Gordon, J. H., Shryne, J. E., & Southam, A. M. (1978). Evidence for a morphological sex difference within the medial preoptic area of the rat brain. *Brain Research*, *148*(2), 333–346.

Gotsiridze, T., Kang, N., Jacob, D., & Forger, N. G. (2007). Development of sex differences in the principal nucleus of the bed nucleus of the stria terminalis of mice: role of Bax-dependent cell death. *Developmental Neurobiology*, *67*(3), 355–362.

Grabowska, M. M., Elliott, A. D., DeGraff, D. J., Anderson, P. D., Anumanthan, G., Yamashita, H., Sun, Q., Friedman, D. B., Hachey, D. L., Yu, X., Sheehan, J. H., Ahn, J.-M., Raj, G. V., Piston, D. W., Gronostajski, R. M., & Matusik, R. J. (2014). NFI transcription factors interact with FOXA1 to regulate prostate-specific gene expression. *Molecular Endocrinology* , *28*(6), 949–964.

Granja, J. M., Corces, M. R., Pierce, S. E., Bagdatli, S. T., Choudhry, H., Chang, H. Y., & Greenleaf, W. J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nature Genetics*, *53*(3), 403–411.

Grant, C. E., Bailey, T. L., & Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* , *27*(7), 1017–1018.

Grattan, D. R. (2002). Behavioural significance of prolactin signalling in the central nervous system during pregnancy and lactation. *Reproduction* , *123*(4), 497–506.

Greig, L. C., Woodworth, M. B., Galazo, M. J., Padmanabhan, H., & Macklis, J. D. (2013). Molecular logic of neocortical projection neuron specification, development and diversity. *Nature Reviews. Neuroscience*, *14*(11), 755–769.

Grove-Strawser, D., Boulware, M. I., & Mermelstein, P. G. (2010). Membrane estrogen receptors activate the metabotropic glutamate receptors mGluR5 and mGluR3 to bidirectionally regulate

CREB phosphorylation in female rat striatal neurons. *Neuroscience*, *170*(4), 1045–1055.

Guertin, M. J., Zhang, X., Coonrod, S. A., & Hager, G. L. (2014). Transient estrogen receptor binding and p300 redistribution support a squelching mechanism for estradiol-repressed genes. *Molecular Endocrinology* , *28*(9), 1522–1533.

Gu, G., Cornea, A., & Simerly, R. B. (2003). Sexual differentiation of projections from the principal nucleus of the bed nuclei of the stria terminalis. *The Journal of Comparative Neurology*, *460*(4), 542–562.

Gustafson, P., Ladyman, S. R., McFadden, S., Larsen, C., Khant Aung, Z., Brown, R. S. E., Bunn, S. J., & Grattan, D. R. (2020). Prolactin receptor-mediated activation of pSTAT5 in the pregnant mouse brain. *Journal of Neuroendocrinology*, *32*(11), e12901.

Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* , *32*(18), 2847–2849.

Hahne, F., & Ivanek, R. (2016). Visualizing Genomic Data Using Gviz and Bioconductor. In E. Mathé & S. Davis (Eds.), *Statistical Genomics: Methods and Protocols* (pp. 335–351). Springer New York.

Hashikawa, K., Hashikawa, Y., Lischinsky, J., & Lin, D. (2018). The Neural Mechanisms of Sexually Dimorphic Aggressive Behaviors. *Trends in Genetics: TIG*, *34*(10), 755–776.

Hashikawa, K., Hashikawa, Y., Liu, Y., Rossi, M. A., Basiri, M. L., Chen, J. Y., Ahmad, O. R., Mukundan, R. V., Johnston, N. L., McHenry, J. A., Palmiter, R. D., Rubinow, D. R., Zweifel, L. S., & Stuber, G. D. (2021). Pubertal sex hormones control transcriptional trajectories in the medial preoptic area. In *bioRxiv* (p. 2021.09.02.458782). https://doi.org/10.1101/2021.09.02.458782

Hashikawa, K., Hashikawa, Y., Tremblay, R., Zhang, J., Feng, J. E., Sabol, A., Piper, W. T., Lee, H., Rudy, B., & Lin, D. (2017). Esr1+ cells in the ventromedial hypothalamus control female aggression. *Nature Neuroscience*, *20*(11), 1580–1590.

Heimovics, S. A., Trainor, B. C., & Soma, K. K. (2015). Rapid Effects of Estradiol on Aggression in Birds and Mice: The Fast and the Furious. *Integrative and Comparative Biology*, *55*(2), 281–293.

Henikoff, S., & Greally, J. M. (2016). Epigenetics, cellular memory and gene regulation. *Current Biology: CB*, *26*(14), R644–R648.

Herber, C. B., Krause, W. C., Wang, L., Bayrer, J. R., Li, A., Schmitz, M., Fields, A., Ford, B., Zhang, Z., Reid, M. S., Nomura, D. K., Nissenson, R. A., Correa, S. M., & Ingraham, H. A. (2019). Estrogen signaling in arcuate Kiss1 neurons suppresses a sex-dependent female circuit promoting dense strong bones. *Nature Communications*, *10*(1), 163.

Hewitt, S. C., Li, L., Grimm, S. A., Chen, Y., Liu, L., Li, Y., Bushel, P. R., Fargo, D., & Korach, K. S. (2012). Research resource: whole-genome estrogen receptor α binding in mouse uterine tissue revealed by ChIP-seq. *Molecular Endocrinology* , *26*(5), 887–898.

He, Y., & Ecker, J. R. (2015). Non-CG Methylation in the Human Genome. *Annual Review of Genomics and Human Genetics*, *16*(1), 150615185749007.

Hill, K. K., Roemer, S. C., Churchill, M. E. A., & Edwards, D. P. (2012). Structural and functional analysis of domains of the progesterone receptor. *Molecular and Cellular Endocrinology*, *348*(2), 418–429.

Hines, M. (2006a). Prenatal testosterone and gender-related behaviour. *European Journal of Endocrinology / European Federation of Endocrine Societies*, *155 Suppl* , S115–S121.

Hines, M. (2006b). Prenatal testosterone and gender-related behaviour. *European Journal of Endocrinology / European Federation of Endocrine Societies*, *155 Suppl 1*, S115–S121.

Hines, M. (2008). Early androgen influences on human neural and behavioural development. *Early Human Development*, *84*(12), 805–807.

Hines, M. (2010). Sex-related variation in human behavior and the brain. *Trends in Cognitive Sciences*, *14*(10), 448–456.

Hines, M., Allen, L. S., & Gorski, R. A. (1992a). Sex differences in subregions of the medial nucleus of the amygdala and the bed nucleus of the stria terminalis of the rat. *Brain Research*, *579*(2), 321–326.

Hines, M., Allen, L. S., & Gorski, R. A. (1992b). Sex differences in subregions of the medial nucleus of the amygdala and the bed nucleus of the stria terminalis of the rat. *Brain Research*, *579*(2), 321–326.

Hines, M., Pasterski, V., Spencer, D., Neufeld, S., Patalay, P., Hindmarsh, P. C., Hughes, I. A., & Acerini, C. L. (2016). Prenatal androgen exposure alters girls' responses to information indicating gender-appropriate behaviour. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1688), 20150125.

Hobert, O. (2021). Homeobox genes and the specification of neuronal identity. *Nature Reviews. Neuroscience*, *22*(10), 627–636.

Hobert, O., Carrera, I., & Stefanakis, N. (2010). The molecular and gene regulatory signature of a neuron. *Trends in Neurosciences*, *33*(10), 435–445.

Holding, A. N., Cullen, A. E., & Markowetz, F. (2018). Genome-wide Estrogen Receptor-α activation is sustained, not cyclical. *eLife*, *7*. https://doi.org/10.7554/eLife.40854

Hon, G. C., Rajagopal, N., Shen, Y., McCleary, D. F., Yue, F., Dang, M. D., & Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nature Genetics*, *45*(10), 1198–1206.

Horstmann, A., Kovacs, P., Kabisch, S., Boettcher, Y., Schloegl, H., Tönjes, A., Stumvoll, M., Pleger, B., & Villringer, A. (2013). Common genetic variation near MC4R has a sex-specific impact on human brain structure and eating behavior. *PloS One*, *8*(9), e74362.

Howard, A. D., Wang, R., Pong, S. S., Mellin, T. N., Strack, A., Guan, X. M., Zeng, Z., Williams, D. L., Jr, Feighner, S. D., Nunes, C. N., Murphy, B., Stair, J. N., Yu, H., Jiang, Q., Clements, M. K., Tan, C. P., McKee, K. K., Hreniuk, D. L., McDonald, T. P., … Liu, Q. (2000). Identification

of receptors for neuromedin U and its role in feeding. *Nature*, *406*(6791), 70–74.

Hsieh, C.-L., Fei, T., Chen, Y., Li, T., Gao, Y., Wang, X., Sun, T., Sweeney, C. J., Lee, G.-S. M., Chen, S., Balk, S. P., Liu, X. S., Brown, M., & Kantoff, P. W. (2014). Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(20), 7319–7324.

Hsu, P. Y., Hsu, H. K., Singer, G. A. C., Yan, P. S., Rodriguez, B. A. T., Liu, J. C., Weng, Y. I., Deatherage, D. E., Chen, Z., Pereira, J. S., Lopez, R., Russo, J., Wang, Q., Lamartiniere, C. A., Nephew, K. P., & Huang, T. H. M. (2010). Estrogen-mediated epigenetic repression of large chromosomal regions through DNA looping. *Genome Research*, *20*(6), 733–744.

Huang, N., Li, S., Xie, Y., Han, Q., Xu, X.-M., & Sheng, Z.-H. (2021). Reprogramming an energetic AKT-PAK5 axis boosts axon energy supply and facilitates neuron survival and regeneration after injury and ischemia. *Current Biology: CB*, *31*(14), 3098–3114.e7.

Huang, Z. J., & Paul, A. (2019). The diversity of GABAergic neurons and neural communication elements. *Nature Reviews. Neuroscience*, *20*(9), 563–572.

Hu, Q., Zhang, B., Chen, R., Fu, C., A, J., Fu, X., Li, J., Fu, L., Zhang, Z., & Dong, J.-T. (2019). ZFHX3 is indispensable for ERβ to inhibit cell proliferation via MYC downregulation in prostate cancer cells. *Oncogenesis*, *8*(4), 28.

Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D., & Carroll, J. S. (2011). FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nature Genetics*, *43*(1), 27–33.

Huszar, D., Lynch, C. A., Fairchild-Huntress, V., Dunmore, J. H., Fang, Q., Berkemeier, L. R., Gu, W., Kesterson, R. A., Boston, B. A., Cone, R. D., Smith, F. J., Campfield, L. A., Burn, P., & Lee, F. (1997). Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell*, *88*(1), 131–141.

Hutton, L. A., Gu, G., & Simerly, R. B. (1998). Development of a sexually dimorphic projection from the bed nuclei of the stria terminalis to the anteroventral periventricular nucleus in the rat. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *18*(8), 3003–3013.

Hu, Y., Park, K. K., Yang, L., Wei, X., Yang, Q., Cho, K.-S., Thielen, P., Lee, A.-H., Cartoni, R., Glimcher, L. H., Chen, D. F., & He, Z. (2012). Differential effects of unfolded protein response pathways on axon injury-induced death of retinal ganglion cells. *Neuron*, *73*(3), 445–452.

Ibanez, M. A., Gu, G., & Simerly, R. B. (2001). Target-dependent sexual differentiation of a limbic-hypothalamic neural pathway. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *21*(15), 5652–5659.

Ishii, K. K., & Touhara, K. (2019). Neural circuits regulating sexual behaviors via the olfactory system in mice. *Neuroscience Research*, *140*, 59–76.

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics*, *33 Suppl*, 245–254.

Janssens, D. H., Otto, D. J., Meers, M. P., Setty, M., Ahmad, K., & Henikoff, S. (2021). Simultaneous CUT&Tag profiling of the accessible and silenced regulome in single cells. In *bioRxiv* (p. 2021.12.19.473377). https://doi.org/10.1101/2021.12.19.473377

Jenuwein, T., & Allis, C. D. (2001). Translating the histone code. *Science*, *293*(5532), 1074–1080.

Jones, K. J., Pfaff, D. W., & McEwen, B. S. (1985). Early estrogen-induced nuclear changes in rat hypothalamic ventromedial neurons: an ultrastructural and morphometric analysis. *The Journal of Comparative Neurology*, *239*(3), 255–266.

Jubb, A. W., Boyle, S., Hume, D. A., & Bickmore, W. A. (2017). Glucocorticoid Receptor Binding Induces Rapid and Prolonged Large-Scale Chromatin Decompaction at Multiple Target Loci. *Cell Reports*, *21*(11), 3022–3031.

Juntti, S. A., Tollkuhn, J., Wu, M. V., Fraser, E. J., Soderborg, T., Tan, S., Honda, S.-I., Harada, N.,

& Shah, N. M. (2010). The androgen receptor governs the execution, but not programming, of male sexual and territorial behaviors. *Neuron*, *66*(2), 260–272.

Karigo, T., Kennedy, A., Yang, B., Liu, M., Tai, D., Wahle, I. A., & Anderson, D. J. (2020). Distinct hypothalamic control of same- and opposite-sex mounting behaviour in mice. *Nature*. https://doi.org/10.1038/s41586-020-2995-0

Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K., & Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications*, *10*(1), 1930.

Kelava, I., Chiaradia, I., Pellegrini, L., Kalinka, A. T., & Lancaster, M. A. (2022). Androgens increase excitatory neurogenic potential in human brain organoids. *Nature*, *602*(7895), 112–116.

Kelly, D. A., Varnum, M. M., Krentzel, A. A., Krug, S., & Forger, N. G. (2013). Differential control of sex differences in estrogen receptor α in the bed nucleus of the stria terminalis and anteroventral periventricular nucleus. *Endocrinology*, *154*(10), 3836–3846.

Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Ustün, T. B. (2007). Age of onset of mental disorders: a review of recent literature. *Current Opinion in Psychiatry*, *20*(4), 359–364.

Kimchi, T., Xu, J., & Dulac, C. (2007). A functional circuit underlying male sexual behaviour in the female mouse brain. *Nature*, *448*(7157), 1009–1014.

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360.

Kim, D.-W., Yao, Z., Graybuck, L. T., Kim, T. K., Nguyen, T. N., Smith, K. A., Fong, O., Yi, L., Koulena, N., Pierson, N., Shah, S., Lo, L., Pool, A.-H., Oka, Y., Pachter, L., Cai, L., Tasic, B., Zeng, H., & Anderson, D. J. (2019). Multimodal Analysis of Cell Types in a Hypothalamic Node Controlling Social Behavior. *Cell*, *179*(3), 713–728.e17.

Kim, J.-Y., Koh, H. C., Lee, J.-Y., Chang, M.-Y., Kim, Y.-C., Chung, H.-Y., Son, H., Lee, Y.-S.,

Studer, L., McKay, R., & Lee, S.-H. (2003). Dopaminergic neuronal differentiation from rat embryonic neural precursors by Nurr1 overexpression. *Journal of Neurochemistry*, *85*(6), 1443–1454.

Kimura, K.-I., Ote, M., Tazawa, T., & Yamamoto, D. (2005). Fruitless specifies sexually dimorphic neural circuitry in the Drosophila brain. *Nature*, *438*(7065), 229–233.

Klemm, S. L., Shipony, Z., & Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nature Reviews. Genetics*, *20*(4), 207–220.

Kohl, J., Babayan, B. M., Rubinstein, N. D., Autry, A. E., Marin-Rodriguez, B., Kapoor, V., Miyamishi, K., Zweifel, L. S., Luo, L., Uchida, N., & Dulac, C. (2018). Functional circuit architecture underlying parental behaviour. *Nature*, *556*(7701), 326–331.

Kokay, I. C., Wyatt, A., Phillipps, H. R., Aoki, M., Ectors, F., Boehm, U., & Grattan, D. R. (2018). Analysis of prolactin receptor expression in the murine brain using a novel prolactin receptor reporter mouse. *Journal of Neuroendocrinology*, *30*(9), e12634.

Korsia, S., & Bottjer, S. W. (1991). Chronic testosterone treatment impairs vocal learning in male zebra finches during a restricted period of development. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *11*(8), 2362–2371.

Kos, M., Reid, G., Denger, S., & Gannon, F. (2001). Minireview: genomic organization of the human ERalpha gene promoter region. *Molecular Endocrinology* , *15*(12), 2057–2063.

Krashes, M. J., Koda, S., Ye, C., Rogan, S. C., Adams, A. C., Cusher, D. S., Maratos-Flier, E., Roth, B. L., & Lowell, B. B. (2011). Rapid, reversible activation of AgRP neurons drives feeding behavior in mice. *The Journal of Clinical Investigation*, *121*(4), 1424–1428.

Krause, W. C., & Ingraham, H. A. (2017). Origins and Functions of the Ventrolateral VMH: A Complex Neuronal Cluster Orchestrating Sex Differences in Metabolism and Behavior. *Advances in Experimental Medicine and Biology*, *1043*, 199–213.

Krishnan, S., Intlekofer, K. A., Aggison, L. K., & Petersen, S. L. (2009). Central role of TRAF-

interacting protein in a new model of brain sexual differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(39), 16692–16697.

Kritzer, M. F. (2002). Regional, Laminar, and Cellular Distribution of Immunoreactivity for ERalpha and ERbeta in the Cerebral Cortex of Hormonally Intact, Adult Male and Female Rats. *Cerebral Cortex* , *12*, 116–128.

Kurian, J. R., Olesen, K. M., & Auger, A. P. (2010). Sex differences in epigenetic regulation of the estrogen receptor-alpha promoter within the developing preoptic area. *Endocrinology*, *151*(5), 2297–2305.

Laman-Maharg, A., & Trainor, B. C. (2017). Stress, sex, and motivated behaviors. *Journal of Neuroscience Research*, *95*(1-2), 83–92.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359.

Le Dily, F., & Beato, M. (2018). Signaling by steroid hormones in the 3D nuclear space. *International Journal of Molecular Sciences*, *19*(2), 1–16.

Le Dily, F., Vidal, E., Cuartero, Y., Quilez, J., Nacht, A. S., Vicent, G. P., Carbonell-Caballero, J., Sharma, P., Villanueva-Cañas, J. L., Ferrari, R., De Llobet, L. I., Verde, G., Wright, R. H. G., & Beato, M. (2019). Hormone-control regions mediate steroid receptor-dependent genome organization. *Genome Research*, *29*(1), 29–39.

Lee, H., Kim, D.-W., Remedios, R., Anthony, T. E., Chang, A., Madisen, L., Zeng, H., & Anderson, D. J. (2014). Scalable control of mounting and attack by Esr1+ neurons in the ventromedial hypothalamus. *Nature*, *509*(7502), 627–632.

Lee, S. J., Krieger, J.-P., Vergara, M., Quinn, D., McDougle, M., de Araujo, A., Darling, R., Zollinger, B., Anderson, S., Pan, A., Simonnet, E. J., Pignalosa, A., Arnold, M., Singh, A., Langhans, W., Raybould, H. E., & de Lartigue, G. (2020). Blunted Vagal Cocaine- and Amphetamine-Regulated Transcript Promotes Hyperphagia and Weight Gain. *Cell Reports*,

30(6), 2028–2039.e4.

Liao, Y., Smyth, G. K., & Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* , *30*(7), 923–930.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25*(16), 2078–2079.

Lim, L., Mi, D., Llorca, A., & Marín, O. (2018). Development and Functional Diversification of Cortical Interneurons. *Neuron*, *100*(2), 294–313.

Lin, D., Boyle, M. P., Dollar, P., Lee, H., Lein, E. S., Perona, P., & Anderson, D. J. (2011). Functional identification of an aggression locus in the mouse hypothalamus. *Nature*, *470*(7333), 221–226.

Lister, R., Mukamel, E. a., Nery, J. R., Urich, M., Puddifoot, C. a., Johnson, N. D., Lucero, J., Huang, Y., Dwork, A. J., Schultz, M. D., Yu, M., Tonti-Filippini, J., Heyn, H., Hu, S., Wu, J. C., Rao, A., Esteller, M., He, C., Haghighi, F. G., … He, C. (2013). Global Epigenomic Reconfiguration During Mammalian Brain Development. *Science* , *341*(mC), 629–643.

Liu, M., Kim, D.-W., Zeng, H., & Anderson, D. J. (2022). Make war not love: The neural substrate underlying a state-dependent switch in female social behavior. *Neuron*, *110*(5), 841–856.e6.

Li, Y., & Dulac, C. (2018). Neural coding of sex-specific social information in the mouse brain. *Current Opinion in Neurobiology*, *53*, 120–130.

Li, Y. E., Preissl, S., Hou, X., Zhang, Z., Zhang, K., Qiu, Y., Poirion, O. B., Li, B., Chiou, J., Liu, H., Pinto-Duarte, A., Kubo, N., Yang, X., Fang, R., Wang, X., Han, J. Y., Lucero, J., Yan, Y., Miller, M., … Ren, B. (2021). An atlas of gene regulatory elements in adult mouse cerebrum. *Nature*, *598*(7879), 129–136.

Lonstein, J. S., & Gammie, S. C. (2002). Sensory, hormonal, and neural control of maternal aggression. *Neuroscience and Biobehavioral Reviews*, *26*, 869–888.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550.

MacLusky, N. J., & Naftolin, F. (1981). Sexual Differentiation of the Central Nervous System. *Science*, *211*(4488), 1294–1303.

Maekawa, R., Sato, S., Okada, M., Lee, L., Tamura, I., Jozaki, K., Kajimura, T., Asada, H., Yamagata, Y., Tamura, H., Yamamoto, S., & Sugino, N. (2016). Tissue-Specific Expression of Estrogen Receptor 1 Is Regulated by DNA Methylation in a T-DMR. *Molecular Endocrinology* , *30*(3), 335–347.

Maggi, A. (2011). Liganded and unliganded activation of estrogen receptor and hormone replacement therapies. *Biochimica et Biophysica Acta*, *1812*(8), 1054–1060.

Manoli, D. S., & Tollkuhn, J. (2018a). Gene regulatory mechanisms underlying sex differences in brain development and psychiatric disease. *Annals of the New York Academy of Sciences*. https://doi.org/10.1111/nyas.13564

Manoli, D. S., & Tollkuhn, J. (2018b). Gene regulatory mechanisms underlying sex differences in brain development and psychiatric disease. *Annals of the New York Academy of Sciences*, *1420*(1), 26–45.

Marín, O. (2016). Developmental timing and critical windows for the treatment of psychiatric disorders. *Nature Medicine*, *22*(11), 1229–1238.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, *17*(1), 10–12.

Ma, S., Zhang, B., LaFave, L. M., Earl, A. S., Chiang, Z., Hu, Y., Ding, J., Brack, A., Kartha, V. K., Tay, T., Law, T., Lareau, C., Hsu, Y.-C., Regev, A., & Buenrostro, J. D. (2020). Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and Chromatin. *Cell*, *183*(4), 1103–1116.e20.

Matharu, N., Rattanasopha, S., Tamura, S., Maliskova, L., Wang, Y., Bernard, A., Hardin, A.,

Eckalbar, W. L., Vaisse, C., & Ahituv, N. (2019). CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency. *Science*, *363*(6424). https://doi.org/10.1126/science.aau0629

Matsuda, K. I., Mori, H., Nugent, B. M., Pfaff, D. W., McCarthy, M. M., & Kawata, M. (2011). Histone deacetylation during brain development is essential for permanent masculinization of sexual behavior. *Endocrinology*, *152*(7), 2760–2767.

Mayer, C., Hafemeister, C., Bandler, R. C., Machold, R., Batista Brito, R., Jaglin, X., Allaway, K., Butler, A., Fishell, G., & Satija, R. (2018). Developmental diversification of cortical inhibitory interneurons. *Nature*, *555*(7697), 457–462.

McCarthy, M. M. (2008a). Estradiol and the developing brain. *Physiological Reviews*, *88*, 91–124.

McCarthy, M. M. (2008b). Estradiol and the developing brain. *Physiological Reviews*, *88*, 91–124.

McCarthy, M. M., Auger, A. P., Bale, T. L., De Vries, G. J., Dunn, G. A., Forger, N. G., Murray, E. K., Nugent, B. M., Schwarz, J. M., & Wilson, M. E. (2009). The epigenetics of sex differences in the brain. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(41), 12815–12823.

McCarthy, M. M., & Nugent, B. M. (2015). At the frontier of epigenetics of brain sex differences. *Frontiers in Behavioral Neuroscience*, *9*(August), 221.

McCarthy, M. M., Nugent, B. M., & Lenz, K. M. (2017). Neuroimmunology and neuroepigenetics in the establishment of sex differences in the brain. *Nature Reviews. Neuroscience*, *18*(8), 471–484.

McCarthy, M. M., Wright, C. L., & Schwarz, J. M. (2009). New tricks by an old dogma: mechanisms of the Organizational/Activational Hypothesis of steroid-mediated sexual differentiation of brain and behavior. *Hormones and Behavior*, *55*(5), 655–665.

McEwen, B. S. (1981). Neural gonadal steroid actions. *Science*, *211*(4488), 1303–1311.

McGinnis, C. S., Murrow, L. M., & Gartner, Z. J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Systems*, *8*(4), 329–

337.e4.

McKenna, N. J., & O'Malley, B. W. (2002). Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell*, *108*(4), 465–474.

Métivier, R., Penot, G., Hübner, M. R., Reid, G., Brand, H., Kos, M., & Gannon, F. (2003). Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell*, *115*(6), 751–763.

Micevych, P. E., & Kelly, M. J. (2012). Membrane estrogen receptor regulation of hypothalamic function. *Neuroendocrinology*, *96*(2), 103–110.

Miyoshi, G., Young, A., Petros, T., Karayannis, T., McKenzie Chang, M., Lavado, A., Iwano, T., Nakajima, M., Taniguchi, H., Huang, Z. J., Heintz, N., Oliver, G., Matsuzaki, F., Machold, R. P., & Fishell, G. (2015). Prox1 Regulates the Subtype-Specific Development of Caudal Ganglionic Eminence-Derived GABAergic Cortical Interneurons. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(37), 12869–12889.

Mo, A., Mukamel, E. A., Davis, F. P., Luo, C., Henry, G. L., Picard, S., Urich, M. A., Nery, J. R., Sejnowski, T. J., Lister, R., Eddy, S. R., Ecker, J. R., & Nathans, J. (2015). Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron*, *86*(6), 1369–1384.

Moffitt, J. R., Bambah-Mukku, D., Eichhorn, S. W., Vaughn, E., Shekhar, K., Perez, J. D., Rubinstein, N. D., Hao, J., Regev, A., Dulac, C., & Zhuang, X. (2018a). Molecular, spatial and functional single-cell profiling of the hypothalamic preoptic region. *Science*. https://doi.org/10.1126/science.aau5324

Moffitt, J. R., Bambah-Mukku, D., Eichhorn, S. W., Vaughn, E., Shekhar, K., Perez, J. D., Rubinstein, N. D., Hao, J., Regev, A., Dulac, C., & Zhuang, X. (2018b). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science*, *362*(6416). https://doi.org/10.1126/science.aau5324

Mohammed, H., D'Santos, C., Serandour, A. A., Ali, H. R., Brown, G. D., Atkins, A., Rueda, O. M.,

Holmes, K. A., Theodorou, V., Robinson, J. L. L., Zwart, W., Saadi, A., Ross-Innes, C. S., Chin, S.-F., Menon, S., Stingl, J., Palmieri, C., Caldas, C., & Carroll, J. S. (2013). Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Reports*, *3*(2), 342–349.

Molenda, H. A., Griffin, A. L., Auger, A. P., McCarthy, M. M., & Tetel, M. J. (2002). Nuclear receptor coactivators modulate hormone-dependent gene expression in brain and female reproductive behavior in rats. *Endocrinology*, *143*(2), 436–444.

Mortensen, L. S., Park, S. J. H., Ke, J.-B., Cooper, B. H., Zhang, L., Imig, C., Löwel, S., Reim, K., Brose, N., Demb, J. B., Rhee, J.-S., & Singer, J. H. (2016). Complexin 3 Increases the Fidelity of Signaling in a Retinal Circuit by Regulating Exocytosis at Ribbon Synapses. *Cell Reports*, *15*(10), 2239–2250.

Mountjoy, K. G., Mortrud, M. T., Low, M. J., Simerly, R. B., & Cone, R. D. (1994). Localization of the melanocortin-4 receptor (MC4-R) in neuroendocrine and autonomic control circuits in the brain. *Molecular Endocrinology* , *8*(10), 1298–1308.

Murray, E. K., Hien, A., de Vries, G. J., & Forger, N. G. (2009). Epigenetic control of sexual differentiation of the bed nucleus of the stria terminalis. *Endocrinology*, *150*(9), 4241–4247.

Murr, S. M., Stabenfeldt, G. H., Bradford, G. E., & Geschwind, I. I. (1974). Plasma progesterone during pregnancy in the mouse. *Endocrinology*, *94*(4), 1209–1211.

Nery, S., Fishell, G., & Corbin, J. G. (2002). The caudal ganglionic eminence is a source of distinct cortical and subcortical cell populations. *Nature Neuroscience*, *5*(12), 1279–1287.

Newman, S. W. (1999). The medial extended amygdala in male reproductive behavior. A node in the mammalian social behavior network. *Annals of the New York Academy of Sciences*, *877*, 242–257.

Nguyen, T.-V. V., Yao, M., & Pike, C. J. (2009). Dihydrotestosterone activates CREB signaling in cultured hippocampal neurons. *Brain Research*, *1298*, 1–12.

Nordeen, E. J., Nordeen, K. W., Sengelaub, D. R., & Arnold, A. P. (1985). Androgens prevent normally occurring cell death in a sexually dimorphic spinal nucleus. *Science*, *229*(4714), 671–673.

Nugent, B. M., Wright, C. L., Shetty, A. C., Hodes, G. E., Lenz, K. M., Mahurkar, A., Russo, S. J., Devine, S. E., & Mccarthy, M. M. (2015). Brain feminization requires active repression of masculinization via DNA methylation. *Nature Neuroscience*, *18*(August 2014). https://doi.org/10.1038/nn.3988

O'Connell, L. A., & Hofmann, H. A. (2011). The Vertebrate mesolimbic reward system and social behavior network: A comparative synthesis. *The Journal of Comparative Neurology*, *519*(18), 3599–3639.

Ogawa, S., Lubahn, D. B., Korach, K. S., & Pfaff, D. W. (1997). Behavioral effects of estrogen receptor gene disruption in male mice. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(4), 1476–1481.

Overstreet-Wadiche, L., & McBain, C. J. (2015). Neurogliaform cells in cortical circuits. *Nature Reviews Neuroscience*, *16*(8), 458–468.

Paakinaho, V., Swinstead, E. E., Presman, D. M., Grøntved, L., & Hager, G. L. (2019). Meta-analysis of Chromatin Programming by Steroid Receptors. *Cell Reports*, *28*(13), 3523–3534.e2.

Palaniappan, M., Nguyen, L., Grimm, S. L., Xi, Y., Xia, Z., Li, W., & Coarfa, C. (2019). The genomic landscape of estrogen receptor α binding sites in mouse mammary gland. *PloS One*, *14*(8), e0220311.

Palierne, G., Fabre, A., Solinhac, R., Le Péron, C., Avner, S., Lenfant, F., Fontaine, C., Salbert, G., Flouriot, G., Arnal, J.-F., & Métivier, R. (2016). Changes in Gene Expression and Estrogen Receptor Cistrome in Mouse Liver Upon Acute E2 Treatment. *Molecular Endocrinology* , *30*(7), 709–732.

Panda, S., Dohare, P., Jain, S., Parikh, N., Singla, P., Mehdizadeh, R., Klebe, D. W., Kleinman, G.

M., Cheng, B., & Ballabh, P. (2018). Estrogen treatment reverses prematurity-induced disruption in cortical interneuron population. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 0478–0418.

Papachristou, E. K., Kishore, K., Holding, A. N., Harvey, K., Roumeliotis, T. I., Chilamakuri, C. S. R., Omarjee, S., Chia, K. M., Swarbrick, A., Lim, E., Markowetz, F., Eldridge, M., Siersbaek, R., D'Santos, C. S., & Carroll, J. S. (2018). A quantitative mass spectrometry-based approach to monitor the dynamics of endogenous chromatin-associated protein complexes. *Nature Communications*, *9*(1), 2311.

Paul, A., Crow, M., Raudales, R., He, M., Gillis, J., & Huang, Z. J. (2017a). Transcriptional Architecture of Synaptic Communication Delineates GABAergic Neuron Identity. *Cell*, *171*(3), 522–539.e20.

Paul, A., Crow, M., Raudales, R., He, M., Gillis, J., & Huang, Z. J. (2017b). Transcriptional Architecture of Synaptic Communication Delineates GABAergic Neuron Identity. *Cell*, *171*(3), 522–539.e20.

Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? *Nature Reviews. Neuroscience*, *9*(12), 947–957.

Paxinos, G., & Franklin, K. B. J. (2019). *Paxinos and Franklin's the Mouse Brain in Stereotaxic Coordinates*. Academic Press.

Peden, E., Kimberly, E., Gengyo-Ando, K., Mitani, S., & Xue, D. (2007). Control of sex-specific apoptosis in C. elegans by the BarH homeodomain protein CEH-30 and the transcriptional repressor UNC-37/Groucho. *Genes & Development*, *21*(23), 3195–3207.

Perissi, V., & Rosenfeld, M. G. (2005). Controlling nuclear receptors: the circular logic of cofactor cycles. *Nature Reviews. Molecular Cell Biology*, *6*, 542.

Pfisterer, U., & Khodosevich, K. (2017). Neuronal survival in the brain: neuron type-specific mechanisms. *Cell Death & Disease*, *8*(3), e2643.

Phillipps, H. R., Yip, S. H., & Grattan, D. R. (2020). Patterns of prolactin secretion. *Molecular and Cellular Endocrinology*, *502*, 110679.

Phoenix, C. H., Goy, R. W., Gerall, A. A., & Young, W. C. (1959). Organizing action of prenatally administered testosterone propionate on the tissues mediating mating behavior in the female guinea pig. *Endocrinology*, *65*, 369–382.

Piekarski, D. J., Boivin, J. R., & Wilbrecht, L. (2017). Ovarian Hormones Organize the Maturation of Inhibitory Neurotransmission in the Frontal Cortex at Puberty Onset in Female Mice. *Current Biology: CB*, *27*(12), 1735–1745.e3.

Piekarski, D. J., Johnson, C. M., Boivin, J. R., Thomas, A. W., Lin, W. C., Delevich, K., M Galarce, E., & Wilbrecht, L. (2017). Does puberty mark a transition in sensitive periods for plasticity in the associative neocortex? *Brain Research*, *1654*(Pt B), 123–144.

Pjanic, M., Schmid, C. D., Gaussin, A., Ambrosini, G., Adamcik, J., Pjanic, P., Plasari, G., Kerschgens, J., Dietler, G., Bucher, P., & Mermod, N. (2013). Nuclear Factor I genomic binding associates with chromatin boundaries. *BMC Genomics*, *14*, 99.

Power, R. F., Mani, S. K., Codina, J., Conneely, O. M., & O'Malley, B. W. (1991). Dopaminergic and ligand-independent activation of steroid hormone receptors. *Science*, *254*(5038), 1636–1639.

Ptashne, M. (2007). On the use of the word "epigenetic." *Current Biology: CB*, *17*(7), R233–R236.

Ptashne, M. (2013). Epigenetics: core misconcept [Review of *Epigenetics: core misconcept*]. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(18), 7101–7103.

Qi, L., Kraft, P., Hunter, D. J., & Hu, F. B. (2008). The common obesity variant near MC4R gene is associated with higher intakes of total energy and dietary fat, weight change and diabetes risk in women. *Human Molecular Genetics*, *17*(22), 3502–3508.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* , *26*(6), 841–842.

Quinn, J. C., Molinek, M., Martynoga, B. S., Zaki, P. A., Faedo, A., Bulfone, A., Hevner, R. F., West, J. D., & Price, D. J. (2007). Pax6 controls cerebral cortical cell number by regulating exit from the cell cycle and specifies cortical cell identity by a cell autonomous mechanism. *Developmental Biology*, *302*(1), 50–65.

Rafique, S., Thomas, J. S., Sproul, D., & Bickmore, W. A. (2015). Estrogen-induced chromatin decondensation and nuclear re-organization linked to regional epigenetic regulation in breast cancer. *Genome Biology*, *16*, 145.

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., Heyne, S., Dündar, F., & Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160–W165.

Ratnu, V. S., Emami, M. R., & Bredy, T. W. (2017). Genetic and epigenetic factors underlying sex differences in the regulation of gene expression in the brain. *Journal of Neuroscience Research*, *95*(1-2), 301–310.

Raznahan, A., Lue, Y. H., Probst, F., Greenstein, D., Giedd, J., Wang, C., Lerch, J., & Swerdloff, R. (2015). Triangulating the sexually dimorphic brain through high-resolution neuroimaging of murine sex chromosome aneuploidies. *Brain Structure & Function*, *220*(6), 3581–3593.

Reyes, F. I., Boroditsky, R. S., Winter, J. S., & Faiman, C. (1974). Studies on human sexual development. II. Fetal and maternal serum gonadotropin and sex steroid concentrations. *The Journal of Clinical Endocrinology and Metabolism*, *38*(4), 612–617.

Reznikoff, W. S. (2003). Tn5 as a model for understanding DNA transposition. *Molecular Microbiology*, *47*(5), 1199–1206.

Rissman, E. F., Wersinger, S. R., Fugger, H. N., & Foster, T. C. (1999). Sex with knockout models: Behavioral studies of estrogen receptor. *Brain Research*, *835*(1), 80–90.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* , *26*(1), 139–140.

Ross-Innes, C. S., Brown, G. D., & Carroll, J. S. (2011). A co-ordinated interaction between CTCF and ER in breast cancer cells. *BMC Genomics*, *12*, 593.

Ross-Innes, C. S., Stark, R., Teschendorff, A. E., Holmes, K. A., Ali, H. R., Dunning, M. J., Brown, G. D., Gojis, O., Ellis, I. O., Green, A. R., Ali, S., Chin, S.-F., Palmieri, C., Caldas, C., & Carroll, J. S. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*, *481*(7381), 389–393.

Safe, S., & Kim, K. (2008). Non-classical genomic estrogen receptor (ER)/specificity protein and ER/activating protein-1 signaling pathways. *Journal of Molecular Endocrinology*, *41*(5), 263–275.

Sahni, V., Itoh, Y., Shnider, S. J., & Macklis, J. D. (2021). Crim1 and Kelch-like 14 exert complementary dual-directional developmental control over segmentally specific corticospinal axon projection targeting. *Cell Reports*, *37*(3), 109842.

Sandberg, M., Flandin, P., Silberberg, S., Su-Feher, L., Price, J. D., Hu, J. S., Kim, C., Visel, A., Nord, A. S., & Rubenstein, J. L. R. (2016). Transcriptional Networks Controlled by NKX2-1 in the Development of Forebrain GABAergic Neurons. *Neuron*, *91*(6), 1260–1275.

Sanders, L. E., & Arbeitman, M. N. (2008). Doublesex establishes sexual dimorphism in the Drosophila central nervous system in an isoform-dependent manner by directing cell number. *Developmental Biology*, *320*(2), 378–390.

Sanz, E., Yang, L., Su, T., Morris, D. R., McKnight, G. S., & Amieux, P. S. (2009). Cell-type-specific isolation of ribosome-associated mRNA from complex tissues. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(33), 13939–13944.

Satterstrom, F. K., Kosmicki, J. A., Wang, J., Breen, M. S., De Rubeis, S., An, J.-Y., Peng, M., Collins, R., Grove, J., Klei, L., Stevens, C., Reichert, J., Mulhern, M. S., Artomov, M., Gerges, S., Sheppard, B., Xu, X., Bhaduri, A., Norman, U., … Buxbaum, J. D. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the

Neurobiology of Autism. *Cell*, *180*(3), 568–584.e23.

Savell, K. E., Tuscher, J. J., Zipperly, M. E., Duke, C. G., Phillips, R. A., 3rd, Bauman, A. J., Thukral, S., Sultan, F. A., Goska, N. A., Ianov, L., & Day, J. J. (2020). A dopamine-induced gene expression signature regulates neuronal function and cocaine response. *Science Advances*, *6*(26), eaba4221.

Schep, A. N., Wu, B., Buenrostro, J. D., & Greenleaf, W. J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nature Methods*, *14*(10), 975–978.

Schulz, K. M., & Sisk, C. L. (2016). The organizing actions of adolescent gonadal steroid hormones on brain and behavioral development. *Neuroscience and Biobehavioral Reviews*, *70*, 148–158.

Schwartz, H. T., & Horvitz, H. R. (2007). The C. elegans protein CEH-30 protects male-specific neurons from apoptosis independently of the Bcl-2 homolog CED-9. *Genes & Development*, *21*(23), 3181–3194.

Schwarz, J. M., Nugent, B. M., & McCarthy, M. M. (2010). Developmental and hormone-induced epigenetic changes to estrogen and progesterone receptor genes in brain are dynamic across the life span. *Endocrinology*, *151*(10), 4871–4881.

Scordalakes, E. M., & Rissman, E. F. (2003). Aggression in Male Mice Lacking Functional Estrogen Receptor alpha. *Behavioral Neuroscience*, *117*(1), 38–45.

Scott, N., Prigge, M., Yizhar, O., & Kimchi, T. (2015). A sexually dimorphic hypothalamic circuit controls maternal care and oxytocin secretion. *Nature*, *525*(7570), 519–522.

Searles, S., Makarewicz, J. A., & Dumas, J. A. (2018). The role of estradiol in schizophrenia diagnosis and symptoms in postmenopausal women. *Schizophrenia Research*, *196*, 35–38.

Serrano-Saiz, E., Poole, R. J., Felton, T., Zhang, F., De La Cruz, E. D., & Hobert, O. (2013). Modular control of glutamatergic neuronal identity in C. elegans by distinct homeodomain proteins. *Cell*, *155*(3), 659–673.

Shaywitz, A. J., & Greenberg, M. E. (1999). CREB: a stimulus-induced transcription factor activated by a diverse array of extracellular signals. *Annual Review of Biochemistry*, *68*, 821–861.

Sherwin, B. B., & Gelfand, M. M. (1985). Sex steroids and affect in the surgical menopause: a double-blind, cross-over study. *Psychoneuroendocrinology*, *10*(3), 325–335.

Shilatifard, A. (2006). Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annual Review of Biochemistry*, *75*, 243–269.

Shi, L., Zhang, Z., & Su, B. (2016). Sex Biased Gene Expression Profiling of Human Brains at Major Developmental Stages. *Scientific Reports*, *6*, 21181.

Shughrue, P. J., Lane, M. V., & Merchenthaler, I. (1997). Comparative distribution of estrogen receptor-alpha and -beta mRNA in the rat central nervous system. *The Journal of Comparative Neurology*, *388*(4), 507–525.

Simeone, A. (2005). Genetic control of dopaminergic neuron differentiation. *Trends in Neurosciences*, *28*(2), 62–65; discussion 65–66.

Simerly, R. B. (2002a). Wired For Reproduction : Organization and Development of Sexually Dimorphic Circuits in the Mammalian Forebrain. *Annual Review of Neuroscience*, *25*(1), 507–536.

Simerly, R. B. (2002b). Wired for reproduction: organization and development of sexually dimorphic circuits in the mammalian forebrain. *Annual Review of Neuroscience*, *25*, 507–536.

Simerly, R. B., & Swanson, L. W. (1988). Projections of the medial preoptic nucleus: a Phaseolus vulgaris leucoagglutinin anterograde tract-tracing study in the rat. *The Journal of Comparative Neurology*, *270*(2), 209–242.

Simerly, R. B., Swanson, L. W., Handa, R. J., & Gorski, R. A. (1985). Influence of perinatal androgen on the sexually dimorphic distribution of tyrosine hydroxylase-immunoreactive cells and fibers in the anteroventral periventricular nucleus of the rat. *Neuroendocrinology*, *40*(6), 501–510.

Sina, M., Hinney, A., Ziegler, A., Neupert, T., Mayer, H., Siegfried, W., Blum, W. F., Remschmidt, H., & Hebebrand, J. (1999). Phenotypes in three pedigrees with autosomal dominant obesity caused by haploinsufficiency mutations in the melanocortin-4 receptor gene. *American Journal of Human Genetics*, *65*(6), 1501–1507.

Sisk, C. L., & Foster, D. L. (2004). The neural basis of puberty and adolescence. *Nature Neuroscience*, *7*(10), 1040–1047.

Sizemore, M., & Perkel, D. J. (2011). Premotor synaptic plasticity limited to the critical period for song learning. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(42), 17492–17497.

Skene, P. J., Henikoff, J. G., & Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nature Protocols*, *13*(5), 1006–1019.

Skene, P. J., & Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife*, *6*. https://doi.org/10.7554/eLife.21856

Sohn, J.-W., Elmquist, J. K., & Williams, K. W. (2013). Neuronal circuits that regulate feeding behavior and metabolism. *Trends in Neurosciences*, *36*(9), 504–512.

Srivastava, D. P., Woolfrey, K. M., & Penzes, P. (2013). Insights into rapid modulation of neuroplasticity by brain estrogens. *Pharmacological Reviews*, *65*(4), 1318–1350.

Stark, R., Brown, G., & Others. (2011). DiffBind: differential binding analysis of ChIP-Seq peak data. *R Package Version*, *100*, 4–3.

Ste Marie, L., Miura, G. I., Marsh, D. J., Yagaloff, K., & Palmiter, R. D. (2000). A metabolic defect promotes obesity in mice lacking melanocortin-4 receptors. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(22), 12339–12344.

Stincic, T. L., Rønnekleiv, O. K., & Kelly, M. J. (2018). Diverse actions of estradiol on anorexigenic and orexigenic hypothalamic arcuate neurons. *Hormones and Behavior*. https://doi.org/10.1016/j.yhbeh.2018.04.001

Stroud, H., Su, S. C., Hrvatin, S., Greben, A. W., Renthal, W., Boxer, L. D., Nagy, M. A., Hochbaum, D. R., Kinde, B., Gabel, H. W., & Greenberg, M. E. (2017). Early-Life Gene Expression in Neurons Modulates Lasting Epigenetic States. *Cell*, *171*(5), 1151–1164.e16.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, *177*(7), 1888–1902.e21.

Stuart, T., Srivastava, A., Madad, S., Lareau, C. A., & Satija, R. (2021). Single-cell chromatin state analysis with Signac. *Nature Methods*, *18*(11), 1333–1341.

Swinstead, E. E., Miranda, T. B., Paakinaho, V., Baek, S., Goldstein, I., Hawkins, M., Karpova, T. S., Ball, D., Mazza, D., Lavis, L. D., Grimm, J. B., Morisaki, T., Gr??ntved, L., Presman, D. M., & Hager, G. L. (2016). Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. *Cell*, *165*(3), 593–605.

Takeda, Y., & Jetten, A. M. (2013). Prospero-related homeobox 1 (Prox1) functions as a novel modulator of retinoic acid-related orphan receptors α- and γ-mediated transactivation. *Nucleic Acids Research*, *41*(14), 6992–7008.

Tasic, B., Menon, V., Nguyen, T. N., Kim, T. K., Jarsky, T., Yao, Z., Levi, B., Gray, L. T., Sorensen, S. A., Dolbeare, T., Bertagnolli, D., Goldy, J., Shapovalova, N., Parry, S., Lee, C., Smith, K., Bernard, A., Madisen, L., Sunkin, S. M., … Zeng, H. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature Neuroscience*, *19*(2), 335–346.

Tasic, B., Yao, Z., Graybuck, L. T., Smith, K. A., Nguyen, T. N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M. N., Viswanathan, S., Penn, O., Bakken, T., Menon, V., Miller, J., Fong, O., Hirokawa, K. E., Lathia, K., Rimorin, C., Tieu, M., … Zeng, H. (2018). Shared and distinct transcriptomic cell types across neocortical areas. *Nature*, *563*(7729), 72–78.

Taylor, S. R., Santpere, G., Weinreb, A., Barrett, A., Reilly, M. B., Xu, C., Varol, E., Oikonomou, P., Glenwinkel, L., McWhirter, R., Poff, A., Basavaraju, M., Rafi, I., Yemini, E., Cook, S. J.,

Abrams, A., Vidal, B., Cros, C., Tavazoie, S., … Miller, D. M., 3rd. (2021). Molecular topography of an entire nervous system. *Cell*, *184*(16), 4329–4347.e23.

Tobet, S. A., Henderson, R. G., Whiting, P. J., & Sieghart, W. (1999). Special relationship of gamma-aminobutyric acid to the ventromedial nucleus of the hypothalamus during embryonic development. *The Journal of Comparative Neurology*, *405*(1), 88–98.

Tsukahara, S., Kakeyama, M., & Toyofuku, Y. (2006). Sex differences in the level of Bcl-2 family proteins and caspase-3 activation in the sexually dimorphic nuclei of the preoptic area in postnatal rats. *Journal of Neurobiology*, *66*(13), 1411–1419.

Tsukahara, S., Tsuda, M. C., Kurihara, R., Kato, Y., Kuroda, Y., Nakata, M., Xiao, K., Nagata, K., Toda, K., & Ogawa, S. (2011). Effects of aromatase or estrogen receptor gene deletion on masculinization of the principal nucleus of the bed nucleus of the stria terminalis of mice. *Neuroendocrinology*, *94*(2), 137–147.

Tsuneoka, Y., Tsukahara, S., Yoshida, S., Takase, K., Oda, S., Kuroda, M., & Funato, H. (2017). Moxd1 Is a Marker for Sexual Dimorphism in the Medial Preoptic Area, Bed Nucleus of the Stria Terminalis and Medial Amygdala. *Frontiers in Neuroanatomy*, *11*(March), 1–13.

Tulchinsky, D., Hobel, C. J., Yeager, E., & Marshall, J. R. (1972). Plasma estrone, estradiol, estriol, progesterone, and 17-hydroxyprogesterone in human pregnancy. I. Normal pregnancy. *American Journal of Obstetrics and Gynecology*, *112*(8), 1095–1100.

Turano, A., Osborne, B. F., & Schwarz, J. M. (2018). Sexual Differentiation and Sex Differences in Neural Development. *Current Topics in Behavioral Neurosciences*. https://doi.org/10.1007/7854_2018_56

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A. J., Burdziak, C., Moon, K. R., Chaffer, C. L., Pattabiraman, D., Bierie, B., Mazutis, L., Wolf, G., Krishnaswamy, S., & Pe'er, D. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell*, *174*(3), 716–729.e27.

van Veen, J. E., Kammel, L. G., Bunda, P. C., Shum, M., Reid, M. S., Massa, M. G., Arneson, D., Park, J. W., Zhang, Z., Joseph, A. M., Hrncir, H., Liesa, M., Arnold, A. P., Yang, X., & Correa, S. M. (2020). Hypothalamic estrogen receptor alpha establishes a sexually dimorphic regulatory node of energy expenditure. *Nature Metabolism*, *2*(4), 351–363.

Varshney, M. K., Inzunza, J., Lupu, D., Ganapathy, V., Antonson, P., Rüegg, J., Nalvarte, I., & Gustafsson, J.-Å. (2017). Role of estrogen receptor beta in neural differentiation of mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(48), E10428–E10437.

Vong, L., Ye, C., Yang, Z., Choi, B., Chua, S., Jr, & Lowell, B. B. (2011). Leptin action on GABAergic neurons prevents obesity and reduces inhibitory tone to POMC neurons. *Neuron*, *71*(1), 142–154.

Wade, K. H., Lam, B. Y. H., Melvin, A., Pan, W., Corbin, L. J., Hughes, D. A., Rainbow, K., Chen, J.-H., Duckett, K., Liu, X., Mokrosiński, J., Mörseburg, A., Neaves, S., Williamson, A., Zhang, C., Farooqi, I. S., Yeo, G. S. H., Timpson, N. J., & O'Rahilly, S. (2021). Loss-of-function mutations in the melanocortin 4 receptor in a UK birth cohort. *Nature Medicine*, *27*(6), 1088–1096.

Waki, H., Nakamura, M., Yamauchi, T., Wakabayashi, K.-I., Yu, J., Hirose-Yotsuya, L., Take, K., Sun, W., Iwabu, M., Okada-Iwabu, M., Fujita, T., Aoyama, T., Tsutsumi, S., Ueki, K., Kodama, T., Sakai, J., Aburatani, H., & Kadowaki, T. (2011). Global mapping of cell type-specific open chromatin by FAIRE-seq reveals the regulatory role of the NFI family in adipocyte differentiation. *PLoS Genetics*, *7*(10), e1002311.

Walker, D. M., Bell, X. M. R., Flores, X. C., Gulley, X. J. M., Willing, J., & Paul, M. J. (2017). Adolescence and Reward : Making Sense of Neural and Behavioral Changes Amid the Chaos. *The Journal of Neuroscience*, *37*(45), 10855–10866.

Walker, D. M., Kirson, D., Perez, L. F., & Gore, A. C. (2012). Molecular profiling of postnatal

development of the hypothalamus in female and male rats. *Biology of Reproduction*, *87*(6), 129.

Wang, L., Andersson, S., Warner, M., & Gustafsson, J.-Å. (2001). Morphological abnormalities in the brains of estrogen receptor β knockout mice. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(5), 2792–2796.

Wang, L., Andersson, S., Warner, M., & Gustafsson, J.-A. (2003). Estrogen receptor (ER)beta knockout mice reveal a role for ERbeta in migration of cortical neurons in the developing brain. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(2), 703–708.

Wang, L., Ozark, P. A., Smith, E. R., Zhao, Z., Marshall, S. A., Rendleman, E. J., Piunti, A., Ryan, C., Whelan, A. L., Helmin, K. A., Morgan, M. A., Zou, L., Singer, B. D., & Shilatifard, A. (2018). TET2 coactivates gene expression through demethylation of enhancers. *Science Advances*, *4*(11), eaau6986.

Wang, M., & Zhang, Y. (2021). Tn5 transposase-based epigenomic profiling methods are prone to open chromatin bias. In *bioRxiv* (p. 2021.07.09.451758). https://doi.org/10.1101/2021.07.09.451758

Wang, Q., Carroll, J. S., & Brown, M. (2005). Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Molecular Cell*, *19*(5), 631–642.

Wang, S., Sun, H., Ma, J., Zang, C., Wang, C., Wang, J., Tang, Q., Meyer, C. A., Zhang, Y., & Liu, X. S. (2013). Target analysis by integration of transcriptome and ChIP-seq data with BETA. *Nature Protocols*, *8*(12), 2502–2515.

Wei, Y.-C., Wang, S.-R., Jiao, Z.-L., Zhang, W., Lin, J.-K., Li, X.-Y., Li, S.-S., Zhang, X., & Xu, X.-H. (2018). Medial preoptic area in mice is capable of mediating sexually dimorphic behaviors regardless of gender. *Nature Communications*, *9*(1), 279.

Welch, J. D., Kozareva, V., Ferreira, A., Vanderburg, C., Martin, C., & Macosko, E. Z. (2019).

Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell*, *177*(7), 1873–1887.e17.

Wersinger, S. R., Sannen, K., Villalba, C., Lubahn, D. B., Rissman, E. F., & De Vries, G. J. (1997). Masculine sexual behavior is disrupted in male and female mice lacking a functional estrogen receptor alpha gene. *Hormones and Behavior*, *32*(3), 176–183.

Westberry, J. M., Prewitt, A. K., & Wilson, M. E. (2008). Epigenetic regulation of the estrogen receptor alpha promoter in the cerebral cortex following ischemia in male and female rats. *Neuroscience*, *152*(4), 982–989.

Westberry, J. M., Trout, A. L., & Wilson, M. E. (2010). Epigenetic regulation of estrogen receptor alpha gene expression in the mouse cortex during early postnatal development. *Endocrinology*, *151*(2), 731–740.

Westberry, J. M., & Wilson, M. E. (2012). Regulation of estrogen receptor alpha gene expression in the mouse prefrontal cortex during early postnatal development. *Neurogenetics*, *13*(2), 159–167.

Whaling, C. S., Nelson, D. A., & Marler, P. (1995). Testosterone-induced shortening of the storage phase of song development in birds interferes with vocal learning. *Developmental Psychobiology*, *28*(7), 367–376.

White, S. A., Livingston, F. S., & Mooney, R. (1999). Androgens Modulate NMDA Receptor–Mediated EPSCs in the Zebra Finch Song System. *Journal of Neurophysiology*, *82*(5), 2221–2234.

Wilson, M. E., Westberry, J. M., & Prewitt, A. K. (2008). Dynamic regulation of estrogen receptor-alpha gene expression in the brain: a role for promoter methylation? *Frontiers in Neuroendocrinology*, *29*(3), 375–385.

Wilson, M. E., Westberry, J. M., & Trout, A. L. (2011). Estrogen receptor-alpha gene expression in the cortex: Sex differences during development and in adulthood. *Hormones and Behavior*, *59*(3), 353–357.

Winick-Ng, W., Kukalev, A., Harabula, I., Zea-Redondo, L., Szabó, D., Meijer, M., Serebreni, L., Zhang, Y., Bianco, S., Chiariello, A. M., Irastorza-Azcarate, I., Thieme, C. J., Sparks, T. M., Carvalho, S., Fiorillo, L., Musella, F., Irani, E., Torlai Triglia, E., Kolodziejczyk, A. A., … Pombo, A. (2021). Cell-type specialization is encoded by specific chromatin topologies. *Nature*, *599*(7886), 684–691.

Woolley, C. S. (2007). Acute Effects of Estrogen on Neuronal Physiology. *Annual Review of Pharmacology and Toxicology*, *47*(1), 657–680.

Wu, M. V., Manoli, D. S., Fraser, E. J., Coats, J. K., Tollkuhn, J., Honda, S.-I., Harada, N., & Shah, N. M. (2009a). Estrogen masculinizes neural pathways and sex-specific behaviors. *Cell*, *139*(1), 61–72.

Wu, M. V., Manoli, D. S., Fraser, E. J., Coats, J. K., Tollkuhn, J., Honda, S.-I., Harada, N., & Shah, N. M. (2009b). Estrogen masculinizes neural pathways and sex-specific behaviors. *Cell*, *139*(1), 61–72.

Wu, M. V., & Tollkuhn, J. (2017a). Estrogen receptor alpha is required in GABAergic, but not glutamatergic, neurons to masculinize behavior. *Hormones and Behavior*, *95*, 3–12.

Wu, M. V., & Tollkuhn, J. (2017b). Estrogen receptor alpha is required in GABAergic, but not glutamatergic, neurons to masculinize behavior. *Hormones and Behavior*, *95*, 3–12.

Wu, S. J., Furlan, S. N., Mihalas, A. B., Kaya-Okur, H. S., Feroze, A. H., Emerson, S. N., Zheng, Y., Carson, K., Cimino, P. J., Keene, C. D., Sarthy, J. F., Gottardo, R., Ahmad, K., Henikoff, S., & Patel, A. P. (2021). Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nature Biotechnology*, *39*(7), 819–824.

Xu, X., Coats, J. K., Yang, C. F., Wang, A., Ahmed, O. M., Alvarado, M., Izumi, T., & Shah, N. M. (2012). Modular Genetic Control of Sexually Dimorphic Behaviors. *Cell*, *148*(3), 596–607.

Xu, Y., Huangyang, P., Wang, Y., Xue, L., Devericks, E., Nguyen, H. G., Yu, X., Oses-Prieto, J. A., Burlingame, A. L., Miglani, S., Goodarzi, H., & Ruggero, D. (2021). ERα is an RNA-binding

protein sustaining tumor cell survival and drug resistance. *Cell*. https://doi.org/10.1016/j.cell.2021.08.036

Yague, J. G., Muñoz, A., de Monasterio-Schrader, P., Defelipe, J., Garcia-Segura, L. M., & Azcoitia, I. (2006). Aromatase expression in the human temporal cortex. *Neuroscience*, *138*(2), 389–401.

Yao, G., Hu, S., Yu, L., Ru, Y., Chen, C. D., Liu, Q., & Zhang, Y. (2017). Genome-Wide Mapping of In Vivo ERα-Binding Sites in Male Mouse Efferent Ductules. *Endocrinology*, *158*(11), 3724–3737.

Yao, Z., Nguyen, T. N., van Velthoven, C. T. J., Goldy, J., Sedeno-Cortes, A. E., Baftizadeh, F., Bertagnolli, D., Casper, T., Crichton, K., Ding, S.-L., Fong, O., Garren, E., Glandon, A., Gray, J., Graybuck, L. T., Hirschstein, D., Kroll, M., Lathia, K., Levi, B., … Zeng, H. (2020). A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. In *Cold Spring Harbor Laboratory* (p. 2020.03.30.015214). https://doi.org/10.1101/2020.03.30.015214

Yao, Z., van Velthoven, C. T. J., Nguyen, T. N., Goldy, J., Sedeno-Cortes, A. E., Baftizadeh, F., Bertagnolli, D., Casper, T., Chiang, M., Crichton, K., Ding, S.-L., Fong, O., Garren, E., Glandon, A., Gouwens, N. W., Gray, J., Graybuck, L. T., Hawrylycz, M. J., Hirschstein, D., … Zeng, H. (2021). A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell*, *184*(12), 3222–3241.e26.

Yeo, G. S., Farooqi, I. S., Aminian, S., Halsall, D. J., Stanhope, R. G., & O'Rahilly, S. (1998). A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nature Genetics*, *20*(2), 111–112.

Yi, P., Wang, Z., Feng, Q., Pintilie, G. D., Foulds, C. E., Lanz, R. B., Ludtke, S. J., Schmid, M. F., Chiu, W., & O'Malley, B. W. (2015). Structure of a biologically active estrogen receptor-coactivator complex on DNA. *Molecular Cell*, *57*(6), 1047–1058.

Yu, G., Wang, L.-G., Han, Y., & He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: A Journal of Integrative Biology*, *16*(5), 284–

287.

Yu, G., Wang, L.-G., & He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* , *31*(14), 2382–2383.

Yu, G., Wang, L.-G., Yan, G.-R., & He, Q.-Y. (2015). DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* , *31*(4), 608–609.

Yun, M., Wu, J., Workman, J. L., & Li, B. (2011). Readers of histone modifications. *Cell Research*, *21*(4), 564–578.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., & Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137.

Zhou, H., Liu, J., Zhou, C., Gao, N., Rao, Z., Li, H., Hu, X., Li, C., Yao, X., Shen, X., Sun, Y., Wei, Y., Liu, F., Ying, W., Zhang, J., Tang, C., Zhang, X., Xu, H., Shi, L., … Yang, H. (2018). In vivo simultaneous transcriptional activation of multiple genes in the brain using CRISPR-dCas9-activator transgenic mice. *Nature Neuroscience*, *21*(3), 440–446.

Zhou, J. N., Hofman, M. A., Gooren, L. J., & Swaab, D. F. (1995). A sex difference in the human brain and its relation to transsexuality. *Nature*, *378*(6552), 68–70.

Zhu, C., Zhang, Y., Li, Y. E., Lucero, J., Behrens, M. M., & Ren, B. (2021). Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nature Methods*, *18*(3), 283–292.

Zhu, Q., Liu, N., Orkin, S. H., & Yuan, G.-C. (2019). CUT&RUNTools: a flexible pipeline for CUT&RUN processing and footprint analysis. *Genome Biology*, *20*(1), 192.

Zuloaga, D. G., Zuloaga, K. L., Hinds, L. R., Carbone, D. L., & Handa, R. J. (2014a). Estrogen receptor β expression in the mouse forebrain: age and sex differences. *The Journal of Comparative Neurology*, *522*(2), 358–371.

Zuloaga, D. G., Zuloaga, K. L., Hinds, L. R., Carbone, D. L., & Handa, R. J. (2014b). Estrogen

receptor β expression in the mouse forebrain: age and sex differences. *The Journal of Comparative Neurology*, *522*(2), 358–371.

Appendix

Method

Copolymerization of single-cell nucleic acids into balls of acrylamide gel

Siran Li,[1] Jude Kendall,[1] Sarah Park,[1] Zihua Wang,[1] Joan Alexander,[1] Andrea Moffitt,[1] Nissim Ranade,[1] Cassidy Danyko,[1] Bruno Gegenhuber,[1] Stephan Fischer,[1] Brian D. Robinson,[2] Herbert Lepor,[3] Jessica Tollkuhn,[1] Jesse Gillis,[1] Eric Brouzes,[4] Alex Krasnitz,[1] Dan Levy,[1] and Michael Wigler[1]

[1]*Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA;* [2]*Department of Pathology and Laboratory Medicine, Weill Medical College of Cornell University, New York, New York 10021, USA;* [3]*Department of Urology, New York University Langone Medical Center, New York, New York 10017, USA;* [4]*Department of Biomedical Engineering, Stony Brook University, Stony Brook, New York 11794, USA*

We show the use of 5′-Acrydite oligonucleotides to copolymerize single-cell DNA or RNA into balls of acrylamide gel (BAGs). Combining this step with split-and-pool techniques for creating barcodes yields a method with advantages in cost and scalability, depth of coverage, ease of operation, minimal cross-contamination, and efficient use of samples. We perform DNA copy number profiling on mixtures of cell lines, nuclei from frozen prostate tumors, and biopsy washes. As applied to RNA, the method has high capture efficiency of transcripts and sufficient consistency to clearly distinguish the expression patterns of cell lines and individual nuclei from neurons dissected from the mouse brain. By using varietal tags (UMIs) to achieve sequence error correction, we show extremely low levels of cross-contamination by tracking source-specific SNVs. The method is readily modifiable, and we will discuss its adaptability and diverse applications.

[Supplemental material is available for this article.]

Single-cell analyses are increasingly used for understanding the patterns of gene expression and genomic variation in complex populations of cells and tissues (Navin et al. 2011; Patel et al. 2014; Tirosh et al. 2016; Villani et al. 2017). Many droplet-based technologies have emerged as high-throughput ways to study DNA (Lan et al. 2017; Pellegrino et al. 2018) or RNA (Klein et al. 2015; Macosko et al. 2015) of single cells. However, these methods often lack the breadth of coverage (see Supplemental Text; Ding et al. 2019). Droplet merging and breakage give rise to cross-contamination. Moreover, some droplet-based methods suffer from inefficient use of samples. Therefore they are not the ideal choice for analyzing rare and valuable samples, such as cells from biopsy washes or cells microdissected from tissue samples. To address these and other needs, we developed and describe here a method that has advantages in coverage, quantitation, the efficient use of samples, sequence accuracy, and flexibility without sacrificing scalability. The set-up requires only inexpensive standard equipment and reagents, and the cost of preparing single-cell libraries is negligible compared with sequencing.

The central concept in this protocol has broad applicability. The underlying principle is the encapsulation of single cells or single nuclei in aqueous droplets containing acrylamide monomer in an oil emulsion, followed by conversion of each droplet into a ball of acrylamide gel (BAG) by polymerization. Primers containing 5′-Acrydite copolymerize with the acrylamide. Through annealing and extension, the information content of the cell is captured as nucleic acids covalently bound to the polyacrylamide matrix. After removing the oil, each BAG serves as an independent reaction vessel, accessible by diffusion in an aqueous environment to polymerases and other reagents. BAGs are then individually barcoded by split-and-pool methods, first used during the production of peptide libraries (Fodor et al. 1991), then used as a method to encode beads (Ohlmeyer et al. 1993), and finally for single-cell analysis (Cusanovich et al. 2015; Rosenberg et al. 2018). Our method has great flexibility. By varying designs of primers, enzymes, and conditions, the BAGs can be used as sources for libraries for single-cell DNA or RNA, or possibly even proteins. In this report, we show and characterize the applications for single-cell DNA copy number and RNA profiling from simple and complex mixed populations.

Results

Converting single cells into BAG libraries

Figure 1 illustrates our protocol, which we outline here. First, we create a suspended aqueous droplet in oil containing single-cell contents and reagents and then convert that droplet into a polyacrylamide bead. By using 5′-Acrydite primers, some of the contents of single cells become linked to the bead matrix. To achieve this, we use a single-cell DroNc device (Habib et al. 2017), with one stream (aqueous phase 1) carrying the single cells or nuclei and another stream (aqueous phase 2) carrying reagents, combining both as an aqueous droplet in oil. Aqueous phase 2 contains acrylamide monomers, bis-acrylamide cross-linker,

Corresponding author: wigler@cshl.edu
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.253047.119. Freely available online through the *Genome Research* Open Access option.

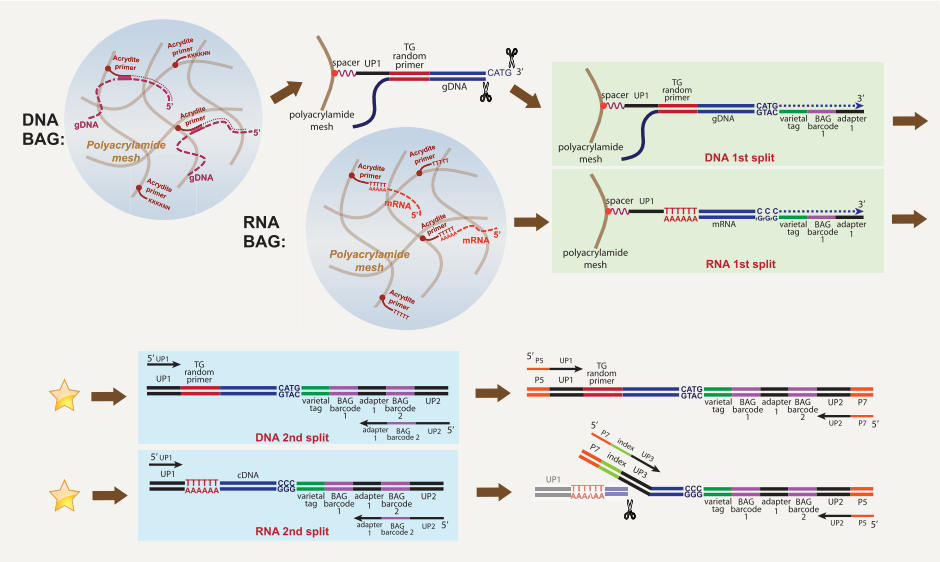

Figure 1. Schematic of single-cell DNA or RNA BAG-seq workflow. The star between first split and second split indicating the place where more cycles of split-and-pool can be added.

ammonium persulfate, 5′-Acrydite capture primers, and detergents in a buffer. For single-cell DNA analysis, we also include Proteinase K in aqueous phase 2. For RNA analysis, we include RNase inhibitor and omit Proteinase K. The oil phase contains TEMED, an accelerator of polymerization. During incubation in oil, the aqueous droplet forms a gel ball with the Acrydite primer covalently incorporated into the matrix.

For DNA, we use 5′-Acrydite primers made of essentially random Ts and Gs. We tested other primers, including *Alu* repeat sequence and random N primers, but these T and G primers performed best (Supplemental Fig. S1). We melt and cool to allow annealing to the 5′-Acrydite primers. For RNA, the Acrydite primers are poly(T) (see Methods; Fig. 1). We remove the oil, and the BAGs are subsequently processed in the aqueous phase. Annealing to primers is essential, as without linkage to the matrix, all nucleic acids leak out of the bead.

After removal from the oil, each BAG functions as a reaction vessel, with the nucleic acid tethered to the bead matrix. Once in the aqueous phase, we extend the primers on the captured templates, and thereby link copies of templates to the bead matrix. In subsequent steps, including split-and-pool, we add varietal tags (unique molecular identifiers) (Kivioja et al. 2012; Hicks et al. 2016) and BAG barcodes, but the method details differ if the initial template is DNA or RNA.

If the template is DNA, we make the second strand in the pool stage and cleave with a restriction endonuclease to make an adaptable end. In the first split, we use the cleavage site to add a varietal tag and the first split barcode. We pool the BAGs and redistribute them into wells for the second split, during which the second split barcode is added by PCR. We pool the PCR product from the second split and amplify using modified Illumina sequencing adapters to make the final sequencing library (for details, see Supplemental Method S1).

If the template is RNA, copying takes place in the first split yielding a DNA–RNA hybrid, and by using a "template switch oligo," the DNA strand acquires a varietal tag and first-split barcode. The second split can be performed as with DNA (see Methods), or additional cycles of split can be performed through a denaturation–hybridization–extension procedure and the final split is performed by PCR (see Supplemental Method S2). After pooling the PCR product, tagmentation followed by PCR is used to make the final sequencing library.

For DNA or RNA, sequencing libraries are prepared from pooled BAGs. The combination of first and second split barcodes gives almost all BAGs a unique bead barcode. We partition all the reads by this barcode. We then tally the captured templates with a given barcode by counting the nearly unique combination of varietal tag and captured sequence.

Not all detected barcodes derive from BAGs with single-cell contents. To determine which barcodes correspond to BAGs with cells, we only use barcodes with high read count. To do this, we plot a cumulative sum of the barcode counts, ordered by their magnitude. We typically observe a sharp inflection point, as illustrated in Supplemental Figure S2, and use the barcodes to the left of the inflection point.

In general, we observe efficient use of input cells (see Methods). Upon loading into the microfluidic device anywhere from 0.3 to 1 mL, we can recover up to 85% of cells in the final library. Some of the input fluid is retained by the device, so yields are higher for larger input volume.

Copy number profiles from mixed populations

As a demonstration of our method, we applied it to mixtures of four cell sources: three breast tumor cell lines (SK-BR-3, MCF-7, and BT-20) and one normal cell strain (SKN1). We computed the copy number profiles of each BAG using empirically normalized bins (see Methods). We display the results in Figure 2, A and B, at a resolution 20,000 bins or ~150 kbp per bin. After hierarchical clustering, we observe four clusters (Fig. 2A). One representative of each cluster is displayed as a conventional copy number profile in Figure 2B. Supplemental Figure S3 shows profiles from two BAGs with the SK-BR-3 pattern, illustrating the consistency of the method.

The method works on nuclei isolated from frozen tumor biopsies, and we illustrate this using previously published clinical material (Alexander et al. 2018). Figure 2C shows the pathology image of a region from a prostate with a Gleason 9 lesion, assessed by pathologists as 60% cancer. The BAG profiles from that region are displayed in a hierarchical cluster (Supplemental Fig. S4A). There are two clusters, one with 26 "normal" profiles and one with 39 tumor profiles. Figure 2, D through F, shows the representative copy number profiles from each clone of this region using BAG technology at

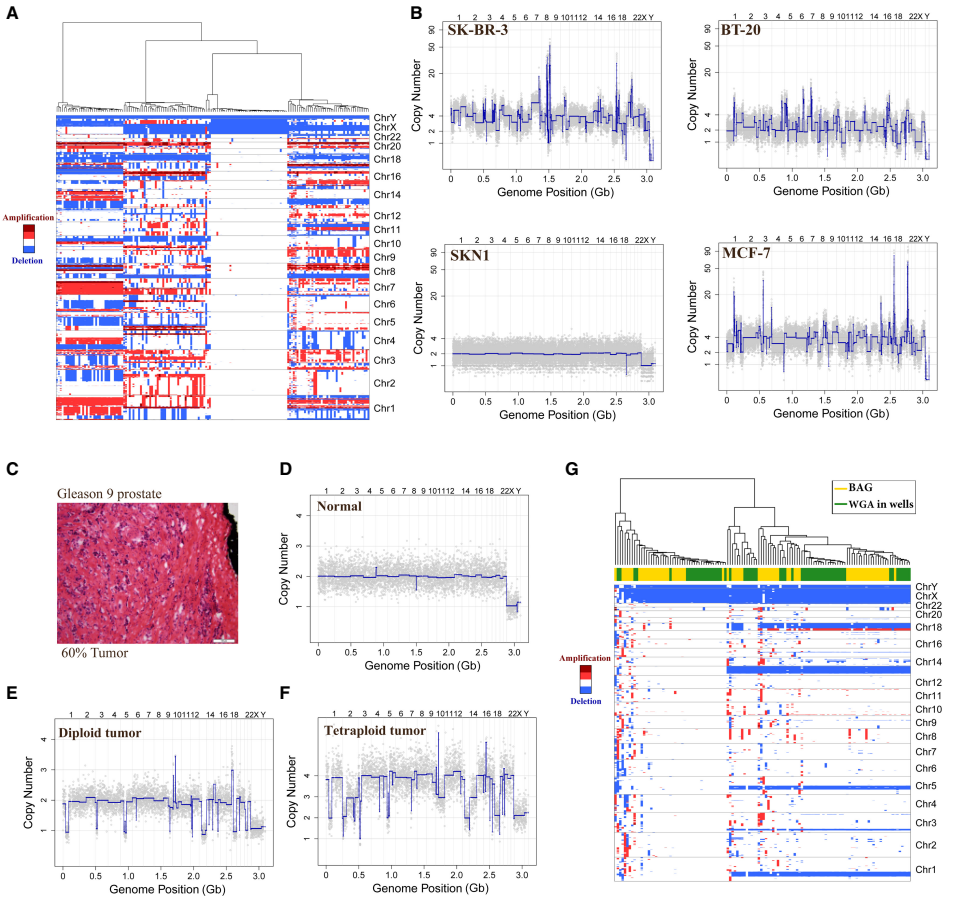

Figure 2. Copy number variation analysis of single-nucleus DNA (snDNA) BAG in cell lines and frozen prostate tumor. (*A*) Hierarchical clustering of four cell types SKN1, SK-BR-3, MCF-7, and BT-20 at a resolution of 20,000 bins (150 kbp per bin). Red indicates amplification, whereas blue indicates deletion. (*B*) The 20,000-bin copy number profiles from each of the four clusters in *A*. (*C*) Pathology image showing the region of Gleason 9 prostate cancer, which was estimated by pathologist as 60% tumor. Scale bar, 100 µm. (*D–F*) Representative snDNA BAG copy number profiles from this region: (*D*) a representative normal copy number profile; (*E*) a representative diploid tumor profile; and (*F*) a representative tetraploid tumor profile. (*G*) Hierarchical clustering of this region by combining data from both the BAG method and 96-well WGA method.

a resolution of 5000 bins (Supplemental Data S1, S2). Supplemental Figure S4B compares three BAG profiles with three single-nucleus profiles obtained from an earlier published whole genome amplification (WGA) method in 96-well plates. We combined profiles from our current (BAG) and previous (WGA) methods and display the hierarchical clustering in Figure 2G. The individual profiles obtained by the two methods are largely indistinguishable.

To illustrate that the method works with small amounts of precious sample, we also studied two biopsy wash samples from one patient. One sample was defined by the pathologist as benign (Fig. 3A; Supplemental Data S3, S4), and the other region was Gleason 6 cancer (Fig. 3B; Supplemental Data S5, S6). We examined 75 nuclei from the biopsy wash of this benign region, all normal profiles, and 269 nuclei from the Gleason 6 region. Hierarchical clustering trees from these two biopsy wash samples are shown in Figure 3, C and D. From the Gleason 6 region, we detected one major tumor clone as 35% of the cells, and among them possibly a minor clone (seven nuclei) that possesses all the features of the major clone but also has additional unique features

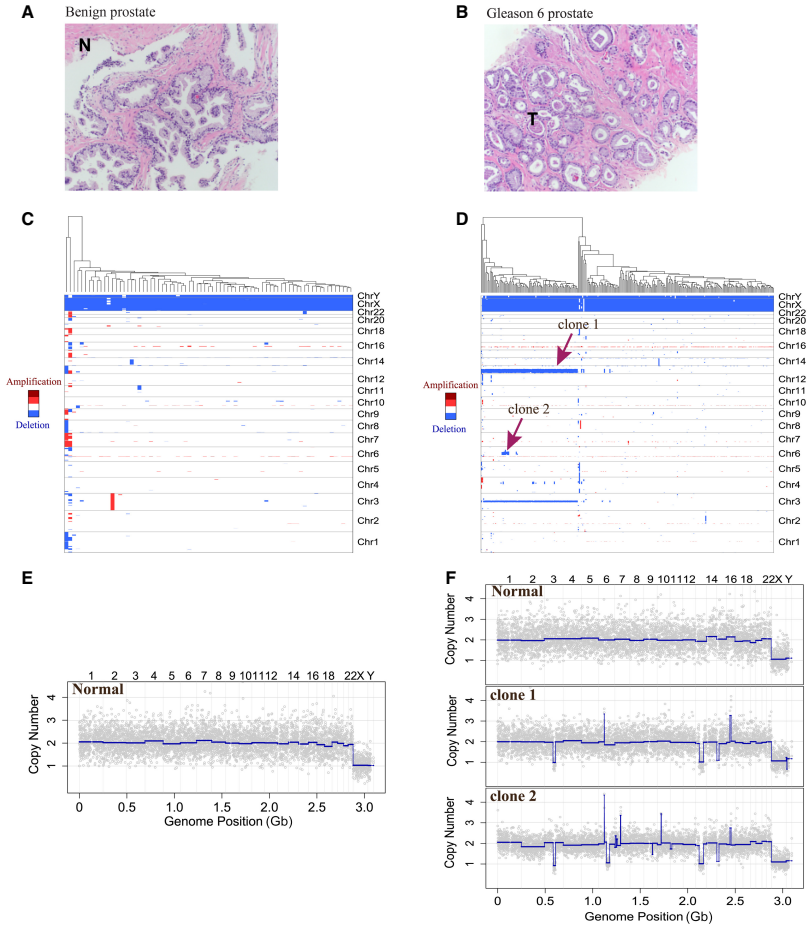

Figure 3. CNV study of prostate tumor biopsy wash samples from a benign region and a Gleason 6 cancer region. (*A*) A 20× magnification pathology image of a benign region of the prostate. (*B*) Pathology image of a Gleason 6 cancer region from the same patient at the same resolution. (*C*) Hierarchical clustering of biopsy wash sample from the benign region. (*D*) Hierarchical clustering of biopsy wash sample from the Gleason 6 region showing a normal clone and two tumor clones based on CNV patterns. Red arrows indicate the major (clone 1) and minor (clone 2) tumor clones. (*E*) A representative normal single-nucleus copy number profile from the biopsy wash of this benign region. (*F*) Representative single-nucleus copy number profiles from one normal clone and two tumor clones from the biopsy wash of the Gleason 6 cancer region.

(Fig. 3D). Representative copy number profiles of these two samples are shown in Figure 3, E and F.

Statistics of coverage

To measure the coverage of the method, we used SKN1, a diploid cell strain prepared from a healthy donor. The BAGs had 96 × 96 possible barcodes. We obtained 342 million read pairs, of which 84% were mapped to 88 BAGs with the largest counts (Supplemental Fig. S2A). From these 88, we characterized each BAG with respect to the numbers of read pairs, mapped reads pairs, uniquely tagged templates, and reads per template. We also characterize the total genome coverage per BAG. The data are found in Supplemental Table S1.

We sequenced about 300 million read pairs from the library made from 88 cells. We determined that 53% of paired-end reads could be properly mapped to the genome, meaning that both ends mapped within 2 kb apart had the proper orientation to each other, and Read 2 had the expected restriction endonuclease site. Because each template is captured as a single strand, with the NlaIII cleavage site marking the 3′ end, we could determine that 38.8% of the genome is captured from the plus strand, 38.9% captured from the minus strand, and 15.7% from both strands. In total, uniquely mapped reads cover a total of 60% of the human genome, with 40% of the genome mapping to at least two BAGs (Supplemental Fig. S5A).

We obtained a median number of 1.3 million reads per BAG, and a median absolute deviation (MAD) of 0.5 million. We have a multiplicity of about 3.5 reads per uniquely tagged template, yielding on average 492,490 unique templates. The median BAG barcode covered ~1.6% of the genome. These statistics are a function of read depth, so we sampled from 10% to 100% of reads and recomputed them. Supplemental Figure S5A shows total genome coverage, as well as the proportion of the genome seen in at least two BAGs as a function of reads sampled. Supplemental Figure S5B shows the shape of coverage for each of the 88 BAGs on downsampling, normalized to total counts at 100%. The 88 curves are very similar, as shown by the small error bars at each downsampling position, and indicate the limiting return of additional sequence.

Sequence error correction

Single-cell genome sequencing has been used for variant analysis (Wang et al. 2012; Xu et al. 2012; Zong et al. 2012). To use our method in this fashion requires an understanding of its sequence error rate. To measure error, we examined the single-cell sequence data from SKN1 for differences to the donor genome obtained from his blood DNA. We restricted analysis to those regions of the genome where the donor was well covered and homozygous to the human reference, and we sought variant sequence in the individual reads of the BAG libraries that were not reference bases. We then determined if the variant sequence was seen in multiple reads from the same template and in more than one BAG. We also examined the trinucleotide sequence context and the variant base.

Figure 4A summarizes error rates in single and multiple reads per uniquely tagged template. There are 64 trinucleotide contexts with three possible variants for the central nucleotide. Without error correction, some nucleotide contexts have low error (A or T to G, below 10^{-4}) and others high (G to C, about 10^{-3}). Using only multiple reads for a template, and then only when they are concordant, reduces some error rates on the order of 10-fold (e.g., A or T to C) (Fig. 4A). Some nucleotide contexts are not corrected by multiple reads in consensus (e.g., G to C, and C to G). We assume that if an error is not corrected by the concordance method, the error is due largely to initial template damage, for example, from depurination or deamination. Moreover, we observed that the error rate is lowered by reducing polymerization time or decreasing polymerization temperature. We infer this low-frequency damage may be induced by heat or the chemicals needed for polymerization, in

A

	C	C	G	G	T	T		A	A	G	G	T	T
AAA	2.8E-04	1.8E-05	1.1E-04	3.0E-05	2.0E-04	1.4E-04	ACA	1.3E-04	9.5E-05	3.2E-04	3.0E-04	4.3E-04	3.4E-04
AAC	3.6E-04	2.5E-05	1.3E-04	4.8E-05	2.1E-04	1.1E-04	ACC	9.5E-05	6.1E-05	3.8E-04	3.7E-04	2.7E-04	1.9E-04
AAG	3.5E-04	2.0E-05	1.3E-04	4.9E-05	3.2E-04	2.5E-04	ACG	2.3E-04	1.9E-04	3.0E-04	3.0E-04	4.9E-04	3.6E-04
AAT	3.6E-04	1.8E-05	1.1E-04	4.6E-05	2.0E-04	1.3E-04	ACT	1.7E-04	1.4E-04	3.1E-04	3.0E-04	2.6E-04	2.1E-04
CAA	3.2E-04	2.8E-05	1.7E-04	5.0E-05	4.7E-04	4.0E-04	CCA	2.7E-04	2.0E-04	2.8E-04	2.7E-04	5.2E-04	4.2E-04
CAC	3.3E-04	5.6E-05	1.4E-04	5.5E-05	3.9E-04	2.8E-04	CCC	2.3E-04	1.3E-04	3.7E-04	3.6E-04	3.8E-04	2.8E-04
CAG	3.2E-04	5.1E-05	1.6E-04	5.6E-05	5.3E-04	4.6E-04	CCG	4.6E-04	3.6E-04	3.1E-04	3.0E-04	7.5E-04	6.0E-04
CAT	4.0E-04	3.9E-05	1.4E-04	6.5E-05	5.4E-04	4.6E-04	CCT	3.4E-04	2.6E-04	2.5E-04	2.3E-04	3.1E-04	2.4E-04
GAA	3.8E-04	4.1E-05	1.5E-04	5.7E-05	4.5E-04	3.8E-04	GCA	1.8E-04	1.2E-04	3.8E-04	3.6E-04	4.9E-04	3.9E-04
GAC	4.1E-04	6.0E-05	1.7E-04	7.9E-05	4.0E-04	3.0E-04	GCC	1.4E-04	7.6E-05	4.4E-04	4.3E-04	3.1E-04	2.3E-04
GAG	2.9E-04	5.7E-05	2.0E-04	1.1E-04	5.9E-04	5.3E-04	GCG	3.2E-04	2.5E-04	3.6E-04	3.6E-04	7.1E-04	6.0E-04
GAT	4.0E-04	3.8E-05	1.5E-04	8.0E-05	4.3E-04	3.6E-04	GCT	2.7E-04	2.1E-04	3.4E-04	3.2E-04	2.9E-04	2.4E-04
TAA	3.8E-04	3.5E-05	1.3E-04	4.0E-05	2.4E-04	2.0E-04	TCA	3.3E-04	2.9E-04	1.7E-04	1.5E-04	4.8E-04	4.0E-04
TAC	3.9E-04	4.7E-05	1.4E-04	5.8E-05	2.7E-04	1.8E-04	TCC	2.7E-04	2.0E-04	2.9E-04	2.8E-04	3.2E-04	2.4E-04
TAG	4.1E-04	7.6E-05	1.6E-04	8.1E-05	4.1E-04	3.5E-04	TCG	5.2E-04	4.4E-04	2.2E-04	2.1E-04	7.4E-04	6.5E-04
TAT	4.1E-04	3.1E-05	1.2E-04	5.9E-05	2.9E-04	2.3E-04	TCT	4.1E-04	3.6E-04	1.7E-04	1.6E-04	2.9E-04	2.4E-04
	random	ec	random	ec	random	ec		random	ec	random	ec	random	ec
	A	A	C	C	T	T		A	A	C	C	G	G
AGA	1.2E-04	7.2E-05	7.6E-04	4.8E-04	4.0E-04	3.4E-04	ATA	1.3E-04	5.9E-05	3.8E-04	6.8E-05	4.1E-05	2.7E-05
AGC	1.4E-04	8.6E-05	2.2E-03	1.9E-03	1.8E-04	9.1E-05	ATC	1.3E-04	4.7E-05	4.7E-04	1.1E-04	8.6E-05	7.0E-05
AGG	1.6E-04	1.1E-04	1.5E-03	1.2E-03	2.5E-04	1.5E-04	ATG	1.8E-04	1.0E-04	5.2E-04	1.4E-04	7.8E-05	6.8E-05
AGT	1.2E-04	7.6E-05	1.4E-03	1.0E-03	2.0E-04	1.6E-04	ATT	1.1E-04	3.3E-05	4.6E-04	5.9E-05	6.0E-05	5.0E-05
CGA	3.8E-04	2.7E-04	6.9E-04	3.8E-04	1.7E-03	1.7E-03	CTA	1.9E-04	1.3E-04	4.4E-04	8.5E-05	3.5E-05	2.1E-05
CGC	4.0E-04	2.7E-04	2.1E-03	1.8E-03	4.9E-04	4.1E-04	CTC	2.0E-04	1.3E-04	4.8E-04	1.2E-04	5.0E-05	3.2E-05
CGG	5.0E-04	3.6E-04	1.3E-03	1.1E-03	8.7E-04	7.5E-04	CTG	2.8E-04	2.0E-04	4.7E-04	1.2E-04	5.0E-05	3.1E-05
CGT	2.9E-04	1.7E-04	1.3E-03	9.3E-04	8.2E-04	7.8E-04	CTT	1.7E-04	1.1E-04	5.2E-04	9.0E-05	3.8E-05	2.3E-05
GGA	1.6E-04	8.7E-05	1.1E-03	8.1E-04	1.1E-03	1.1E-03	GTA	2.3E-04	1.4E-04	4.7E-04	7.5E-05	4.7E-05	3.1E-05
GGC	1.7E-04	1.0E-04	2.9E-03	2.6E-03	3.3E-04	2.6E-04	GTC	1.8E-04	8.5E-05	5.1E-04	1.3E-04	1.6E-04	1.0E-04
GGG	2.2E-04	1.4E-04	2.0E-03	1.8E-03	6.2E-04	5.3E-04	GTG	2.5E-04	1.4E-04	4.0E-04	1.0E-04	1.1E-04	7.5E-05
GGT	1.4E-04	8.1E-05	1.7E-03	1.4E-03	6.3E-04	6.1E-04	GTT	1.7E-04	8.3E-05	4.8E-04	7.0E-05	1.0E-04	6.2E-05
TGA	2.0E-04	1.2E-04	9.9E-04	6.4E-04	8.2E-04	7.6E-04	TTA	1.6E-04	1.3E-04	4.9E-04	7.8E-05	3.7E-05	2.3E-05
TGC	2.4E-04	1.5E-04	2.9E-03	2.6E-03	2.8E-04	2.1E-04	TTC	2.1E-04	1.4E-04	5.5E-04	1.2E-04	5.7E-05	4.0E-05
TGG	3.0E-04	2.1E-04	1.8E-03	1.5E-03	5.0E-04	4.1E-04	TTG	2.7E-04	2.1E-04	5.4E-04	1.3E-04	6.1E-05	4.8E-05
TGT	2.0E-04	1.2E-04	1.8E-03	1.4E-03	4.4E-04	4.1E-04	TTT	1.8E-04	1.2E-04	5.1E-04	7.1E-05	5.3E-05	4.1E-05
	random	ec	random	ec	random	ec		random	ec	random	ec	random	ec

1.0E-03
7.5E-04
5.0E-04
2.5E-04
1.0E-04
5.0E-05
1.0E-05

B

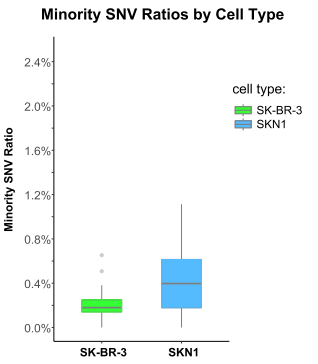

Figure 4. Sequence error correction (ec) and analysis of cross-contamination using error-corrected SNVs. (*A*) Comparison of error rates between random sampling and ec in trinucleotide context. The number in each box indicates the error rate and is colored by its intensity. The middle base in the trinucleotide context is the "source" base, and the single base on *top* of each column is the "destination" base. For each "destination" base, the first column corresponds to random sampling method, and the second column corresponds to the ec method. (*B*) Minority SNV ratios of SK-BR-3 nuclei and SKN1 nuclei from the four-nuclei mixing experiment using the ec method showing very low contaminations between BAGs.

particular the presence of persulfate, as has been previously noted (Wang et al. 2017).

One can further lower background errors to a large extent by demanding to see a variant as concordant in at least two BAGs. For applications such as studying cell lineage based on variants, one would require mutations shared by at least two cells. In Supplemental Table S2, we note the incidence of variant reads occurring at all homozygous reference positions on all chromosomes, the positions of which are seen in at least two BAGs and in each BAG with concordance. We note that if we restrict to variants appearing in at least two BAGs, then such variants occur with a frequency of 10^{-5} to 10^{-6}.

Let us call a variant that appears in two BAGs, in each with concordance, a "candidate" variant. We have thus put an impressive upper bound on the error rate for candidate variants. But this is not a least upper bound, because some of these variants might actually be somatic variation between these fibroblast cells and the blood DNA, and not sequence error. To explore this further, we consider the two genome strands since we can tell them apart. In this data set, if a position is captured on one strand in at least two BAGs (in each with concordance), ~40% of the time that position is captured on the opposite strand in at least one BAG. When the position of a candidate variant can be observed on the complementary captured strand, the complement of the variant is seen ~56% of the time. Thus, the candidate is most likely a somatic variant between the fibroblast and blood DNA, not sequencing error. Thus, one may make discovery of somatic mutation from single-cell genome data obtained using our current protocol.

Measuring cross-contamination

BAGs are semiclosed systems. They are porous and accessible by diffusion, but the trapped contents of the cell, once covalently linked to the polyacrylamide skeleton, will not leak out. Were this otherwise, we would not obtain the distinct copy number profiles that we see in mixed cell populations. However, we can now make this a quantitatively precise conclusion. To examine just how little cross-contamination does occur, we used SNV analysis from the four genome mixing experiments (illustrated in Fig. 2A), but looking only at the two genomes for which we had complete genome sequence data, SK-BR-3 and SKN1. We considered only "mutually distinct" variants and used the error-correction rules just described. Then we looked for consensus reads of distinct variants found in one or more BAGs that match the wrong genome (Supplemental Table S3). It is evident that there are very few distinct variants from the SK-BR-3 genome that are seen as consensus reads in SKN1 BAGs (110 out of 29,360 observed variant positions), and the reverse (261 from SKN1 out of 125,468 seen in SK-BR-3 BAGs). The average minority SNV ratio is 0.2% for SK-BR-3 nuclei and 0.4% for SKN1 nuclei (Fig. 4B), and the minority SNV ratios (0.2% to 0.4%) agree with the ploidy of SKN1 (diploid) versus SK-BR-3 (tetraploid).

Single-cell mRNA BAG libraries

BAG technology is flexible and can also be used for single-cell mRNA analysis using a 5′-Acrydite primer containing a 5′ PCR primer 1 (UP1) with a poly(T) tail (see Fig. 1). After capture and pooling, we split BAGs into a 96-well plate and used a reverse transcriptase with terminal deoxynucleotidyl transferase (TdT) function to add to the cDNA a varietal tag, first BAG barcode, and a common adapter sequence. We pooled BAGs again and split into another 96-well plate, where the second round of BAG barcodes with Primer 2 (UP2) was hybridized to molecules with the common adapter, and amplified by PCR (see Methods; Fig. 1; for details, Supplemental Methods). Following amplification, we fragmented the cDNA amplification product using the Nextera XT kit. After final amplification using the Nextera primer and P5-UP2 primer, "Read 1" corresponds to the 5′ end of the captured nucleic acid, the varietal tag, and the BAG barcodes and is used for deconvolution and counting of templates. "Read 2" is used for mapping to the genome.

To show the method works for RNA, we executed an assay on mixed cell populations of SKN1 and SK-BR-3 cells. From Read 2, we found that 70% of the bases were mapped to the exons (40% coding and 30% UTR), 17% were mapped to the introns, and 13% were mapped to the intergenic regions. These figures are similar to those we obtain using bulk RNA sequencing data. We used the cell-specific SNVs found in exons from these two cell sources to identify the captured cell, shown in Figure 5A. From 235 cells (Supplemental Fig. S2B), 233 (>99%) had a clear major source for its SNVs. We observe little to no cross-contamination judged by the SNV analysis. From these 233 BAGs, the median minority SNV ratio is 0.5% whether the minority source is SK-BR-3 or SKN1.

Two of the BAG barcodes appear to be associated with two cells. We expect this is the result of "barcode collision." The split-and-pool method does not guarantee that each BAG receives a unique barcode. The collision rate is driven by the number of BAGs and the number of possible barcodes. If there were $96 \times 96 \sim 10^4$ possible barcodes and if we picked barcodes at random 235 times, we expect the 2.9 barcodes to be picked twice, producing a collision. Collisions can be reduced by increasing the number of possible barcodes, as we will discuss.

The reads were then collapsed by their varietal tags to count how many uniquely tagged templates and genes were captured in each BAG. The median number of genes captured from 152 single SKN1 cells and 81 single SK-BR-3 cells was 6560 and 6542, respectively (Fig. 5B), and the median number of uniquely tagged templates captured from single SKN1 cells and SK-BR-3 cells was 37,429 and 51,128, respectively (Fig. 5C). These numbers compare favorably to what we obtained from previously used or commercially available methods.

The average number of reads per uniquely tagged template (RPT) was 5.9 for SK-BR-3 and 6.1 for SKN1. To estimate whether we have sequenced to saturation, we downsampled the reads and recomputed the unique templates and genes detected from these reads. From the shape of the downsampling curves (see Supplemental Fig. S6), more new templates would be observed by deeper sequencing of the libraries.

To study the consistency of gene expression between single cells of the same type, we performed PCA analysis after normalizing and centering the expression matrix. The first principal component (PC1) dominates and clearly separates SKN1 cells from SK-BR-3 cells (Fig. 5D). Two SKN1 and SK-BR-3 bulk RNA expression profiles from the conventional RNA sequencing method fit well among the single-cell RNA (scRNA) expression profiles (Fig. 5D). We calculated the correlation coefficients between gene expression and PC1 and plotted a heatmap composed of the top 20 positively correlated genes and the top 20 negatively correlated genes (Fig. 5E). Among the genes most correlated with PC1 in the fibroblast cells are collagen genes, and in the epithelial cancer cell line, SK-BR-3, are the keratin 8 (*KRT8*), *ERBB2*, and *ERBB2* signaling pathway genes.

To show scalability, we implemented three cycles of split-and-pool by adding a denaturation–hybridization–extension step

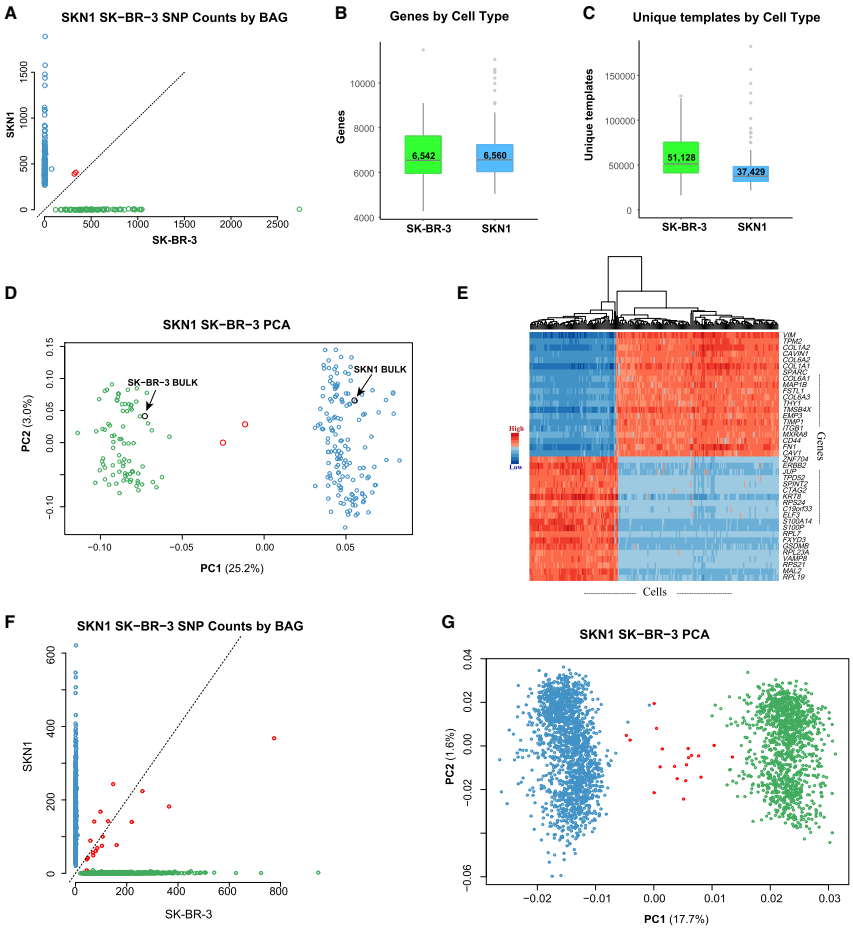

Figure 5. Single-cell RNA (scRNA) BAG showing high yield, low contamination, and consistent expression profiles. (*A–E*) A two-cycle split-pool experiment including 235 cells. (*F,G*) A three-cycle split-pool experiment including 2875 cells. (*A*) Scatter plot showing the number of SKN1-specific and SK-BR-3–specific SNVs found in exons for each cell. BAGs with majority SKN1 or SK-BR-3 SNVs are colored blue or green. Two (0.85% of total) BAGs without clear majority SNVs are labeled as red. (*B*) Boxplot showing the number of genes detected per cell. (*C*) Boxplot showing the number of unique templates captured per cell. (*D*) Scatter plot of PC1 versus PC2. The scRNA BAGs are colored by their majority SNVs defined in *A*. Two bulk RNA data sets for SKN1 and SK-BR-3 clusters with their respective single-cell data. The contribution of PC1 is more than eight times that of PC2 (25.2%/3.0%). (*E*) Heatmap based on 40 (20+, 20−) genes with the most positive and negative correlations to PC1. (*F*) Scatter plot showing the number of SKN1-specific and SK-BR-3–specific SNVs found in exons for each cell in the three-cycle split-pool experiments. Nineteen (0.66% of total) cells without clear majority SNVs are labeled as red. (*G*) PC1 versus PC2 from the 2875 cells in the three-cycle split-pool experiment illustrated in *F*.

after the first split-and-pool (see Supplemental Method S2). This generates around 1 million (96 × 96 × 96) different BAG barcodes. We showed performance in a mixture of SKN1 and SK-BR-3 cells surveying a total of approximately 3000 cells (Supplemental Fig. S2C). After counting cell line–specific SNPs and removing BAGs with fewer than 5000 unique tags, we identified 1663 BAGs as SKN1 cells and 1193 BAGs as SK-BR-3 cells. We identified 19 BAGs as having mixed identity, showing SNP ratios between 15% and 85% (Fig. 5F). The observed barcode collision rate is 0.66%. The SKN1 and SK-BR-3 populations are easily separable in the first component of PCA analysis (Fig. 5G).

One key advantage of BAG-seq is its high cell-capture efficiency, making it an ideal technique for studying a rare cell population. To show this important feature and also to show its performance

on complex cell populations, we performed single-nuclei RNA BAG-seq on a specific region of mouse brain. In this experiment, we collected a subset of the bed nucleus of the stria terminalis (BNST) called the principal nucleus (BNSTp) from five male mice and five female mice and pooled the nuclei from male or female mice, respectively. The area is sexually dimorphic in both mice and humans, with ~45% more neurons present in the BNSTp of male mice compared with females (Allen and Gorski 1990; Forger et al. 2004; Sokolowski and Corbin 2012; Welch et al. 2019). We studied 540 male mouse nuclei and 320 female mouse nuclei in this experiment. After clustering, we obtained eight cell populations (Fig. 6A–C). By comparing the number of nuclei between sexes in each cluster, we identified that in one of the eight clusters, the number of nuclei from males is significantly larger than that from females (Supplemental Table S4). We identified several marker genes that distinguish between each cluster and showed that this method distinguishes between neuronal and nonneuronal cells and between excitatory and inhibitory neurons (Fig. 6D).

Discussion

We have continued the evolution of methods for single-cell analysis. The first methods required isolating single cells into microwells (Navin et al. 2011). This idea was then extended by robotics (Wang et al. 2012; Gao et al. 2017; Gierahn et al. 2017). The method also evolved by using droplets in oil to isolate cells and then transferring the cell contents to beads coencapsulated with the cells (Klein et al. 2015; Macosko et al. 2015). Other methods used nuclei or fixed cells as vehicles for combinatorial indexing (Cusanovich et al. 2015; Rosenberg et al. 2018). Each of these methods has drawbacks, in the form of either expensive machinery, low yield per cell, limitations of Poisson sampling, requirements of nuclei isolation or cell fixation, or costly barcoded beads and reagents. In our present embodiment, we use aqueous droplets to isolate the cells, but the cell and its contents are polymerized and form the bead itself. By including Acrydite-modified primers, the nucleic acid sequences of the cell become bound to the acrylamide matrix. We call these cells in a ball of acrylamide gel (C-BAGs). Unlike other "cell bead" ideas (Tamminen and Virta 2015; Andor et al. 2018), in C-BAGs the first copy of nuclei acids is covalently bound to the gel matrix and individually tagged by a split-and-pool strategy.

The apparatus is inexpensive, and we estimate the cost per cell to be $0.50 when performed on the scale of a few hundred cells, with costs asymptotic to zero on a larger scale, which is a general feature for split-and-pool technologies. Unlike other droplet methods in which the barcoded bead is much smaller than the size

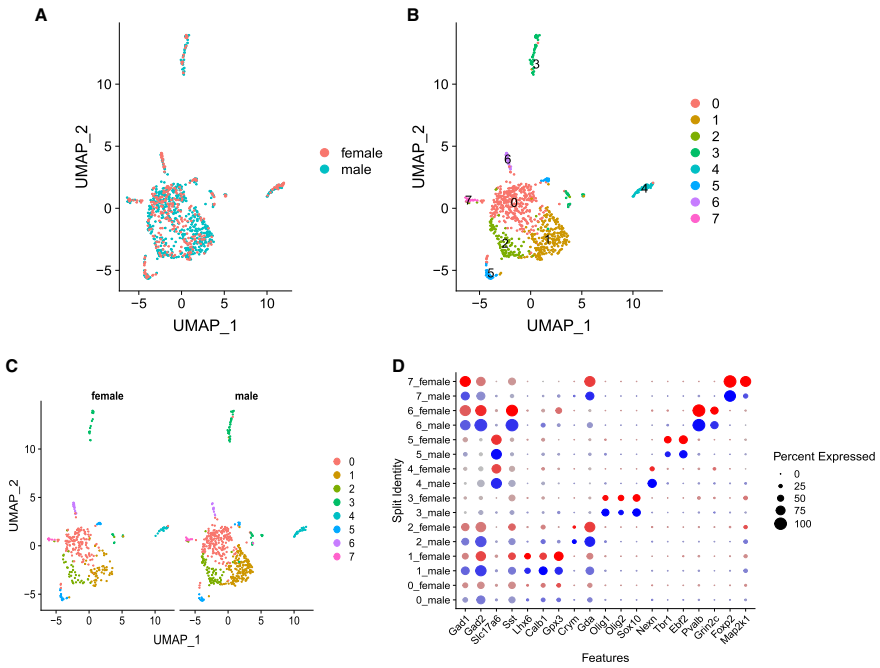

Figure 6. Comparison of single-nuclei RNA clusters distinguishing sexes. (*A*) UMAP clustering of 860 nuclei from brain BNSTp region, and colored by sex. (*B*) Eight clusters in *A* are distinguished and labeled using different colors. (*C*) Nuclei are split by sex. There are 540 nuclei from males and 320 nuclei from females. (*D*) Dotplot showing features expression across all clusters. The size of the dot indicates the percentage of cells within a cluster, and the brightness of color indicates the expression level in a cluster.

of a droplet, in our method, the capture bead becomes the size of the droplet so, in theory, capture should be more efficient. The nucleic acid capture reagents that fix the cell contents to the BAG skeleton can be custom designed. The BAG that forms is permeable, and its contents can then be processed easily and cheaply in an aqueous environment. We exploit this feature to generate BAG barcodes, subsequent to their formation, by split-and-pool synthesis, obviating the need for expensive reagents or kits. We observe virtually no cross-contamination between BAGs. We do have occasional barcode collisions, as expected. These can be minimized by increasing the number of barcodes as we showed in the Results.

We have illustrated the BAG-seq method for single-cell DNA. For this purpose, we used a non-self-annealing TG primer to trap the cell DNA. The sequence distribution is sufficiently uniform (Supplemental Fig. S1B), and we use varietal tags for accurate counting of the initial templates, so that with empirical bin normalization we obtain a genome copy number resolution equivalent to our best previous manual methods. There is virtually no cross-contamination between BAGs. Sequence error can be reduced with template varietal tags and the template concordance method so that single-cell variant calling is feasible.

With minor modifications of the capture primer and the barcoding method, the C-BAGs capture RNA and can produce scRNA libraries. At our sequence depth, about 250 cells per lane of Illumina NextSeq for the two-cycle SKN1-SB-BR-3 mixing BAG-seq experiment, we capture about 50,000 unique templates per cell and on the order of 6500 expressed genes per cell. Deeper sequencing would yield higher numbers (Supplemental Fig. S6). By integrating over all BAGs, we obtain the full spectrum of genes found by bulk sequencing. We observe little to no cross-contamination, as judged by SNV analysis. For comparison to other scRNA technologies, we performed a scRNA BAG-seq using HEK293 cells (Supplemental Fig. S7). We show that the BAG-seq method exceeds the number of genes and unique templates captured per cell and has fewer barcode collisions compared with other high-throughput methods (Supplemental Text). Many other variations in methods for RNA trapping, extension, and barcoding might be tested in the future such as encapsulating beads in BAGs. The present method should keep pace with or exceed future developments in commercially available methods and at lower costs.

The protocol makes efficient use of sample cells since it does not depend on two simultaneous Poisson events: one sampling cells; the other, beads. Thus, we can examine a few hundred cells, which is important if the sample is scarce and precious. By the same token, running devices in parallel could generate millions of C-BAGs and, after split-and-pool tagging, still achieve a one-to-one correspondence between tag and cell.

As this last example illustrates, BAG technology is very flexible. Although we have presented two applications, many other possible uses of BAGs merit development. For some of these, alternative ways of linking the cell nucleic acids, or proteins, to the gel matrix will need to be developed. In principle, BAGs could capture both DNA and RNA for making dual libraries from single cells. Capturing RNA and/or protein, followed by reaction with fluorescent probes, could enable enrichment for BAGs containing the content of rare subtypes of cells, for example, by fluorescence-activated cell sorting (FACS). In principle, the BAGs are reusable, and one can select BAGs for deeper analysis as desired. It should be possible to freeze and store them until one is ready for sequencing. We might be able to reencapsulate BAGs into oil if further biochemical steps required reisolation.

Methods

Samples

In the experiments, we used a normal skin fibroblast cell strain (SKN1) and three breast tumor cell lines (SK-BR-3, BT-20, and MCF-7). Clinical specimens were as previously cited (Alexander et al. 2018). In particular, a tissue biopsy was obtained from a patient (COR003.GS9.2) undergoing radical prostatectomy (RP) at New York–Presbyterian/Weill Cornell Medical Center. The Gleason score (GS) at RP was GS9 (4 + 5). One-millimeter-diameter cores of frozen tissue were placed into a sterile tube and maintained on dry ice to transfer to Cold Spring Harbor Laboratory (CSHL) for further processing. Clinical and pathological data were collected and maintained in a database curated by the Weill Cornell Medical College Center's Prostate Cancer Biobank. Also, tissue biopsies were obtained from a patient (NYU005) undergoing prostate biopsy at the Smilow Comprehensive Prostate Cancer Center (SCPCC) at NYU Langone Medical Center. This patient underwent a systematic transrectal ultrasound prostate biopsy (TRUS-Bx) and an MRI-TRUS fusion-targeted biopsy (TBx). Individual cores of prostate tissue were placed in site-separated vials filled with 5 mL of sterile wash buffer (1× PBS containing 0.5% BSA [Thermo Fisher Scientific B14] and 2 mM EDTA) and gently inverted several times for 60 sec to enhance exfoliation of prostate cells. After inversion, prostate cores were removed from the wash solution using disposable single-use sterile forceps and transferred to site-separated containers with formalin fixative for histological processing and pathological evaluation. The biopsy GS for NYU005 was GS7 (3 + 4). The presence of perineural invasion (PNI) was noted in the final diagnostic pathology report. Prostate biopsy washings were kept on wet ice for 1–2 h during transfer to CSHL, where the cell suspensions were briefly centrifuged to pellet the cells and lysed using NST-DAPI buffer as described in the previous study (Navin et al. 2011). This patient underwent a RP in which the GS7 at biopsy was downgraded to GS6 at RP.

Mouse experiments were performed under the guidelines of the CSHL Institutional Animal Care and Use Committee (IACUC). Esr1cre (Lee et al. 2014) and ROSA26CAG-Sun1/sfGFP (INTACT) (Mo et al. 2015) mice were obtained from the Jackson Laboratory and crossed to generate Esr1cre; INTACT animals.

Isolating nuclei

Nuclei were prepared from four sources. First, nuclei from cell lines were isolated using nuclei EZ Lysis buffer from a nuclei EZ prep kit (Sigma-Aldrich NUC101). Second, nuclei from frozen tissue were isolated using the protocol described by Habib et al. (2017). The nuclei were stored in ice-cold PBS-BSA (0.05%) buffer. Third, nuclei from biopsy washings were isolated using NST-DAPI buffer. Nuclei from prostate biopsy washes were prepared by gently centrifuging washings at 1000 rpm for 5 min to pellet the exfoliated cells followed by removal of supernatant and addition of 1.0 mL of NST-DAPI buffer to the cell pellet. All nuclei suspensions were filtered through a 35-µm cell strainer before flow sorting. Single nuclei regardless of ploidy were sorted into an Eppendorf tube using a BD Biosystems SORP flow cytometer. Fourth, nuclei were isolated from the mouse BNSTp as described previously (Mo et al. 2015) with minor modifications. BNSTp tissue was microdissected from 500-mm brain sections after rapid decapitation of anesthetized animals. Tissue was pooled from five P14 male and female animals heterozygous for the Esr1cre and INTACT alleles. The tissue was dounce homogenized 15× with a loose pestle in a glass homogenizer containing homogenization medium (250 mM sucrose, 25 mM KCl, 5 mM $MgCl_2$, 20 mM Tricine-KOH, 1 mM DTT, 0.15 mM spermine, 0.5 mM spermidine, 1× EDTA-free

protease inhibitor cocktail [Roche], 0.2 U/µL RNaseOUT [Thermo Fisher Scientific], adjusted to pH 7.8). We added 0.3% IGEPAL CA-630, and the tissue was further dounced 5× with a tight pestle. The homogenate was then filtered through a 40-µm strainer and mixed 1:1 with 50% OptiPrep solution (Millipore Sigma), prepared in dilution buffer (150 mM KCl, 30 mM $MgCl_2$, 120 mM Tricine-KOH, adjusted to pH 7.8). The homogenate was underlaid with 5 mL of 30% and 40% OptiPrep solution, respectively, and centrifuged at 9200 RPM for 18 min at 4°C in an ultracentrifuge using a Beckman SW-28 swinging bucket rotor. After centrifugation, ~2 mL of nuclei solution was removed from the 30%–40% OptiPrep interface by tube puncture using a 3-mL syringe attached to an 18-gauge needle.

Single-nucleus/cell encapsulation

Nuclei were encapsulated into aqueous droplets using a 70-µm-channel DroNc-seq device (Nanoshift), and the droplets were polymerized into BAGs. The droplet formation requires two aqueous phases and one oil phase. The oil phase recipe and acrylamide monomer concentration were adapted from the published protocol (Zilionis et al. 2017), but 5% surfactant (Ran Technologies 008-FluoroSurfactant) was used in HFE-7500 oil (Oakwood Chemical 051243). Aqueous phase 1 is nuclei or cells in PBS-BSA (0.05%) buffer.

For the single-cell DNA experiments and for the purpose of forming a BAG, 1 mL of aqueous phase 2 contains 180 µL of 40% AA/bis-acrylamide solution (Sigma-Aldrich A9926), 129 µL of 40% Acrylamide solution (Sigma-Aldrich A4058), 160 µL of 500 µM Acrydite-random (TG) primer ("/5ACryd//iSp18/TGTGT TGGGTGTGTTTGGKKKKKKKGKKKKKKKKNN," Integrated DNA Technologies), 100 µL of 1M Tris-HCl (pH 7.5), 50 µL of 0.5M EDTA, 10 µL of 20% sarkosyl (Sigma-Aldrich L7414), 20 µL of Proteinase K (Sigma-Aldrich P4850), 20 µL of 0.1M DTT, 60 µL of freshly made 10% APS, and 271 µL of H_2O. The output of the device was collected into 1.5-mL Eppendorf tubes preloaded with 300 µL mineral light oil. Upon collection, a milky-colored droplet layer forms between the heavy and light transparent oil layers in the tube.

The tubes were then incubated overnight at 50°C. After incubation, the bottom heavy oil layer was replaced by a new oil layer consisting of FC-40 oil (Sigma-Aldrich F9755) with 5% surfactant. The tubes were then transferred to a heating block preset at 95°C for melting DNA. After heating for 12 min, the tubes were incubated for 1 h at 55°C and then for another 10 min at room temperature to allow annealing. The top and bottom layers of oil were removed, and the BAGs were washed twice using a mixture of 600 µL 6× SSC solution (Thermo Fisher Scientific 15557036) and 150 µL of 1H,1H,2H,2H-Perfluoro-1-octanol (Sigma-Aldrich 370533). The BAGs were then washed once using 6× SSC solution and once with 1× NEBuffer 2.

For the scRNA experiments, 1 mL of aqueous phase 2 contains 180 µL of 40% AA/bis-acrylamide solution, 129 µL of 40% Acrylamide solution, 80 µL of 500 µM Acrydite-poly(T) primer ("/5ACryd//iSp18/AAGCAGTGGTATCAACGCAGAGTNNWNNN STTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT," Integrated DNA Technologies), 70 µL of 1M Tris-HCl (pH 7.5), 50 µL of 0.5M EDTA, 10 µL of 20% sarkosyl (Sigma-Aldrich L7414), 100 µL of SUPERase•In RNase inhibitor (Thermo Fisher Scientific AM2696), 100 µL of 10% NP-40 (Thermo Fisher Scientific 28324), 30 µL of 0.1M DTT, 60 µL of freshly made 10% APS, and 191 µL of H_2O.

The tubes were incubated for 2.5 h at room temperature for polymerization and then transferred to a heat block for 5 min at 50°C. The tubes were then incubated for another 10 min at room temperature. The BAGs were collected as previously described for DNA experiments but were washed using 5× RT buffer instead of 1× NEBuffer 2 in the DNA experiment.

DNA BAG first split-and-pool barcoding

Immediately after the BAGs were collected and washed, a linear extension step using DNA polymerase I (NEB M0210) was performed for 1.5 h at room temperature and then for 30 min at 37°C. That was followed by a 3′ exonuclease treatment using exonuclease I (NEB, M0293) to chew up unused single-stranded primers. The DNA in BAGs were cut using NlaIII (NEB R0125) to generate a 3′CATG overhang. The BAGs were then distributed into a 96-well plate. In each well, we added dNTP (Sigma-Aldrich 11814362001) and well-specific primers with the following structure: 5′ adapter 1–BAG barcode 1–varietal tag–CATG. The last two bases (T and G) are locked nucleic acid (LNA) to improve annealing to the four-base overhang. We then perform what we call a ligation-extension reaction, in which ligation and extension occur in the same reaction. To be specific, after the BAGs were first incubated with 1 µL of 100-µM well-specific primers in 1× Quick Ligation buffer (NEB M2200) in a total volume of 14 µL for 20 min at 4°C with rotation, 0.75 µL of quick ligase (NEB M2200) and 0.75 µL of Klenow fragment (NEB M0212) in 4.5 µL of 1× Quick Ligation buffer were added. The plate was incubated for another 20 min at 4°C with rotation and then for 30 min at 10°C to promote ligation. Following that, the plate was then rotated for 40 min at room temperature and then for another 40 min at 37°C to allow linear extension. The reaction was stopped by high-EDTA buffer STOP-25. The beads were then pooled together and washed by STOP-10 (Supplemental Method S1).

RNA BAG first split-and-pool barcoding

For scRNA experiments, immediately after we collected and washed the gel balls from oil, we distributed them into 96-well plates for reverse transcription using Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific EP0751). During reverse transcription, each well has a well-specific template-switch-oligo with the structure: 5′ adapter 1–BAG barcode 1–varietal tag–rGrGrG. Here "rG" means a ribonucleotide guanine base. The BAG barcode 1 sequence and varietal tag sequence were copied to the cDNA by reverse transcriptase. The reaction was stopped by high-SDS buffer TE-SDS (Supplemental Method S2) and then pooled and washed by STOP-10 buffer. An exonuclease reaction was followed to chew up the free primers in the BAGs.

The last split-and-pool barcoding

For both single-cell DNA and RNA experiments, the last round of BAG barcode was added in the same way. Many strategies can be used to perform split-and-pool (Cao et al. 2017; Vitak et al. 2017; Rosenberg et al. 2018; Cao et al. 2019), and additional rounds of BAG split-barcodes can be added based on the common sequence from the previous round (Supplemental Method S2). Here we used PCR to add 96 different BAG barcodes in the last split. In each well, there is a universal PCR primer and a well-specific primer containing different barcodes. BAGs from the last split-and-pool were evenly distributed into 96 wells. DNA BAGs were amplified using NEBNext ultra II Q5 master mix (NEB M0544). RNA BAGs were amplified using KAPA HiFi HotStart ReadyMix (Roche KK2602). The PCR product was pooled together and purified using AMPure XP magnetic beads (Bechman Coulter A63881).

Single-nuclei WGA method

We used single-nuclei DNA data previously generated from a previous WGA method (Alexander et al. 2018). Briefly, single nuclei were deposited into individual wells in a 96-well plate and amplified using GenomePlex WGA4 kit (Sigma-Aldrich WGA4-50RXN) according to the manufacturer's instructions. WGA DNA was sonicated using a Covaris focus acoustics system. The Covaris E210 300 ± sonication program generated WGA DNA inserts of the desired length, ~300 bp (range 200–400 bp), for library construction. Customized well-specific barcodes were ligated to the fragments in each well. Multiple libraries were combined into pools ranging from eight to 12 libraries to pools of 96 libraries for 76-bp single-read sequencing on single lanes of Illumina's GAIIx and HiSeq flowcells, respectively. The first 30 bases of each read were trimmed to remove any WGA primer sequence.

Bulk RNA sequencing

Total RNA from each cell line was extracted using Direct-zol RNA MiniPrep plus kit (Zymo Research R2070). mRNA isolation was performed using a NEBNext Poly(A) mRNA magnetic isolation module (NEB E7490). Sequencing library was prepared using a NEBNext ultra II directional RNA library prep kit for Illumina (NEB E7760).

Estimating cell capture efficiency

First, we measured the cell solution volume and cell concentration using a cell counter. We loaded a one:one ratio of cell solution (aqueous phase 1) and aqueous phase 2 into the microfluidic device. After collecting and polymerizing the droplets, we measured the volume of the recovered aqueous phase containing BAGs, as well as the total number of BAGs by counting under a stereo-microscope. At this time point, we measured the cell occupancy rate in BAGs by DAPI staining and counting through a fluorescent microscope. BAGs with a cell in them were much brighter than empty BAGs in DAPI channel. By using the cell occupancy rate, we estimated the total number of recovered cells in BAGs.

We also measured the number of BAGs by counting under a stereo-microscope before we made the cDNA library, and recorded the number of cells in the final sequencing library with high read counts. There was little loss of BAGs during the split-and-pool procedure, and essentially every BAG with a cell yielded a single-cell library.

Read alignment

For DNA data, Illumina sequence files were preprocessed before mapping to remove reads that do not conform to expectation, to retain those that do, and to trim away sequences not needed for mapping. In particular, read pairs were removed if the CATG NlaIII cut site sequence was not in base positions 31 to 34 on Read 2 with a maximum of one mismatched base. Reads 1 and 2 were then both trimmed to remove 3′ bases that matched Illumina adapter sequence or universal primer sequence. Read pairs with both ends at least 100 bases after trimming were retained for mapping. Reads were mapped to the UCSC hg19 reference genome using HISAT2 version 2.1.0 (Kim et al. 2015) with default parameters except for the following: -3 25, -X 2000, --no-spliced-alignment. Aligning reads to genome assembly GRCh38 would not impact our study, as the updates to the genome assembly are primarily related to population variation and filling of gaps (Schneider et al. 2017). The identifying sequence in the DNA protocol is on Read 2. The BAG barcode, base positions one through six appended to base positions 22 through 26, and the varietal tag, base positions 27 through 30, were appended the read ID in the FASTQ file. This allowed read pair identity to be tracked through subsequent processing.

The first three steps in the RNA data processing pipeline are as follows: (1) select reads with valid sequence structure on Read 1 to be included in the analysis, (2) extract identifying BAG barcode and varietal tag sequences and append these to the read ID in the FASTQ files, and (3) map reads. The different RNA libraries in the paper were processed in slightly different ways.

Step 1 (check sequence structure)

For the SKN1, SK-BR-3 2 split-pool library, reads with GGG in positions 38 through 40 and positions seven through 25 matching the primer sequence, AGTGGAAAAGGAAGGTGGT, up to two mismatches, were included in further processing. For the HEK293 and the BNSTp libraries, reads with GGG in positions 38 through 40, allowing up to one mismatch, and positions one through six and 26 through 31 having valid BAG barcodes were included for further processing. For the SKN1, SK-BR-3 3 split-pool sample, reads with GGG in positions 67 through 69, allowing up to one mismatch, and positions one through six, 30 through 35, and 55 through 60 having valid BAG barcodes were included for further processing.

Step 2 (extract BAG barcode and varietal tag)

For the SKN1, SK-BR-3 2 split-pool, HEK293, and BNSTp libraries, BAG barcodes are positions one through six and 26 through 31 from Read 1. Varietal tags are positions 32 through 37 and 43 through 48 from Read 1. For the SKN1, SK-BR-3 3 split-pool, BAG barcodes are positions one through six, 30 through 35, and 55 through 60 from Read 1. Varietal tags are positions 61 through 66 and 72 through 77 from Read 1.

Step 3 (map to reference genome)

For the SKN1, SK-BR-3 2 split-pool, SKN1, SK-BR-3 3 split-pool, and HEK293 libraries, 76 bases of Read 2 were mapped to the UCSC hg19 reference genome with UCSC refGene annotations for known splice sites using HISAT2 version 2.1.0 with default parameters. For the BNSTp library, 50 bases of Read 2 were mapped to the UCSC mm9 reference genome with UCSC refGene annotations for known splice sites using HISAT2 version 2.1.0 with default parameters. The updates in GRCm38 (mm10) from mm9 mainly filled gaps and finished the sequence of repetitive genomic regions, so the genic sequence in mm9 is sufficiently complete for the analysis presented here.

Copy number analysis

For the purpose of copy number analysis, the genome was divided into either 5000 or 20,000 bins. The bin boundaries were determined empirically from the data to generate a uniform distribution for the number of tags mapping to each bin assuming a constant copy number (Supplemental Fig. S1B). For this purpose, all the reads from the good SKN1 single-cell libraries were used.

Bincount data for all BAG barcodes with at least 100,000 unique tags, based on varietal tag and mapping location, were normalized by first computing log(bincount + 1)/mean(bincount + 1) and then further normalized for GC content by lowess normalization in R programming language (R Core Team 2018) with parameter $f = 0.05$. The normalized bincount vectors were then segmented using DNAcopy version 1.50.5 (Olshen et al. 2004). DNAcopy parameters used were alpha = 0.02, nperm = 1000, undo.SD = 0.5, and min.width = 3. Copy number heatmaps

were made using the heatmap function in R using the segmented bin values. The distance function used was "manhattan," and the hierarchical clustering agglomeration method used was "ward.D2."

Individual genome plots were made after estimating ploidy. After normalization, the segmentation vectors have a mean value of one. Ploidy was estimated by multiplying these vectors by 1.5, 1.55, 1.6, …, 4.5 and using the multiplier that minimizes the sum of square error from the multiplied vector to the multiplied vector rounded to nearest integers. This multiplier is the ploidy estimate. The segmentation vector is multiplied by the ploidy estimate to get a segmentation that has as much as possible of the genome on segments close to integer values. For SKN1, the frozen tumor sample, and the biopsy washes, these segmentation values are the y-axis values on the genome plots. For the cell line samples in the four-nuclei mixing library, there are copy number values ranging from zero to almost 100. To visualize these values more clearly, the y-axis values are $\log(y+1)$ with horizontal lines corresponding to copy numbers 1, 2, 3, 4, 20, 50, and 80 displayed.

Base calling error rate analysis

To assess the sequence error in libraries made by the BAG method, we used assays based on SKN1. This is a cell strain for which we have the whole-genome sequence (WGS) of the donor from his blood DNA. Illumina sequence data from the donor were mapped to the reference genome using Bowtie 2 version 2.3.2-legacy (Langmead and Salzberg 2012) with indel realignment using GATK version 1.6-13 (McKenna et al. 2010). Mapped reads were selected for error rate analysis provided they had a read mapping quality at least 30, bases called with a base quality at least 30, genome positions with a read depth of at least 20, and no SNPs or indels called in this region. This resulted in 1.25 Gb (1.25 billion bases) used for this analysis. At these 1.25 Gb, the BAG data were evaluated for mismatches to the reference genome, as follows. We first determined "template read sets" as the set of reads sharing identical BAG barcode, varietal tags, and map position. We called positions from template read sets of at least two members and then only if at least 80% of reads agreed. Most of the nonconsensus base positions had exactly two reads, one of which did not match the reference genome.

Cell source–specific SNVs

Cell-specific SNVs were called as follows. Illumina WGS data for the cell sources SKN1 and SK-BR-3 were mapped using Bowtie 2 version 2.3.2-legacy. Variants are called using reads with mapping quality of at least 30 where a nonreference base (with base quality at least 20) is seen at least three times and in at least 5% of reads covering this position. A variant is considered to be specific to one cell line if the variant is not seen in the other cell line where there are at least 12 reads of mapping quality at least 30 and requiring a base quality at least 20. There were 617,608 SNVs specific to SK-BR-3 and 561,443 specific to SKN1. BAG data were evaluated at these SNV sites after removing six SNV sites prone to anomalous mapping artifacts. These positions were Chr 1: 569874, Chr 6: 58777419, Chr 6: 58778584, Chr 6: 58779097, Chr 7: 61969087, and Chr 10: 42385520.

Gene expression analysis

For bulk RNA analysis, reads that mapped completely within exons for a transcript in UCSC refGene annotations were counted and assigned to that transcript. Values at reads per kilobase per million reads (RPKM) were then computed for all transcripts. To get RPKM values for a gene with multiple transcripts, the transcript with the highest RPKM value was used.

For single-cell analysis, reads that mapped with ≥50% of the read length within exons for a transcript in UCSC refGene annotations were counted for that transcript's gene.

RNA principal component analysis and clustering

For the SKN1, SK-BR-3 experiments, expression level values were first normalized by the mean for each sample, and then the log of the expression level +1 was centered for each gene using the center function in R with the scale parameter = F. The principal components were computed using the "prcomp" function in R with parameters center = T and scale = T. The coefficients of the first two principal components were plotted in a scatter plot with points colored according to their cell type as assessed by SNV analysis. The genes selected for the heatmap were the 40 genes with the most extreme correlations (20 most positive and 20 most negative) to the loadings on principal component 1. The data were clustered by sample and displayed using the R heatmap function with clustering parameters, distance function "Euclidean," and hierarchical clustering agglomeration method "complete."

For the mouse BNSTp experiment, we used R package Seurat v3 (Stuart et al. 2019) and UMAP clustering method (McInnes et al. 2018) to cluster digital expression data for the 860 nuclei. We used the default parameters for the Seurat package and used the first 30 PCA components for the UMAP function.

All aspects of the research were performed with Institutional Review Board approval.

Data access

The raw sequencing reads from the cell lines in this study have been submitted to the NCBI Sequence Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra/) under accession number PRJNA566441.

Acknowledgments

We thank L. Pachter, J. Gehring, and A. Tambe at Caltech for guidance on Drop-seq apparatus. We thank B. Isaila and E. Becher at the New York University Langone Medical Center for assistance of pathology information. We also thank A. Brady for cell culture; P. Moody and C. Kanzler for cell sorting; E. Ghiban for Illumina NextSeq sequencing; and J. Preall, C. Wunderlich, B. Ma, B. Stillman, and X. Wang for helpful discussion. This work was supported by grants to M.W. from the Simons Foundation, Life Sciences Founders Directed Giving-Research (award numbers 519054), and The Breast Cancer Research Foundation (award number 18-174). This work was also supported by the National Institutes of Health (NIH) grant 5U01MH105991 to Daniel Geschwind, the NIH–National Cancer Institute grant (R01 CA181595) to E.B., and R01 MH113628 and Simons Foundation SFARI600568 to J.T. This work was performed with assistance from CSHL Shared Resources, which are funded, in part, by the Cancer Center Support grant 5P30CA045508.

Author contributions: S.L. and M.W. designed the study. J.K. developed analysis programs. S.L., S.P., Z.W., J.A., and C.D. performed the experiments. S.L., J.K., and M.W. performed analysis. D.L., A.K., A.M., N.R., S.F., J.G., and E.B. assisted in analyzing the data. B.G. and J.T. provided mouse brain nuclei. B.D.R. and H.L. provided clinical samples. S.L., J.K., and M.W. wrote the manuscript.

References

Alexander J, Kendall J, McIndoo J, Rodgers L, Aboukhalil R, Levy D, Stepansky A, Sun G, Chobardjiev L, Riggs M, et al. 2018. Utility of single-cell genomics in diagnostic evaluation of prostate cancer. *Cancer Res* **78:** 348–358. doi:10.1158/0008-5472.CAN-17-1138

Allen LS, Gorski RA. 1990. Sex difference in the bed nucleus of the stria terminalis of the human brain. *Journal of Comparative Neurology* **302:** 697–706. doi:10.1002/cne.903020402

Andor N, Lau BT, Catalanotti C, Kumar V, Sathe A, Belhocine K, Wheeler TD, Price AD, Kang M, Stafford D. 2018. Joint single cell DNA-Seq and RNA-Seq of gastric cancer reveals subclonal signatures of genomic instability and gene expression. bioRxiv doi:10.1101/445932

Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ, et al. 2017. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357:** 661–667. doi:10.1126/science.aam8940

Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, et al. 2019. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566:** 496–502. doi:10.1038/s41586-019-0969-x

Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, Steemers FJ, Trapnell C, Shendure J. 2015. Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348:** 910–914. doi:10.1126/science.aab1601

Ding J, Adiconis X, Simmons SK, Kowalczyk MS, Hession CC, Marjanovic ND, Hughes TK, Wadsworth MH, Burks T, Nguyen LT, et al. 2019. Systematic comparative analysis of single cell RNA-sequencing methods. bioRxiv doi:10.1101/632216

Fodor SP, Read JL, Pirrung MC, Stryer L, Lu AT, Solas D. 1991. Light-directed, spatially addressable parallel chemical synthesis. *Science* **251:** 767–773. doi:10.1126/science.1990438

Forger NG, Rosen GJ, Waters EM, Jacob D, Simerly RB, De Vries GJ. 2004. Deletion of *Bax* eliminates sex differences in the mouse forebrain. *Proc Natl Acad Sci* **101:** 13666–13671. doi:10.1073/pnas.0404644101

Gao R, Kim C, Sei E, Foukakis T, Crosetto N, Chan L-K, Srinivasan M, Zhang H, Meric-Bernstam F, Navin N. 2017. Nanogrid single-nucleus RNA sequencing reveals phenotypic diversity in breast cancer. *Nat Commun* **8:** 228. doi:10.1038/s41467-017-00244-w

Gierahn TM, Wadsworth MH II, Hughes TK, Bryson BD, Butler A, Satija R, Fortune S, Love JC, Shalek AK. 2017. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods* **14:** 395–398. doi:10.1038/nmeth.4179

Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M, Choudhury SR, Aguet F, Gelfand E, Ardlie K, et al. 2017. Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* **14:** 955–958. doi:10.1038/nmeth.4407

Hicks J, Navin N, Troge J, Wang Z, Wigler M. 2016. *Varietal counting of nucleic acids for obtaining genomic copy number information*. U.S. patent no. US20140065609A.

Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12:** 357–360. doi:10.1038/nmeth.3317

Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, Taipale J. 2012. Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods* **9:** 72–74. doi:10.1038/nmeth.1778

Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW. 2015. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161:** 1187–1201. doi:10.1016/j.cell.2015.04.044

Lan F, Demaree B, Ahmed N, Abate AR. 2017. Single-cell genome sequencing at ultra-high-throughput with microfluidic droplet barcoding. *Nat Biotechnol* **35:** 640–646. doi:10.1038/nbt.3880

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359. doi:10.1038/nmeth.1923

Lee H, Kim D-W, Remedios R, Anthony TE, Chang A, Madisen L, Zeng H, Anderson DJ. 2014. Scalable control of mounting and attack by $Esr1^{+}$ neurons in the ventromedial hypothalamus. *Nature* **509:** 627–632. doi:10.1038/nature13169

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. 2015. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161:** 1202–1214. doi:10.1016/j.cell.2015.05.002

McInnes L, Healy J, Melville J. 2018. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. arXiv:1802.03426 [stat.ML].

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20:** 1297–1303. doi:10.1101/gr.107524.110

Mo A, Mukamel EA, Davis FP, Luo C, Henry GL, Picard S, Urich MA, Nery JR, Sejnowski TJ, Lister R, et al. 2015. Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* **86:** 1369–1384. doi:10.1016/j.neuron.2015.05.018

Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, et al. 2011. Tumour evolution inferred by single-cell sequencing. *Nature* **472:** 90–94. doi:10.1038/nature09807

Ohlmeyer M, Swanson RN, Dillard LW, Reader JC, Asouline G, Kobayashi R, Wigler M, Still WC. 1993. Complex synthetic chemical libraries indexed with molecular tags. *Proc Natl Acad Sci* **90:** 10922–10926. doi:10.1073/pnas.90.23.10922

Olshen AB, Venkatraman E, Lucito R, Wigler M. 2004. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5:** 557–572. doi:10.1093/biostatistics/kxh008

Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV, Curry WT, Martuza RL, et al. 2014. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344:** 1396–1401. doi:10.1126/science.1254257

Pellegrino M, Sciambi A, Treusch S, Durruthy-Durruthy R, Gokhale K, Jacob J, Chen TX, Geis JA, Oldham W, Matthews J, et al. 2018. High-throughput single-cell DNA sequencing of acute myeloid leukemia tumors with droplet microfluidics. *Genome Res* **28:** 1345–1352. doi:10.1101/gr.232272.117

R Core Team. 2018. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org/.

Rosenberg AB, Roco CM, Muscat RA, Kuchina A, Sample P, Yao Z, Graybuck LT, Peeler DJ, Mukherjee S, Chen W, et al. 2018. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360:** 176–182. doi:10.1126/science.aam8999

Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen H-C, Kitts PA, Murphy TD, Pruitt KD, Thibaud-Nissen F, Albracht D, et al. 2017. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res* **27:** 849–864. doi:10.1101/gr.213611.116

Sokolowski K, Corbin JG. 2012. Wired for behaviors: from development to function of innate limbic system circuitry. *Front Mol Neurosci* **5:** 55. doi:10.3389/fnmol.2012.00055

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM III, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive integration of single-cell data. *Cell* **177:** 1888–1902.e21. doi:10.1016/j.cell.2019.05.031

Tamminen MV, Virta MP. 2015. Single gene-based distinction of individual microbial genomes from a mixed population of microbial cells. *Front Microbiol* **6:** 195. doi:10.3389/fmicb.2015.00195

Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, Rotem A, Rodman C, Lian C, Murphy G, et al. 2016. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352:** 189–196. doi:10.1126/science.aad0501

Villani A-C, Satija R, Reynolds G, Sarkizova S, Shekhar K, Fletcher J, Griesbeck M, Butler A, Zheng S, Lazo S, et al. 2017. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* **356:** eaah4573. doi:10.1126/science.aah4573

Vitak SA, Torkenczy KA, Rosenkrantz JL, Fields AJ, Christiansen L, Wong MH, Carbone L, Steemers FJ, Adey A. 2017. Sequencing thousands of single-cell genomes with combinatorial indexing. *Nat Methods* **14:** 302–308. doi:10.1038/nmeth.4154

Wang J, Fan HC, Behr B, Quake SR. 2012. Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* **150:** 402–412. doi:10.1016/j.cell.2012.06.030

Wang Y, Liu C, Hong T, Wu F, Yu S, He Z, Mao W, Zhou X. 2017. Application of ammonium persulfate for selective oxidation of guanines for nucleic acid sequencing. *Molecules* **22:** 1222. doi:10.3390/molecules22071222

Welch JD, Kozareva V, Ferreira A, Vanderburg C, Martin C, Macosko EZ. 2019. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell* **177:** 1873–1887.e17. doi:10.1016/j.cell.2019.05.006

Xu X, Hou Y, Yin X, Bao L, Tang A, Song L, Li F, Tsang S, Wu K, Wu H, et al. 2012. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* **148:** 886–895. doi:10.1016/j.cell.2012.02.025

Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, Klein AM, Mazutis L. 2017. Single-cell barcoding and sequencing using droplet microfluidics. *Nat Protoc* **12:** 44–73. doi:10.1038/nprot.2016.154

Zong C, Lu S, Chapman AR, Xie XS. 2012. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338:** 1622–1626. doi:10.1126/science.1229164

Received May 24, 2019; accepted in revised form November 13, 2019.