Title
Identifying and Dissecting Regulatory Elements that Drive Drug Response and Human Evolution

Author
Ryu, Ann Hane

Publication Date
2017

Peer reviewed|Thesis/dissertation

Identifying and Dissecting Regulatory Elements that Drive Drug Response and Human Evolution

by

Ann Hane Ryu

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Sciences and Pharmacogenomics

in the

GRADUATE DIVISION

Dedication

This work is dedicated to my mom Woosoon Ryu, who taught me through example how to be independent, patient, and resilient when facing hardship. Thank you for all the sacrifices you made for my future and for always being proud of me. I hoped that you would be at my graduation but I didn't finish in time.

I'd also like to dedicate this to all the other shoulders I've leaned on throughout graduate school, especially my Dad, Uncle 삼촌, Aunt 이모, Jane, and Phil:

you've been my rock for the past few years. I couldn't have done this without you all.

Acknowledgements

I would like to thank my advisors, Nadav Ahituv and Katherine Pollard, for their support and guidance over the past five years. I feel incredibly grateful for the opportunity to have two advisors who mentored me on both the experimental and computational sides of my projects, which really pushed me to understand all possible ways to optimize an experiment and gain a wider range of skills. They showed through example what it means to maintain healthy collaborations and how to communicate effectively with many groups on a very interdisciplinary project. In addition to my appreciation for their academic support, I am also extremely grateful for their patience and understanding during my time of personal hardship. I hope to emulate both of them in the way they balance their family life with their career. I'd also like to thank John Rubenstein, for giving me feedback and reassurance during my somewhat frequent and long thesis committee meetings.

I'd also like to thank all the members of both the Ahituv and Pollard labs. Both labs are full of driven individuals who have helped me grow intellectually, scientifically, and emotionally through many discussions and their company inside and outside of the lab. I'd like to thank all of my fellow graduate students, especially Genevieve Erwin Haliburton, Tara Friedrich, Aram Avila-Herrera, Nir Oksenberg, Julie Vandermeer, Svetlana Lyalina, Kathleen Keough, and Serena Tamura, for guiding me and giving me feedback on my code, presentations, experimental designs, and career. I would also like to give many thanks to all of the post-doctoral scholars in (or affiliated to) both labs that have either taught me

the skills I needed for my research or contributed heavily to my projects, especially Walter Eckalbar, Fumitaka Inoue, Alex Williams (who is not technically a post doc but was super helpful and taught me a lot about coding!) and Sean Thomas. I'd like to thank Robin Smith for providing me the opportunity to intern with his team at 23andme. In terms of preserving my sanity and confidence throughout my PhD studies, I can't give enough thanks to Navneet Matharu for the many clarifying discussions about science, career, family, and life over the past four years. Another special thanks goes to Alex Pollen, a former post-doc from the Kriegstein lab who transitioned successfully to a faculty position at UCSF. His constant encouragement and insightful feedback on all aspects of my human evolution project was integral to the success of that project.

Contributions

Walter Eckalbar taught me how to implement bioinformatics tools to perform alignment of RNA-seq and ChIP-seq reads to the human genome. He performed the differential H3K27ac peak analysis as well as the revision of the gene expression analysis in DESeq2 for the antibiotics project. Anat Kreimer performed the analysis to find correlations between differentially expressed genes and differentially active regulatory elements for the antibiotics project. Walter, Fumitaka Inoue (Taka), Navneet Matharu, Nadja Makki, and Marcelo Rizzatti Luizon, all provided assistance in teaching me how to perform a successful ChIP-seq experiment. Robin Smith provided the preliminary data that gave rise to the design of this project. Nadav Ahituv conceived of the original ideas for the antibiotics study just described, and obtained the grant funding to support this research.

Taka provided protocols, training, and expertise for working with human iPSCs and differentiating them into neural and glial progenitor cells, as well as optimizing the rate of infection of the HAR MPRA lentiviral library in NPCs and GPCs. Alex Pollen also provided training and expertise for working with chimpanzee and human iPSCs, differentiating them, and characterizing them through immunostains and single cell RNA-sequencing. Beatriz Alvarado performed the library preps and analysis of the human and chimpanzee neural and glial single cell data. The human iPSC lines were provided by the Yamanaka lab and the chimpanzee iPSCs were engineered, characterized, and provided by the Kriegstein lab. Designing oligos for the HAR MPRA library was mostly

performed by Alex Williams. The quantification and modeling of the HAR MPRA data was performed by Katie Pollard, Sean Thomas, and Sean Whalen. Hassan Samee and Kathleen Keough performed the HAR TFBS motif and GWAS SNP analyses, respectively. The troubleshooting, cloning, and processing of raw MPRA data were performed by Taka, Beth Martin, Martin Kircher, and Jay Shendure. Katherine Pollard and Nadav Ahituv conceived of the original ideas for the HAR MPRA study, and obtained the grant funding to support this research.

Identifying and Dissecting Regulatory Elements that Drive Drug Response and Human Evolution

A. Hane Ryu

Abstract

Gene regulation is known to contribute to the wide diversity of biological differences between cell types, individuals, and species. Enhancers are regulatory elements that determine when, where, and how much a protein-coding gene is expressed in every tissue. They contain short motifs called transcription factor binding sites and function through chromatin remodeling and DNA looping to activate transcription of their target genes. Due to their role in activating gene expression across tissues and developmental timepoints, disruption in enhancer function can lead to disease and morphological differences between species. By characterizing enhancers we can learn how genetic changes in non-coding DNA alter gene function and ultimately use this knowledge to diagnose and treat disease.

Using RNA-sequencing and chromatin immunoprecipitation (ChIP) sequencing, I identified genome-wide antibiotic-induced changes in gene expression and regulation in HepG2 cells, a human liver cell line. More specifically, I found 209 genes responsive to penicillin-streptomycin (PenStrep), a commonly used cell culture antibiotic cocktail, and 9,514 H3K27ac peaks that were PenStrep-responsive. I also performed a massively parallel reporter assay (MPRA) to quantify enhancer activity of conserved DNA elements that have rapidly evolved in humans called human accelerated regions (HARs) in human

and chimpanzee iPSC-derived neural and glial progenitor cells. This method allowed us to detect novel brain enhancers with species-specific function and dissect the regulatory architecture of these enhancers. Our results showed that the *cis* features or sequence level changes were greater drivers of differences in enhancer activity than the *trans* environment, or cell species and cell stage, that these sequences were tested in. My research sheds insight on the regulatory code driving drug response to common antibiotics, as well as the uniquely human patterns in early neurodevelopment.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 What are enhancers and why are they important?

Enhancers are non-coding, regulatory sequences that determine the spatiotemporal expression of protein-coding genes, and therefore play an important role in vertebrate development (Visel et al., 2009). They contain short DNA motifs called transcription factor binding sites (TFBS) that attract transcription factors, which recruit histone modifying factors such as histone acetyltransferase (HAT) or histone methyltransferase (HMT). Enhancers also require DNA looping to activate transcription of their target genes' promoters through chromatin remodeling factors and the cohesion complex for their DNA (Schmidt et al., 2010; Euskirchen et al., 2011; Faure et al., 2012; Li et al., 2006; Ptashne et al., 1986). Comparing to other types of regulatory elements (i.e. repressors, promoters, insulators, barriers), the widest functional diversity between tissues has been seen for enhancers, suggesting that enhancers have an especially important role in determining tissue specificity (Heintzman et al., 2009).

1.2 How do we detect enhancers and quantify their regulatory activity?

Conservation of DNA across species is often used as an indicator of functional importance. Previous groups have identified potential regulatory elements by searching for highly conserved, non-coding regions of DNA across genomes (Harmston et al., 2013; Alfoldi et al., 2013). In addition to evolutionary conservation, transcription factor binding and histone marks are also used to annotate active enhancer regions. Researchers often use chromatin immunoprecipitation (ChIP) sequencing using antibodies that target transcription factors (e.g. ATF3, FOXP1, CBP/p300) or histone modifications (e.g. H3K27ac) known to mark the presence of an active enhancer region

(Sandman et al., 2007; Boon et al., 2012). Enhancers and other regulatory elements are also known to function within topologically active domains (TADs), which are discrete regions of the genome that have high interaction frequencies within them and are bordered by low interaction regions called TAD boundaries (Matharu et al., 2015). The interactions between enhancers and their target promoters within TAD boundaries can be identified through chromatin capture methods (i.e. 3C, 4C, 5C, Hi-C), in which chromatin is cross-linked using formaldehyde so that DNA regions within spatial proximity are linked together with protein complexes (Shlyueva et al., 2014).

Since enhancers are known to function independently of distance and orientation relative to their target gene, an enhancer sequence can be tested in various *in vitro* and *in vivo* assays that indicate whether that enhancer is active in a specific cell type and/or timepoint in development. Traditionally, one of the methods used to test enhancers in cell culture is the luciferase assay, which allows for quantification of enhancer activity in an episomal context. The *in vivo* method is the lacZ reporter assay in mouse embryos, which indicates the spatial domain and time point in development at which the enhancer is active. For both of these methods however, a negative result doesn't rule out the possibility that the candidate enhancer could be active in a different cell type or developmental time point. More recently, new methods that parallelize the reporter assay now allow researchers to test and quantify the enhancer activity of thousands of candidate enhancers simultaneously in the same cell type or organism (Melkinov et al., 2012; Patwardhan et al., 2012).

1.3 The role of enhancers in evolution, development and disease

Enhancers tend to be evolutionarily conserved but in general they evolve faster than coding regions (Pollard et al., 2006), which suggests that changes in regulatory DNA play an important role in evolution. A key study providing evidence of the role of

enhancers in evolutionary divergence between mouse and bat showed that swapping of the mouse *Prx1* limb enhancer with the bat sequence results in longer forelimb bones (Cretekos et al., 2008). The evolutionary role of enhancers has been observed for other species as well. Another study between humans and other primates (as well as rodents) revealed that the developmental absence of penile spines in the human lineage is due to a loss of the enhancer for the *androgen receptor (AHR)* gene since the split between humans and chimpanzees (McLean et al., 2011). Other evidence of the role of enhancers in human evolution has been shown in a study that discovered non-coding regions that contain human-specific nucleotide substitutions in the brain developmental transcription factor *neuronal PAS domain containing protein 3* (*NPAS3*) with enhancer activity in early neurodevelopment (Kamm et al., 2013).

Beyond their role in evolution, enhancers have also been associated with a number of developmental disorders, diseases, and variation in drug response. Some notable examples of enhancer mutations that cause developmental disorders include those regulating sonic hedgehog (*SHH)* (preaxial polydactlyl) (Lettice et al., 2003), SRY-box 9 (*SOX9)* (Pierre Robin Syndrome) (Benko et al., 2009), and T-box 5 (*TBX5)* (congenital heart disease) (Smemo et al., 2012). The Ahituv lab has previously shown that a wide range of structural variants or nucleotide changes within enhancers can lead to phenotypic differences such as limb malformations (VanderMeer et al., 2011). Enhancer variants have also been implicated in common diseases. Many genome-wide association studies (GWAS) have identified thousands of loci that affect the susceptibility to common diseases- most of these GWAS risk loci don't contain protein-coding variants, suggesting a role for regulatory variation especially in enhancers (Maurano et al., 2012; Ongen et al., 2014; Farh et al., 2015). In addition to playing a role in development and disease, enhancers have also been linked to inter-individual variation in response to drugs (Luizon et al., 2016). As our knowledge of the regulatory

architecture of enhancers progresses, the closer we are to understanding the molecular basis for evolution, development, disease, and drug response.

Chapter 2

Genome-wide identification of changes in gene expression and regulation induced by cell culture antibiotics

2.1 The known impact of antibiotics on gene expression and regulation

One of the most common cautionary measures taken during *in vitro* studies is the use of antibiotics while culturing cells in order to avoid bacterial contamination. Standard cell culture protocols listed by the American Type Culture Collection (ATCC) explicitly entail the addition of antibiotics, such as penicillin-streptomicin (PenStrep) and gentamicin, as media supplements (https://www.atcc.org/Products/Culture_Reagents/Antibiotics.aspx). Many large-scale genomic projects - such as the Encyclopedia of DNA elements (ENCODE) consortium, an international collaboration aiming to annotate functional elements in the human genome and gene expression patterns across tissues (ENCODE consortium, 2012) - use cell lines in order to understand the diversity of gene expression and regulatory profiles across human cell types and require the routine use of antibiotics in their protocols. The implicit assumption made within the community is that using antibiotics in cell culture has a negligible impact on gene expression.

Previous studies have demonstrated that changes in gene expression and regulation *in vitro* can be induced by antibiotics (Smith et al., 2014). One study has even shown that antibiotics such as rifampin can induce genome-wide, drug-dependent changes in gene regulation and expression patterns in human hepatocytes. However, the molecular consequences of growing human cells with antibiotics at standard cell culture concentrations have yet to be thoroughly investigated.

Penicillin is a group of antibiotics metabolized in the liver that eliminate bacteria by inhibiting the peptidoglycan synthesis necessary to maintain the bacterial cell wall (Mucsi et al., 2013). Streptomycin is an antibiotic that acts as a protein synthesis inhibitor by

binding to the small 16S rRNA of the 30S subunit of the bacterial ribosome, interfering with codon reading and ultimately the death of microbial cells through mechanisms that are still not well understood (Sharma et al., 2007). Although streptomycin is used in combination with penicillin in standard antibiotic cocktails to prevent bacterial infection in cell culture, their mechanism of action in cells other than microbial cells are not well understood.

2.2 Materials and methods

In order to investigate the effects of antibiotics commonly used in cell culture on gene expression and regulation, I performed RNA-seq and H3K27ac ChIP-seq on HepG2 cells, immortalized human liver cells, treated with or without PenStrep (Fig. 1). HepG2 cells (ATCC) were thawed at passage 37 and cultured in two separate batches with D-MEM, high glucose media (catalog # 11965-092, Life Technologies) that contained fetal bovine serum (catalog # CCFAP003, UCSF Cell Culture Facility) at 10%, L-glutamine (catalog # CCFGB002, UCSF Cell Culture Facility) at 1% in 6-well plates (catalog # 3516, Corning). One batch was cultured with this DMEM media supplemented with penicillin-streptomycin (catalog # CCFGK003, UCSF Cell Culture Facility) at the standard ATCC guideline of 1%. The other batch that served as the control group, was cultured in parallel in DMEM media that was not supplemented with penicillin-streptomycin and incubated at 37 degrees Celsius, 5% CO_2 for two passages (or 21 days post-thaw).

PenStrep-cultured and control HepG2 cells were washed with PBS, and lysed directly with Buffer RLT from the RNAeasy mini kit (Qiagen) with the on-column DNase digestion step for collecting total RNA. Libraries were made with Illumina NGS library preparation for polyA tail selection on 5 ug of total RNA per sample by the UCSF Genomic Core Laboratories (http://humangenetics.ucsf.edu/genomics-services/). Three

technical replicates were carried out for each condition. 50 base pair (bp) single end sequencing was carried out on an Illumina Next-Generation Sequencing platform. The resulting reads were demultiplexed and aligned to the human genome (hg19) using STAR (Dobin et al., 2013), which also calculated read counts for each gene in the Ensembl annotation. Analysis for differential expression across the nine replicates was performed using DESeq2 (Love et al., 2014). DESeq2 was chosen due to its stringency in normalization and distribution methods. Gene ontology terms per RNA-seq cluster of differentially expressed (DE) genes were assessed using DAVID (McLean et al., 2010).

Around twelve million cells were cultured separately from the RNA-seq experiments but in identical conditions as previously mentioned (37°C, 5% CO_2 for two passages) on 6 well plates. For each immunoprecipitation, one plate was fixed with 1% formaldehyde for 15 minutes and quenched with 0.125 M glycine. The remainder of the ChIP-seq protocol was carried out using the Diagenode LowCell# ChIP kit (Diagenode; catalog number: C01010070) following the manufacturer's protocol. Chromatin was isolated by adding lysis buffer and sheared to an average length of 300 bp with a Covaris sonicator. Genomic DNA regions of interest were isolated using 4 ug of antibody against H3K27ac (Abcam; catalog number ab4729). Genomic DNA complexes were washed, eluted from the beads with SDS buffer, and subjected to RNase and proteinase K treatment. Crosslinks were reversed by incubation at 65°C, and ChIP DNA was subsequently isolated and purified. Three technical replicates were carried out for each condition. Input genomic DNA was prepared by treating aliquots of chromatin with RNase, proteinase K and heat for de-crosslinking, followed by ethanol precipitation. Pellets were resuspended and the resulting DNA was quantified on a Bioanalyzer. ChIP and input DNAs were prepared for amplification using a ThruPLEX DNA-seq kit (Rubicon Genomics) following the manufacturer's protocol. Library barcode adaptors were added to each sample during amplification and the library was size-selected (~150-200 bp)

using a Diagenode iPure kit v2. The resulting amplified DNA was purified, quantified, and tested by a Bioanalyzer reading to assess the quality of the amplification reactions. Amplified DNA libraries were sequenced on and Illumina HiSeq 4000. All reads were mapped to the human genome using Bowtie (Langmead et al., 2012). H3K27ac peaks were called against input using MACS (Chen et al., 2009) and peaks consistent across replicates identified using the ENCODE Irreproducibility Discovery Rate (IDR) pipeline (Landt et al., 2012). For differential peak intensity analysis, peaks across conditions were merged and reads coverage obtained using HTSeq (Anders et al., 2014). H3K27ac peaks differentially enriched in the PenStrep condition were then identified through DESeq2 (Love et al., 2014) using a custom normalization matrix to correct for differing signal to noise ratios between replicates (Shao et al., 2012). As with the RNA-seq analysis, PenStrep dependent regions were identified through clustering analysis on all differentially enriched H3K27ac regions with adjusted p-values (FDR) <0.1 between the PenStrep-treated and non-treated conditions using the R package hclust and displayed in a heatmap.

2.3 Drug-associated genes are differentially expressed due to PenStrep

To systematically identify genes that are differentially expressed due to cell culture antibiotics, I compared gene expression levels of HepG2 cells cultured with standard 1% PenStrep-supplemented media and HepG2 cells cultured with media not supplemented with PenStrep. Using DESeq2 to perform differential expression analysis (Love et al., 2014), I identified 209 differentially expressed (DE) genes using a q-value cutoff, after adjustment for multiple testing of less than or equal to 0.1 (Fig. 2a). Amongst these DE genes, 156 were significantly upregulated due to PenStrep treatment and 49 were downregulated. These include a set of transcription factors- activating transcription factor 3 (*ATF3),* SRY-box 4 *(SOX4),* forkhead box O4 *(FOXO4),* TGFB induced factor

homeobox 1 (*TGIF1),* homeobox D1 *(HOXD1) ,*forkhead box C1 *(FOXC1),* general transcription factor IIIC subunit 6 *(GTF3C6)* – some of which are known to play a significant role in drug and stress response (Luizon et al., 2016; Liu et al., 2016; Kuo et al., 2016; Bar et al., 2016; Shen et al., 2016; Went et al., 2016; Liu et al., 2016; Denk et al., 2015; Zhang et al., 2011; Tothova et al., 2007; Brenkman et al., 2010; Araujo et al., 2011; Sun et al., 2015; Hneino et al., 2012; Berry et al., 2009). Pathway analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Huang et al., 2009) found that the cluster of 156 DE genes that were upregulated due to PenStrep are enriched for gene ontology terms associated with apoptosis (p-value = 1.91E-05), drug response (p-value = 1.58E-04), unfolded protein response (p-value = 3.84E-04), and nitrosative stress (p-value = 3.98E-04) (Fig. 2a). DE genes that were downregulated in response to PenStrep were enriched for gene ontology categories related to insulin response (p-value = 6.85E-04), cell growth and proliferation (p-value = 0.012), toxic substance (p-value = 0.018) and drug response (p-value = 0.012) (Fig. 2a). Further analyses of PenStrep DE genes using Ingenuity pathway analysis (IPA), found canonical pathway enrichment for PXR/RXR activation (p-value = 9.43E-05), a known drug response pathway associated with antibiotic treatment (Di Masi et al., 2009) (Fig. 2b). Further IPA analyses for upstream regulators enriched for DE genes, identified a significant enrichment for gentamicin (p-value = 2.93E-13), an aminoglycoside (like streptomycin) that is associated with nephrotoxicity as well as ototoxicity in human patients (Edson et al., 1983; Germovsek et al., 2016; Hailey et al., 2016) but also commonly used in cell culture on target genes (Fig. 2b and c). The overlap between target genes of gentamicin and the PenStrep-dependent genes in this analysis demonstrates a similar mechanism of action across antibiotics in human cells. Overall, the diversity of gene pathways activated in PenStrep-treated cells not only suggests that

PenStrep induces a systemic change in gene expression in human cell lines, but that PenStrep may also be inducing broader changes at the gene regulatory level.

2.4 PenStrep induces differential enrichment of active promoters and enhancers marked by H3K27ac

To determine whether culturing cells with PenStrep also leads to chromatin landscape changes that can alter gene regulation, I performed ChIP-seq for H3K27ac on both HepG2 cells cultured with and without PenStrep. Using DESeq2 (Love et al., 2014), we annotated a total of 9,514 peaks that are higher in eiither the PenStrep or the control conditions at a q-value cutoff, after adjustment for multiple testing, of less than or equal to 0.1. Of these peaks, 5,087 were highly enriched in the PenStrep condition and 4,427 peaks were highly enriched in the control treatment (Fig. 1 and Fig. 3a). Using the Genomic Regions Enrichment Annotation Tool (GREAT; (McLean et al., 2010)), I identified genes nearby each cluster of DE regions separately (up or down) for gene ontology enrichment. For the cluster of enriched H3K27ac peaks induced by PenStrep, we observed a significant association with genes involved in tRNA modification (p-value= 2.0E-08), regulation of nuclease activity (p-value= 2.0E-08), cellular response to misfolded protein (p-value= 1.1E-07), and regulation of protein dephosphorylation (p-value= 1.9E-07) (Fig. 3a). As streptomycin is known to act as a protein synthesis inhibitor by binding to the small 16S rRNA of the 30S subunit of the bacterial ribosome, this suggests that the known mechanism of action for streptomycin in bacterial cells may also affect mammalian cells (Sharma et al., 2007; Llobet et al., 2015). For the cluster of H3K27ac peaks that were significantly enriched in the control treatment, GREAT identified an enrichment for genes involved in stem cell differentiation (p-value= 6.8E-22), actin depolymerization (p-value=1.3E-21), negative regulation of transcription factor activity (p-value=2.0E-19), response to reactive oxygen species and positive regulation

of cell cycle (p-value=1.2E-14) (Fig. 3a). Combined, these results demonstrate a global change in the regulatory landscape induced by PenStrep that corroborate some of the gene pathways enriched in our RNA-seq results as well as toxicity pathways that are associated with streptomycin.

2.5 PenStrep-dependent regulatory regions overlap or reside near PenStrep-responsive genes

In order to test for a correlation between the DE H3K27ac peaks and DE genes found via ChIP-seq and RNA-seq respectively, we took all 9,514 DE regions and matched the closest gene expressed (i.e. had a normalized FPKM greater than zero across replicates for all conditions). We found a positive Spearman correlation between regions having a DE H3K27ac signal and expression of DE genes that are enriched in the PenStrep condition (r=0.21) as well as a positive Spearman correlation between DE H3K27ac regions and DE genes that are depleted in the PenStrep condition (r=0.15). Additionally, we performed hyper geometric tests on matching clusters of PenStrep dependent DE genes and DE H3K27ac regions in order to determine significance of enrichment of DE genes within DE H3K27ac regions. We found that the cluster of DE genes and DE H3K27ac regions that are up in PenStrep have significant overlap (p-value=$1.21E^{-07}$). Conversely, the cluster of DE genes and DE H3K27ac regions that are down in PenStrep also have significant overlap (p-value=$2.75E^{-08}$). These findings corroborate the similar gene ontology term enrichment found for both our RNA-seq and ChIP-seq DE genes and peaks. GO terms related to misfolded protein response were found for both clusters of genes and H3K27ac peaks that showed an increase upon PenStrep treatment (Fig. 2a and Fig. 3a). Similarly, the clusters of DE genes and H3K27ac peaks that were depleted upon PenStrep treatment showed GO term enrichment for cell cycle regulation and cell growth.

Many of these DE H3K27ac regions were found to overlap DE genes that are transcription factors, as is underscored by enrichment of regions near genes important for transcription factor activity found through our GREAT analysis (Fig. 3a). One notable example of this is *ATF3*, which is more highly expressed in the PenStrep condition and has a regulatory region that is also enriched in the PenStrep condition that overlaps its third exon (Fig. 3b). *ATF3* is a transcription factor that is known to play a role in pathways underlying cell differentiation and proliferation (Fukasawa et al., 2016; Wang et al., 2015; Kim et al., 2015; Jang et al., 2015), response to unfolded proteins (Clement et al., 2016), inflammation and immune response (Kwon et al., 2015; Whitmore et al., 2007; Labzin et al., 2015), and regulating hepatic gluconeogenesis and insulin resistance (Tsai et al., 2015; Kim et al., 2013; Qi et al., 2009; Koh et al., 2010; Peter et al., 2009). *ATF3* is also known to be a major player in drug response (Luizon et al., 2016; Liu et al., 2016; Kuo et al., 2016; Bar et al., 2016; Shen et al., 2016; Weng et al., 2016; Liu et al., 2016; Denk et al., 2015). Combined, these results show correlation between DE regulatory elements and DE genes and highlight several pathways and genes that could be associated with PenStrep response.

2.6 Conclusion

Despite their known toxicity in humans, there have been only a few reports that have demonstrated how antibiotics impact gene expression. These studies have shown the impact of protein synthesis and enzyme activity in rat cell cultures (Schwarze et al., 1981; Goldstein et al., 1982; Martinez-Liarte et al., 1995) and that these antibiotics produce side effects during human stem cell differentiation into adipocytes (Llobet et al,. 2015). However, the systematic effects of using antibiotics at the standard cell culture concentrations on gene expression and regulation in human cell lines have not been thoroughly investigated (Llobet et al,. 2015). Through our RNA-seq and ChIP-seq data

from PenStrep- and vehicle- treated HepG2 cells, we show that antibiotics can induce a global change in gene expression and chromatin landscape in a human cell line.

We find that PenStrep responsive genes are not only involved in pathways related to drug response, but also to insulin response, fatty acid activation, mitochondrial l-carnitine shuttle pathways, apoptosis, cell growth, and unfolded protein response. We also observe that many of these PenStrep-dependent DE genes are also known targets of gentamicin, some of which are nuclear receptors and transcription factors (e.g. *ATF3*). Gentamicin and other members of the aminoglycoside family, such as streptomycin, are associated with both nephrotoxicity and ototoxicity in humans (Edson et al., 1983; Germovsek et al., 2016). However, the molecular mechanisms underlying death of proximal tubule kidney cells and mechanosensory hair cells due to aminoglycosides are not well understood (Hailey et al., 2016). Given that both types of toxicity have been associated with most aminoglycosides, streptomycin could be an activator of toxicity pathways across cell types. The overlap of PenStrep-activated DE genes with known targets of gentamicin suggests that toxicity pathways known to be induced by antibiotics could exist not only in microbial cells and human patients, but also in human cells *in vitro*. Furthermore, our results may also reveal some patterns in gene expression and regulatory regions underlying toxicity pathways induced by aminoglycosides as a group.

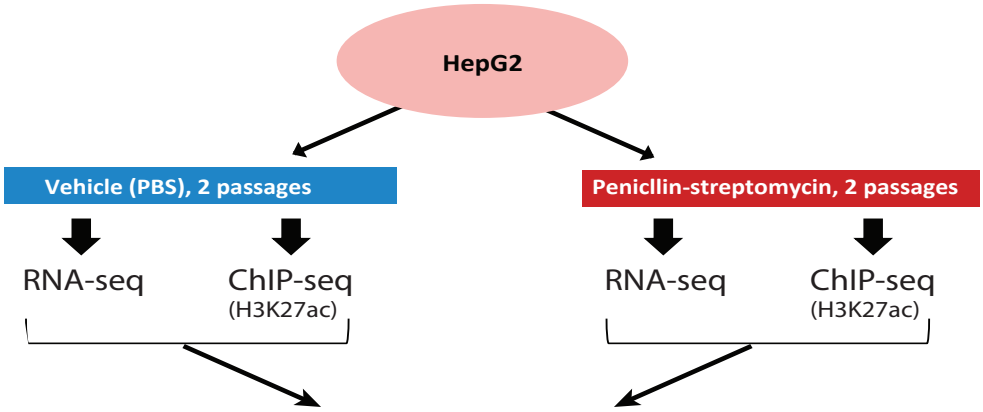
Changes in gene expression were further corroborated by changes in regulatory regions observed in our H3K27ac ChIP-seq data from the same PenStrep treatment of HepG2 cells. Not only was there significant overlap between DE regions found in our H3K27ac data, but when we looked at genes nearby DE regions dependent on PenStrep treatment, we found enrichment for several similar pathways. Regulation of cell cycle and cellular response to misfolded protein reappeared as potential pathways regulated by H3K27ac regions activated by PenStrep. Some pathways based on H3K27ac DE region-gene associations were reminiscent of the known mechanism of action of

streptomycin in bacteria, such as tRNA modification. Other pathways found to be enriched in our DE region-gene associations, such as response to reactive oxygen species and cytoskeleton depolymerization, were previously shown to be disrupted by antibiotics in the differentiation of human stem cells into adipocytes (Llobet et al., 2015) and mouse embryonic stem cells into neurons (Zur Nieden et al., 2004). The recapitulation of pathways disrupted by antibiotics at standard cell culture concentrations in mouse stem cells, human stem cells, human adipocytes and the human liver cell line we used for this study, suggests that the mechanism of action of antibiotics is likely to be detectable and similarly perturbed across other human cell lines commonly used for research.

Several transcription factors were also observed to have enhanced expression due to PenStrep treatment, which could potentially alter the expression of additional genes and pathways. Some of these transcription factors are known to be key regulators of differentiation, drug response, and cell cycle and growth. One of the most notable transcription factors amongst our observed PenStrep –dependent genes, is *ATF3*, a transcriptional repressor that is known to play a role in cell differentiation and proliferation (Fukasawa et al., 2016; Wang et al., 2015; Kim et al., 2015; Jang et al., 2015), the unfolded protein response (Clement et al., 2016), inflammation and immune response (Kwon et al., 2015; Whitmore et al., 2007; Labzin et al., 2015), hepatic gluconeogenesis and insulin resistance (Tsai et al., 2015; Kim et al., 2013; Qi et al., 2009; Koh et al., 2010; Peter et al., 2009), as well as drug response (Luizon et al., 2016; Liu et al., 2016; Kuo et al., 2016; Bar et al., 2016; Shen et al., 2016; Weng et al., 2016; Liu et al., 2016; Denk et al., 2015). *ATF3* is well-studied in a wide diversity of human cell types including hepatocytes, activated T-cells, skeletal muscle cells, macrophages, as well as cancer cell lines and immortalized cell lines. In our study, we find that *ATF3* expression is dependent on PenStrep and that it contains a PenStrep dependent

regulatory region that may compound the regulatory effect of *ATF3* accessibility and expression. Given that *ATF3* expression may be significantly enhanced by PenStrep treatment, studies that examine *ATF3*'s role in pathways related to toxicity, cell proliferation and differentiation should avoid use of antibiotics in cell culture if possible. The impact of culturing liver cells with antibiotics, as shown in our study on HepG2 cells, could also apply to other immortalized human cell lines that are commonly used to assess gene expression patterns at baseline and in drug-induced conditions and even primary cells. It is possible that antibiotics such as penicillin-streptomycin and gentamicin also induce a functional state that is significantly different from the basal state of these cell types. Further evaluation of the biological impact of antibiotic treatment across cell lines is highly warranted. However, we provide some evidence that using antibiotics in cell culture should be avoided- especially in studies focused on drug response as well as cell cycle regulation, differentiation, and growth. Data from studies in which antibiotics are used for cell culture should be examined with caution.

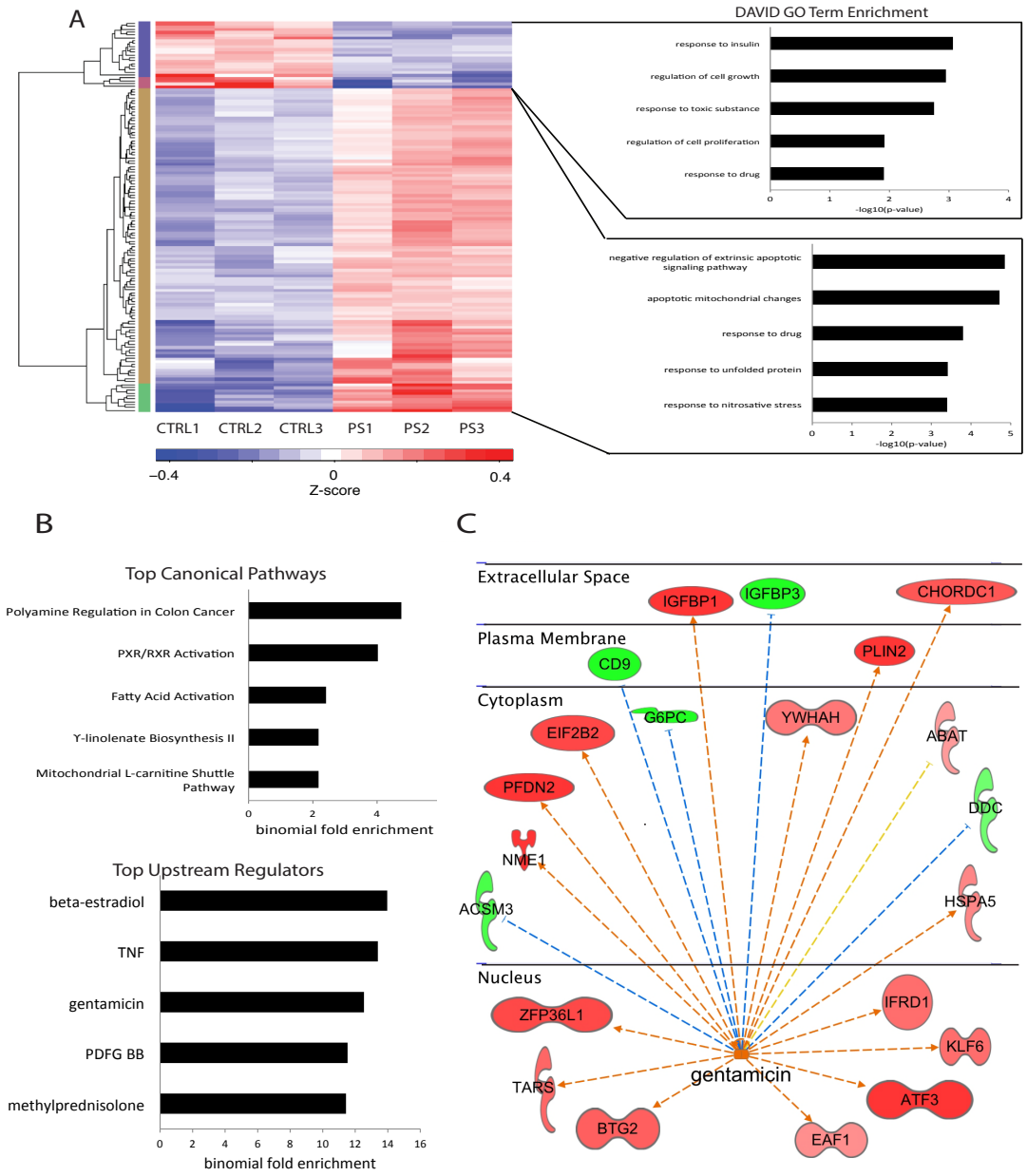
Figure 1. Experimental design of cell culture antibiotics project.



	Pen Strep Up	Pen Strep Down	Total
Differentially expressed genes	157	52	209
Differentially enriched H3K27ac peaks	5087	4427	9514

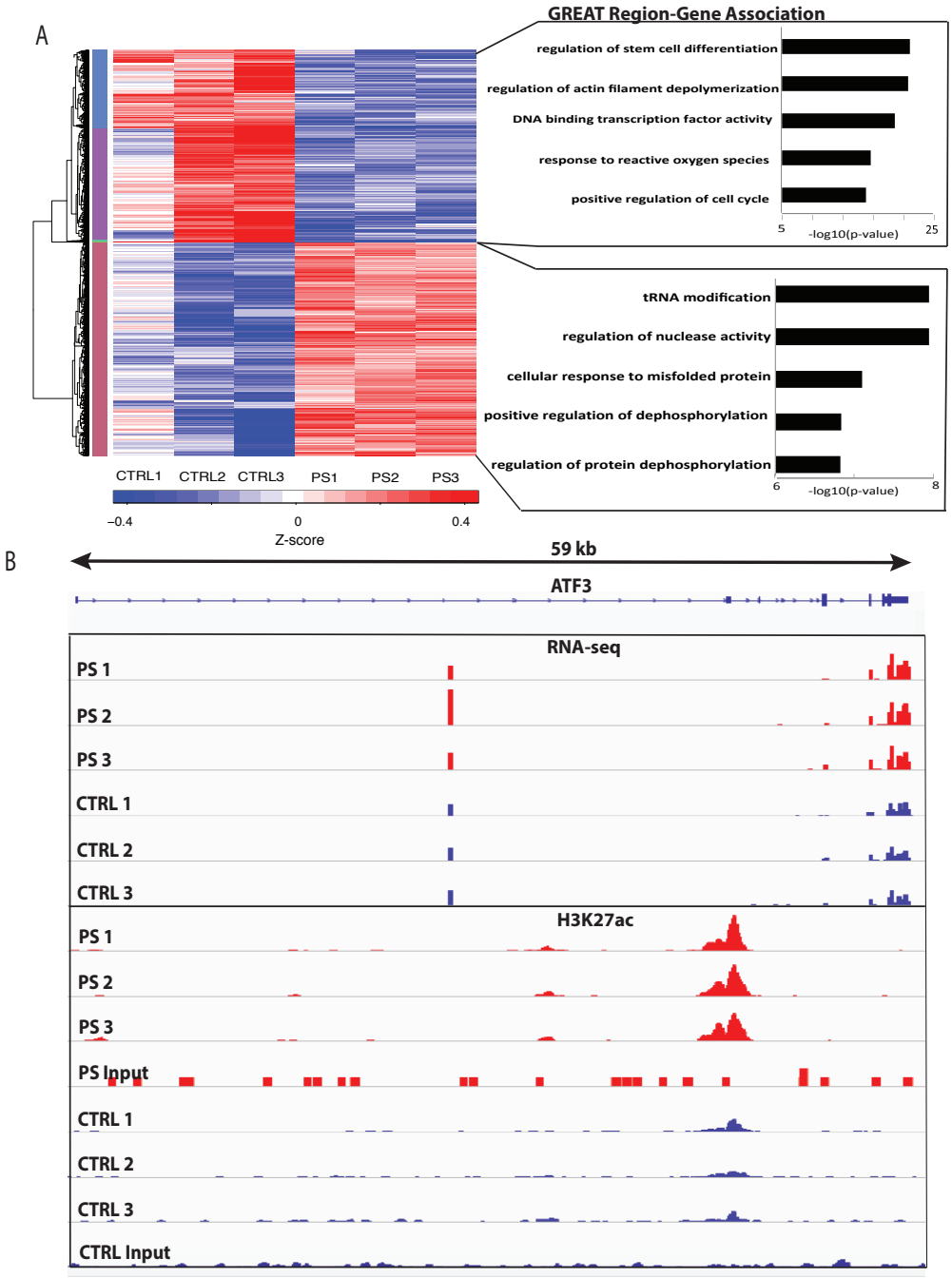
Schematic of the RNA-seq and ChIP-seq assays performed on HepG2 cells treated with and without PenStrep. The number of differentially expressed genes and differentially enriched H3K27ac peaks are listed in the table below the diagram.

Figure 2. RNA-seq analysis on HepG2 cells cultured with and without PenStrep.



(A) Shown in the top left is a heatmap depicting relative expression levels for all 209 differentially expressed genes across all three replicates per treatment. To the right of each cluster, the top gene ontology terms as determined by DAVID v6.8 [24] are shown. **(B)** PenStrep DE genes canonical pathways and upstream regulators with the top five highest binomial fold enrichment values and most significant p-values as determined by IPA. **(C)** IPA network analysis for gentamicin, one of the top most significant upstream regulators for PenStrep DE genes. This network is drawn based on calculated z-scores for published gene expression patterns under gentamicin as determined by IPA.

Figure 3. ChIP-seq analysis of PenStrep responsive peaks in HepG2 cells.



(A) The top right panel shows a heatmap depicting relative expression levels for all 9,514 differentially enriched H3K27ac peaks across all three replicates per treatment. To the right, are shown the top region-gene association pathways according to GREAT [29]. **(B)** Integrative genomic viewer snapshot showing RNA-seq and H3K27ac ChIP-seq results in the *ATF3* locus. This locus shows increased *ATF3* expression in the PenStrep-treated HepG2 cells (red) versus vehicle control treated cells (blue) observed through RNA-seq, as well as ChIP-based enrichment of H3K27ac in PenStrep-treated cells (red) versus vehicle control treated cells (blue).

Chapter 3

Massively parallel dissection of human accelerated regions in primate neural progenitor cells

3.1 The role of human accelerated regions in development and evolution

The genetic changes underlying the many morphological and cognitive differences between humans and other primates are largely unknown. Gene regulatory changes can play an important functional role in these differences, but haven't been well studied. Human accelerated regions (HARs) are highly conserved sequences that acquired many nucleotide substitutions since humans diverged from our common ancestor with chimpanzees. The Pollard group previously identified a set of 721 HARs using PhyloP, a method based on likelihood ratio tests for accelerated sequence divergence on the human lineage (Pollard et al., 2006; Prabhakar et al., 2008; Capra et al., 2013). Even though this approach doesn't constrain HARs to be non-coding, 92% of them are. The genetic signature of these HARs - non-coding, highly conserved but evolved in humans- suggests that these regions are likely to be enhancers or other regulatory elements that play a role in shaping human-specific traits.

Some of the most notable human-specific traits, such as our unique linguistic and cognitive skills, relate back to the neurodevelopmental changes that led to an expanded and more complex forebrain. Of the 92% of HARs that reside in noncoding regions, 30% of these HARs are predicted to be developmental enhancers, the majority of which are predicted to be involved in brain development (Babbitt et al., 2011; Capra et al., 2013; Kamm et al., 2013; Pollard et al., 2006; Prabhakar et al., 2006). Only a small number of HARs have been analyzed and functionally characterized for their regulatory activity via mouse enhancer assays (Capra et al., 2013; Prabhakar et al., 2008). One example is HAR2, also known as HACNS1, where the human sequence was found to gain

enhancer activity in the limb compared to the chimpanzee sequence when assayed in lacZ stainings performed in mouse embryos (Prabhakar et al., 2008).

None of these experiments were high-throughput nor did they analyze HARs at multiple developmental time points. The Ahituv lab, in collaboration with the Shendure lab at the University of Washington, developed and optimized the massively parallel reporter assay (MPRA) as a high-throughput sequencing method that can assay for enhancer function *en masse* (Inoue and Ahituv, 2015). One group has performed MPRA to analyze the effect of ASD-associated biallelic mutations on the enhancer activity of the human sequences of 278 (less than half of all) HARs in an episomal context in primary mouse neurospheres (Doan et al., 2016). This study however, does not examine species-specificity in enhancer activity for all HARs in a cell type relevant for understanding uniquely human patterns in neurodevelopment.

To answer the evolutionary questions about which nucleotide and cellular changes drive differences in human versus chimpanzee neurodevelopment, I assayed both human and chimpanzee sequences of all HARs in human and chimpanzee neural cell types at multiple time developmental points. I used lentivirus-based MPRA (lentiMPRA), which provides the ability to test regulatory sequences in a genomically integrated manner and were shown to be more reproducible and strongly predictable using various enhancer annotations (Inoue et al., 2017). I performed lentiMPRA on neural progenitor cells (NPCs) and glial progenitor cells (GPCs) that I differentiated from human and chimpanzee iPSCs, which allowed us to assay the enhancer activity of all HARs in early neural cell types fated for the forebrain and to understand the contribution of cell species and cell stage in driving HAR enhancer activity (Fig. 4).

3.2 Materials and methods

MPRA library design

All human and chimpanzee sequences for 714 HARs were used to design 171bp-long MPRA probes. I also picked ENCODE positive and negative controls that are commonly used in various cell lines for luciferase assays (provided by the Myers lab), as well as H3K27ac and H3K27me3 ChIP-seq peaks in human iPSC-derived NPCs and GPCs (this data was provided by Dr. Fumitaka Inoue from the Ahituv lab), to design probes against as putative negative and positive controls in the MPRA library. Flanking regions were added to HARs shorter than 171bp, and HARs longer than 171bp were tiled across with variable overlap depending on the length of the HAR. All regions for the human and chimpanzee sequences were designed to be completely orthologous. All HAR and control sequences were scanned for homopolymers and restriction sites (for SbfI and EcoRI) and modified to avoid problems in synthesis and cloning. The final array design included 2,440 different target sequences, each with 100 uniquely assigned 15bp-long barcodes for a total of 244,000 oligos.

MPRA library synthesis and cloning

All MPRA sequences were array-synthesized as 230 bp oligos (Agilent Technologies) containing an enhancer, spacer, and barcode. The amplification and cloning of the enhancers and barcodes into the pLSmp-SV40-mP lentiviral vector was performed by Beth Martin in the Shendure lab by the same method described in a recent study performed by the Ahituv and Shendure labs (Inoue et al, 2017).

Cell lines

The UCSF Committee on Human Research and the UCSF GESCR (Gamete, Embryo, and Stem Cell Research) Committee approved all human iPSC experiments. We performed MPRA in NPC and GPC cells derived from four separate iPSC lines, originating from two human males (WTc and HS1 provided by the Yamanaka lab) and

two chimpanzee males (Pt2A and Pt5C provided by Dr. Alex Pollen). All lines were reprogrammed from fibroblasts using episomal plasmids according to a recently published protocol (Okita et al, 2011). For additional lines, we electroporated three micrograms of episomal expression plasmid mixture encoding OCT3/4, SOX2, KLF4, L-MYC, LIN28, and shRNA for TP53 into 300,000 fibroblasts from each individual with a Neon Electroporation Device (Invitrogen), using a 100 µL kit, with setting of 1,650V, 10ms, and three pulses (Bershteyn et al., 2016). After 5 – 8 days, cells were detached and seeded onto irradiated SNL (derived from mouse fibroblasts) feeder cells. The culture medium was replaced the next day with primate embryonic stem cell (ESC) medium (Reprocell) containing 5 – 20 ng/mL of βFGF. Colonies were picked after 20 – 30 days, and selected for further cultivation. After three to five passages, colonies were transferred to Matrigel-coated dishes and maintained in mTeSR1 medium (Stem Cell Technologies, 05850) supplemented with Penicillin/Streptomycin/Gentomycin. Further passaging was performed using calcium and magnesium free PBS to gently disrupt colonies. Each line showed a normal karyotype (Fig. 5B).

Neural differentiation of human and chimpanzee iPSCs

Human and chimpanzee iPSCs were cultured in Matrigel-coated plates with mTesr media in an undifferentiated state. Cells were propagated at a 1:3 ratio by treatment with 200 U/mL collagenase IV and mechanical dissection. To trigger neural induction, iPSCs were split with EDTA at 1:5 ratios in culture dishes coated with matrigel and culture in N2B27 medium (comprised of DMEM/F12 medium (Invitrogen) supplemented with 1% MEM-nonessential amino acids (Invitrogen), 1 mM L-glutamine, 1% penicillin-streptomycin, 50 ng/mL bFGF (FGF-2) (Millipore), 1x N2 supplement, and 1 x B27 supplement (Invitrogen)) supplemented with 100 ng/ml mouse recombinant Noggin (R&D systems). Cells at passages 1-3 were split by collagenase into small clumps,

similar to hESC culture, and continuously cultured in N2B27 medium with Noggin. After passage 3, cells were plated at the density of 5E4 cells/cm2 after disassociation by TrypLE express (Invitrogen) into single-cell suspension, and cultured in N2B27 medium supplemented with 20 ng/mL bFGF and EGF. Cells were maintained under this culture condition for a minimum of three months with a stable proliferative capacity. More specifically, NPCs were collected at P12-18 and GPCs at P20-28 (Fig. 5A).

Validation of NPC and GPC markers through immunostaining

Human and chimpanzee NPCs and GPCs were validated for NPC and GPC via immunostaining against Nestin (NPC), Pax6 (NPC), and GFAP (GPC) (Fig. 5B). Human and chimpanzee NPCs and GPCs were cultured in chambered Millipore EZ slides, rinsed with PBS, and fixed with 4% paraformaldehyde in PBS for 15 minutes at room temperature. Cells were then washed three times with ice cold PBS then permeabilized through incubation for 10 min with PBS containing 0.1% Triton X-100. Cells were then washed in PBS three times then incubated with 10% donkey serum for 30 minutes to block unspecific binding of antibodies. Cells were then incubated with diluted primary antibodies against Nestin (monoclonal, mouse, Abcam, AB6142), Pax6 (polyclonal, rabbit, Abcam, AB5790), and GFAP (polyclonal, rabbit, Chemicon, AB5804) in 10% donkey serum for 1 hour at room temperature. The cells were then washed three times in PBS, 5 minutes each wash, then incubated with secondary antibody (Alexa 488 donkey anti rabbit, Life technologies; Alexa 546 donkey anti mouse, Life technologies) in donkey serum for 1 hour at room temperature in the dark. Cells were then washed three times with PBS in the dark, then covered with a cover slip in Cytoseal mounting media (Thermo Scientific).

Single cell gene expression analysis of human and chimp NPCs and GPCs

The capture of single cells was done using the C1TM Single-Cell Auto Prep Integrated Fluidic Circuit (IFC), which uses a microfluidic chip to capture the cells, perform lysis, reverse transcription and cDNA amplification in nanoliter reaction volumes. The details of the protocol used are described in protocol PN100-7168 (http://www.fluidigm.com/). Sequencing libraries were prepared after the cDNA was harvested from the C1 microfluidic chip using the Nextera XT Sample Preparation Kit (Illumina), following its protocol with minor modifications. The single cell libraries from each C1 capture were then pooled, cleaned twice with 0.9X Agencourt AMPure XP SPRI beads (Beckman Coulter), eluted in DNA suspension buffer (Teknova) or EB buffer (Qiagen) buffer and quantified using High Sensitivity DNA Chip (Agilent).

Paired-end reads were generated per library. Using cutadapt under the Trim Galore! wrapper with the default settings, reads were trimmed for quality, and Nextera transposase sequences were removed. Reads shorter than 20bp were discarded. Read level quality control was then assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Reads were aligned to the NCBI human reference sequence HG38, by HiSat2 using the --prefilter-multihits option and a guided alignment via the NCBI RefSeq transcriptome reference.

Expression for RefSeq genes was quantified by the featureCounts routine, in the subRead library (Liao et al., 2014), using only uniquely mapping reads and discarding chimeric fragments and unpaired reads. Gene expression values were normalized based on library size as counts per million reads (CPM). We used visual image calls to remove any libraries that originated from C1 chambers with multiple cells. To further identify outlier cells, we removed libraries with fewer than 1,000 genes detected, or with greater than 20% of reads aligning to mitochondrial or ribosomal genes.

To determine the composition of cell types in human and chimpanzee samples used for MPRA, we set a threshold of detection for each gene at 2 CPM. We then

calculated the percentage of cells expressing regional identity genes (e.g., FOXG1 for telencephalon, DLX6-AS1 for GABAergic neurons, MKI67 for dividing cells, SLC1A3 for radial glia). In both human and chimpanzee cell lines at the NPC and GPC stage, 50-90% of cells expressed telencephalon (FOXG1) and radial glia/astrocyte markers.

Lentivirus library infection, RNA & DNA isolations, and sequencing

Lentivirus packaging of the HAR MPRA library was performed by the UCSF Viracore. For all human and chimpanzee cell lines and cell stages, about twelve million cells were plated in 15cm dishes and cultured for 4 days. Infected cells were washed daily with PBS and washed three times before cell lysis for DNA and RNA isolations. Genomic DNA and total RNA were extracted using the AllPrep DNA/RNA mini kit (Qiagen). Messenger RNA was purified from the total RNA using Oligotex mRNA mini kit (Qiagen) and treated with Turbo DNAseq to remove contaminating DNA. The RT-PCR, amplification and sequencing of RNA and DNA were performed (by Beth Martin and Martin Kircher from the Shendure group) as previously described in a recent MPRA study (Inoue et al, 2017). In brief, mRNA was reverse transcribed with SuperScript II (Invitrogen) using a primer downstream from the barcode. The resulting cDNA was amplified in two round PCR setup, as well as the genomic DNA from the same samples. PCR products were cleaned with AMPure XP beads (Beckman Coulter) to remove primers and concentrate samples. All reactions were pooled and run on agarose gels for size selection and submitted for sequencing. RNA and DNA for all three replicates for all samples were sequenced on an Illumina NextSeq instrument (2 x 26 + 10bp index).

Normalization of RNA/DNA ratios and quantification of enhancer activity

Quality control of sequencing data was performed mostly by Martin Kircher. Downstream normalization of RNA/DNA ratios and quantification of enhancer activity were performed

by Sean Thomas and Katherine Pollard. To normalize RNA and DNA for different sequencing depths in each sample, reads were divided by the sum of observed counts and reported as counts per million. Only barcodes observed in RNA and DNA of the same sample were considered. RNA/DNA ratios per HAR were calculated by taking the sum of RNA counts for all barcodes assigned to all oligo(s) tiling across each HAR, divided by the sum of all DNA counts for all barcodes across all oligo(s) per HAR. HAR sequences were defined as active enhancers if they had a human and/or chimpanzee allele with a normalized RNA/DNA ratio above the 75th percentile threshold of the negative control group in all three technical replicates for at least two NPC samples or at least two GPC samples.

Modeling sequence origin, cell species, and cell stage effects on enhancer activity

To test for HARs with human allele versus chimp allele activity, as measured by log2(human [RNA/DNA]/chimpanzee [RNA/DNA]) ratios, significantly different from zero, we used the R limma package to fit a linear model for the log-ratios across conditions. For every HAR with significantly differentially active human and chimpanzee alleles (adjusted p-value< 0.01), the average RNA/DNA ratios per human and chimpanzee allele were calculated across conditions. We also modeled expression in each cell species of origin and cell stage separately and used the R limma package to test for differences in HAR activity in human versus chimpanzee cells or in NPC versus GPC stage.

3.3 Most HARs are active enhancers in primate neural and glial progenitors

Many human-specific traits revolve around the expansion and increased complexity of the human brain (Miller et al, 2012; Somel et al. 2013). Given that almost 30% of HARs have epigenetic signatures that suggest they function as developmental enhancers in

the brain (Capra et al., 2013), we hypothesized that many HARs act as brain enhancers in at least chimpanzee or human neural progenitors fated for the forebrain. Our lentiMPRA results in human and chimpanzee iPSC-derived NPCs and GPCs confirms this hypothesis. We found that 306 out of 714 HARs demonstrate significant enhancer activity in at least two neural progenitor cell lines or at least two glial progenitor cell lines ($p<0.01$) (Fig. 6). The 306 HARs that are active brain enhancers seem to divide fairly evenly into two general clusters: a cluster of 147 HARs with a human allele that is a more active enhancer than the chimpanzee allele, and a cluster of 159 HARs with a chimpanzee allele that is more active than the human allele. Using the Genomic Regions Enrichment Annotation Tool (GREAT; Mclean et al, 2010), I identified genes nearby these 306 HAR brain enhancers for gene ontology enrichment compared to a background of the whole human genome. I found a significant association with genes involved in “negative regulation of gene expression” (p-value= 2.5E-09), “positive regulation of transcription from RNA polymerase II promoter” (p-value=2.6E-09), “brain development” (p-value=1.1E-08), “forebrain development” (p-value=6.4E-07), “telencephalon development” (p-value=2.8E-06), and “hindbrain development” (p-value=1.1E=-05). For the cluster of 159 HARs with a chimp allele that is a stronger enhancer than the human allele, there was a significant association with genes involved in hippocampus morphology (p-value=5.4E-06) and temporal lobe morphology (p-value=6.7E-06). The cluster of 147 HARs with a human allele that is a more active enhancer than the chimp allele had a significant association with genes involved in cerebellar granule cell morphology (p-value=2.8E-06) and brain ventricle/choroid plexus morphology (p-value=4.7E-06).

Four active brain enhancer HARs - 2xHAR.518, 2xHAR.548, HAR152, and 2xHAR.133- were chosen for *in vivo* validation of enhancer activity in mouse embryos based on their magnitude in fold change of human over chimp enhancer activity (Fig. 6),

as well as their predicted gene target. 2xHAR.518 is inside an intron of neurexin 3 (*NRXN3)*, a gene known to play a role in neurodevelopment that has been associated with autism and schizophrenia (Hu et al, 2013; Gauthier et al, 2011; Rujescu et al, 2009). 2xHAR.548 is within the same topologically associating domain (TAD) as *FOXP1*, a gene that has been shown to play a role in radial migration and morphogenesis of cortical neurons, and is associated with autism and speech disorders (Teramitsu et al., 2004; Li et al., 2015; Lozano et al., 2015). HAR152 is in the 3'UTR of *NEUROG2*, a proneural gene that is expressed during the differentiation of the spinal cord and cranial nerves (Aaker et al., 2009; Espinosa-Medina et al., 2014). 2xHAR.133 is in the same TAD as *MEIS2*, a gene important for neural cell proliferation and differentiation, as well as cranial neural crest development (Agoston et al., 2011; Agoston et al., 2014; Machon et al., 2015). Neural expression patterns were validated via lacZ staining in mouse embryos for the human (hg19) and chimpanzee (pt4) sequences for HAR152 (E10.5) and 2xHAR.548 (E13.5) (Fig. 7). Expression patterns for both the human and chimpanzee sequences for HAR152 were consistent in the brain stem. Expression patterns for 2xHAR.548 showed some suggested species-specific enhancer domains. The human sequence for 2xHAR.548 showed more consistent expression patterns in the mid- and hindbrain, whereas the expression pattern for the chimpanzee sequence was more consistent in the eye and neural tube.

3.4 Many HARs have human and chimp alleles with differential enhancer activity

Of the 306 HARs that demonstrated enhancer activity in human and chimp neural and glial progenitors, 204 HARs had human and chimp alleles that were differentially active enhancers across cell species and cell stage at an adjusted p-value<0.01 (Fig. 8). HARs with the greatest differences in enhancer activity between the human allele versus chimp allele are listed in Table 1. Only six HARs showed significant differences in activity

between NPCs and GPCs (adjusted p-value<0.01): 2xHAR.319, HAR5, 2xHAR.28, 2xHAR.238, 2xHAR.1, and 2xHAR.49. Even fewer HARs showed a significant difference in activity between human and chimpanzee cell lines (adjusted p-value<0.01): 2xHAR.518, HAR51, and 2xHAR.264. These results suggest that the sequence origin, or human-chimp fixed differences at the nucleotide level, influence the level of HAR enhancer activity much more than the differences in cellular context in which these HAR sequences are tested.

3.5 *cis* regulatory features are stronger drivers of HAR enhancer activity than the *trans* environment of cell species and cell stage

In order to understand the degree to which HAR enhancer activity depends on *cis* features (i.e. nucleotide-level fixed differences between humans and chimps) or *trans* features (i.e. cell species of origin and cell stage), I performed two types of analyses to look at the contribution of cell species and cell stage to HAR enhancer activity. The first approach was simply an identification of cell conditions in which HAR enhancers were active and the intersections of these conditions using an UpSetR package (Conway et al, 2017). When comparing HAR enhancer activity across four sets of *trans* environments- human NPCs, human GPCs, chimp NPCs, chimp GPC – the largest set of *trans* conditions in which HAR enhancers were active was the intersection of all four conditions (Fig. 9A). 130 HARs demonstrated enhancer activity in all four cell conditions, and the next two largest sets of HARs were active enhancers in at least three cell conditions: 72 HARs were active in chimp NPCs, human GPCs, and human NPCs; 30 HARs were active in chimp GPCs, chimp NPCs, and human NPCs. This result suggests the relatively low contribution of *trans* effects on HAR enhancer activity since the largest sets of active HAR enhancers span cell species and cell stage. This is consistent with previous reports that have shown that humans and chimpanzees are nearly identical at the amino acid level (King et al., 1975), which implies that many of the human-specific

changes in protein-DNA interactions required for gene regulation are driven more by the changes in non-coding DNA than in differences in protein structure of transcription factors or DNA-binding proteins.

I further investigated the contribution of *cis* versus *trans* features in driving HAR enhancer activity through a principal components analysis performed on HAR enhancer activity across all replicates for all cell species and cell stages. Patterns in HAR enhancer activity separate out on the first five principle components (Fig. 9B). The impact of *cis* features (fixed human-chimp nucleotide differences) is highlighted by the patterns in HAR enhancer variation captured by the second principal component, which separates consistently between human and chimpanzee alleles tested across samples. The third and fourth principal components separate based on cell stage and cell species, but these patterns in variation are much less clear. These results further suggests the role of *trans* factors (human versus chimp cells, NPCs versus GPCs) in driving enhancer activity of HARs is relatively small comparing to the influence of *cis* factors, the fixed nucleotide differences between humans and chimpanzees within these regions.

3.6 HARs contain fixed differences that disrupt TFBS motifs

To further dissect the regulatory architecture of the HAR brain enhancers identified in our study, we looked at TFBS motifs that were disrupted by human-chimp fixed differences at the nucleotide level. Using JASPAR core vertebrate motifs and a p-value threshold of 1^{-05} in FIMO, we were able to identify a set of TF binding sites that were only present in the human allele or only in the chimp allele for the 204 differentially active HAR enhancer sequences (Table 2). More chimp-specific TF binding sites were discovered at this p-value threshold- 137 TF motifs were chimp-specific and 98 TF motifs were found to be human-specific – which is consistent with the expectation that recent human changes in these HAR enhancers that have been conserved between

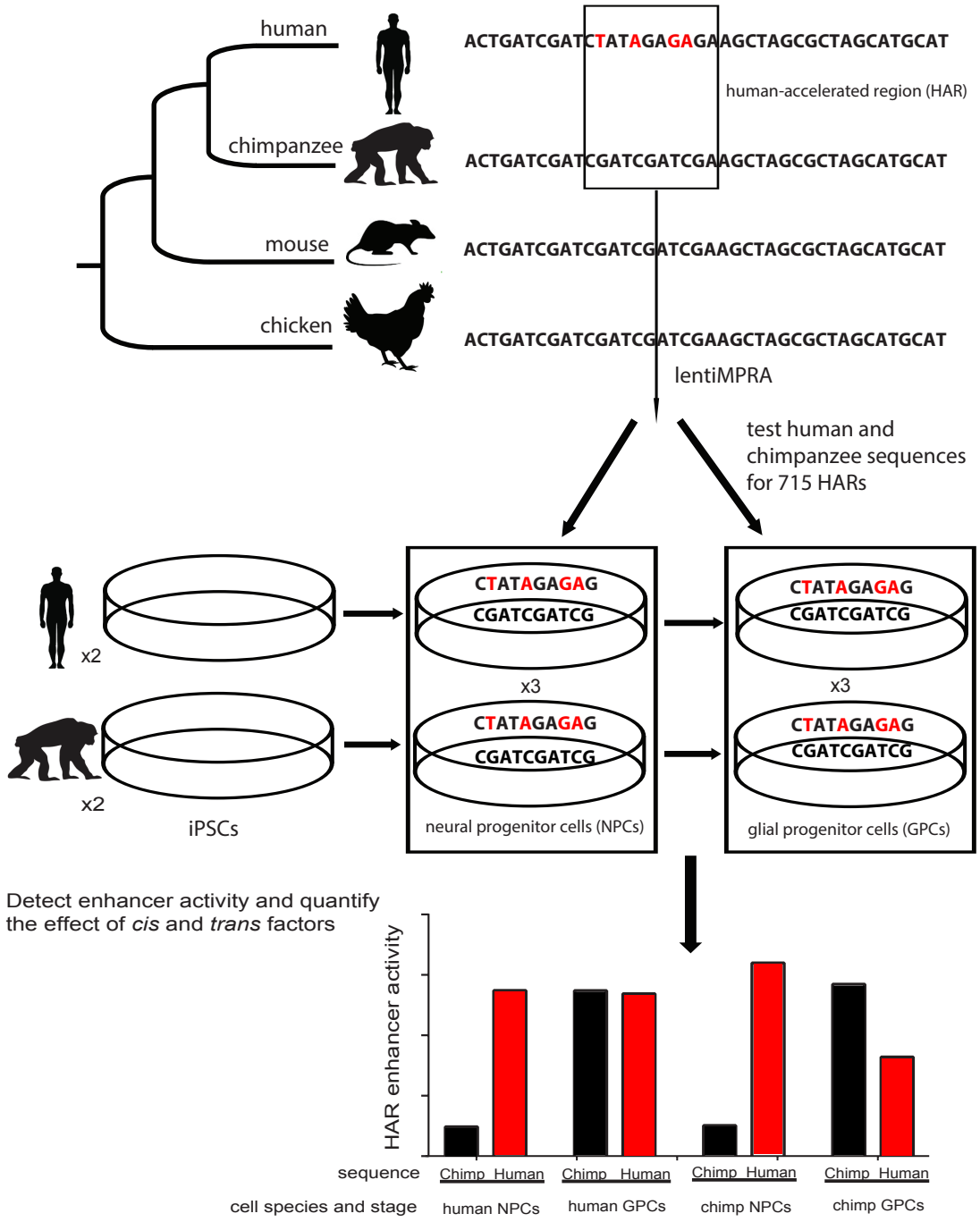
chimp and other vertebrates are likely to disrupt their function as regulatory elements. Some notable transcription factors that have more motifs in human are EGR1/2/3/4, which are zinc finger proteins that are known to play a role in neuronal plasticity (Liu et al., 2000; Knapska et al., 2004; Lu et al., 2011). One other notable TF that has more motifs in human is *FOXP1*, a transcription factor previously mentioned that is in the same TAD as 2xHAR.548 and is associated with autism, speech disorders as well as cortical neuron migration and neurogenesis. TFs with more motifs in chimpanzee included the POU family of neural transcription factors, as well as *ZNF263*. Misregulation of and mutations in *ZNF263* have been linked with autism and hypothalamic hamartoma (Ning et al., 2015; Saitsu et al., 2016). Most of the TFs with species-specific motifs found here are expressed in the brain and known to play a role in neurodevelopment. By overlapping TFBS motifs with human-chimp fixed differences in brain enhancer HARs that are differentially active between human and chimp alleles, we have identified potential TFs that could elucidate the molecular changes that allow for the expansion and increased complexity of the human brain.

3.7 GWAS SNPs linked to neuropsychiatric disorders reside near HAR enhancers

To further evaluate the function of the novel HAR enhancers we identified through our MPRA in neurodevelopment, we investigated the overlap between HAR enhancer loci and GWAS SNPs associated with neuropsychiatric disorders. Many HARs were within a 2MB window of GWAS SNPs associated with schizophrenia (SCZ), autism spectrum disorder (ASD), bipolar disorder (BIP), attention deficit hyperactivity disorder (ADHD), and major depressive disorder (MDD) (Table 3). Only a couple differentially active brain enhancer HARs contain or are in LD with GWAS SNPs that pass a significance threshold of 5^{E-06}. 2xHAR.170 and 2xHAR.502. 2xHAR.170 is in LD with rs2434531 which is associated with schizophrenia (Ripke et al., 2014), a disorder that is speculated

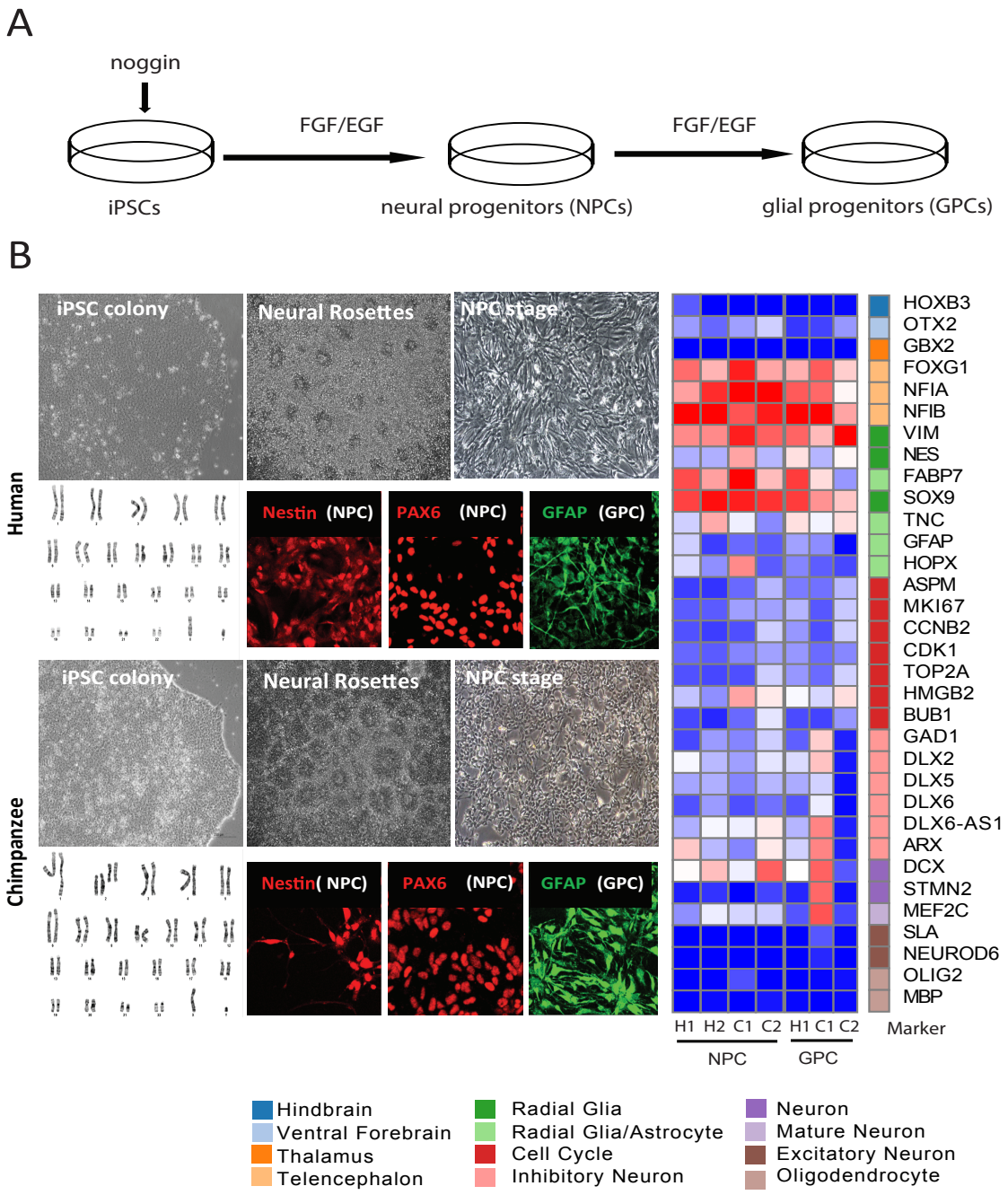
to be human-specific (https://carta.anthropogeny.org/moca/domains/mental-disease). 2xHAR.502 is in LD with rs10249234 which is associated with educational attainment (Okbay et al., 2016) and age of first birth (Barban et al., 2016), both traits that are potentially linked to intelligence. The links discovered between HAR brain enhancers and GWAS SNPs sheds light on potential mechanisms through which these variants impact their associated neurodevelopmental traits.

Figure 4. Graphical abstract of HAR MPRA project.



Human and chimpanzee sequences for all 714 HARs were assayed for enhancer activity through lentiMPRA performed on human and chimpanzee iPSC-derived neural progenitor cells (NPCs) and glial progenitor cells (GPCs). Two biological replicates were tested per species- two human and two chimpanzee individuals- and all samples were performed in triplicate at the NPC and GPC stage.

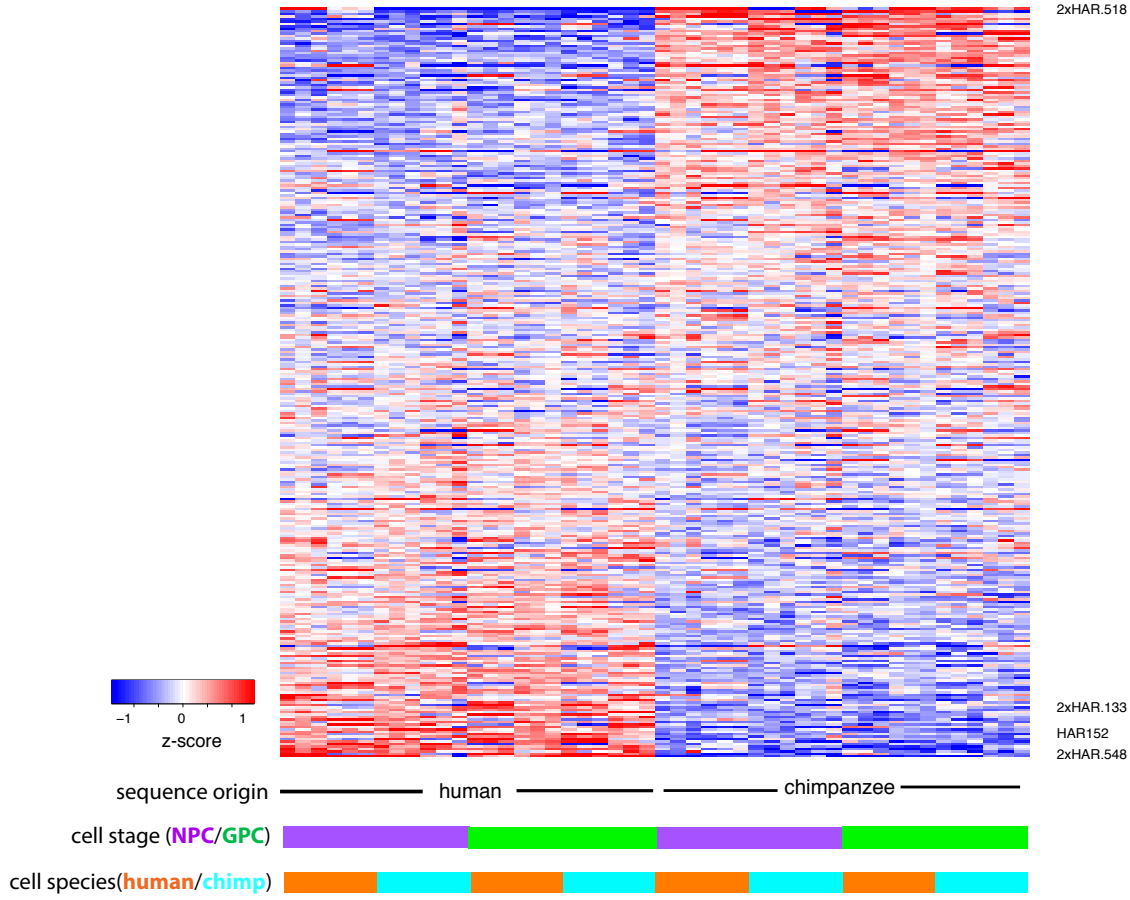
Figure 5. Differentiation and characterization of human and chimpanzee cell lines.



(A) Neural and glial progenitor cells were differentiated from two human and two chimpanzee iPSC lines. iPSCs were cultured with noggin for neural induction, then cultured with FGF and EGF for further differentiation of cells into neural progenitor cells (NPCs) and glial progenitor cells (GPCs). **(B)** Reprogrammed iPS colonies from human and chimpanzee fibroblasts exhibit normal morphology and maintain species-specific karyotypes. Cells were differentiated into neural rosettes and demonstrated typical morphology, i.e. bipolar with small soma, as adherent cultures during the NPC and GPC stages. Characterization by immunoassays showed normal expression of neural stem cell proteins (Nestin and PAX6) and glial progenitor proteins (GFAP).

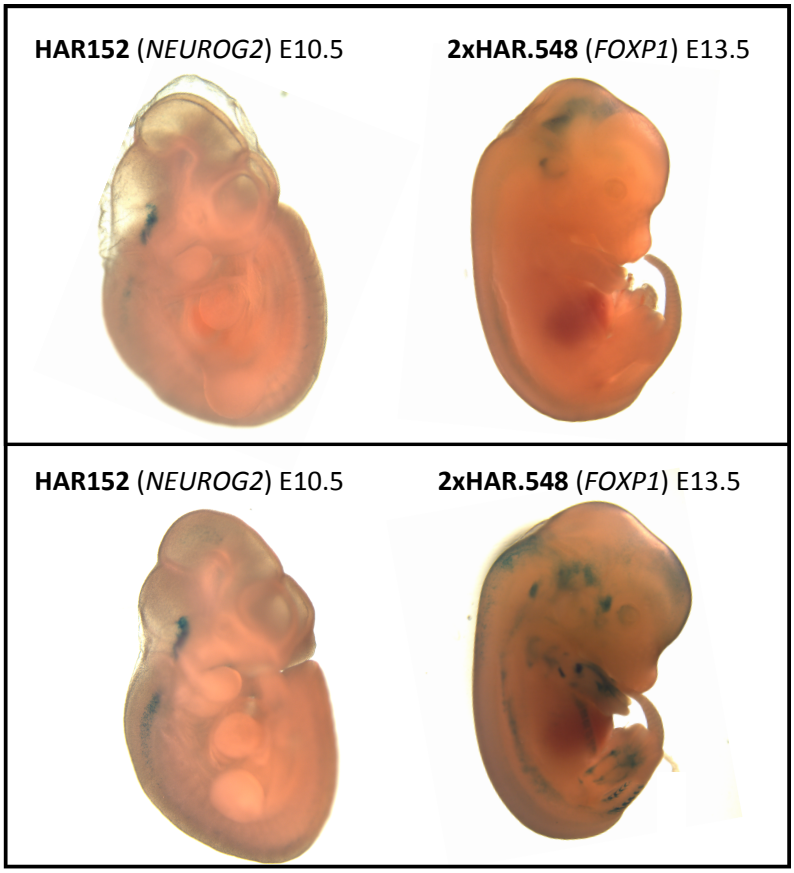
Immunohistochemistry also revealed pluripotency marker expression in chimpanzee iPS colonies (upper right). Single cell gene expression analysis from human and chimpanzee NPCs and GPCs show comparable marker expression for radial glia and telencephalon. In both human and chimpanzee cell lines at the NPC and GPC stage, 50-90% of cells expressed FOXG1, a marker of radial glia fated for the forebrain.

Figure 6. Many HARs are active enhancers in human and chimp NPCs and GPCs.



A heatmap depicting z-scores as relative levels of enhancer activity for all 306 human and chimpanzee HAR sequences with enhancer activity greater than the 75th percentile threshold of the negative control group in human and chimpanzee NPCs and GPCs. Each row represents a HAR that is active in all replicates for at least two NPC samples or at least two GPC samples. Column annotations for sequence origin, cell species and cell stage are show below. Enhancer activity for all human sequences is depicted on the left with the corresponding chimpanzee sequence on the right. Cell stage (purple for NPCs, green for GPCs) and cell species (orange for human, turquoise for chimp) are labeled by the colored bars below sequence origin. Four active HARs (2xHAR.548, HAR152, 2xHAR.133, and 2xHAR.518) were tested *in vivo* to validate enhancer activity and are labeled on the right next to their corresponding rows.

Figure 7. HAR enhancer expression patterns for human and chimp sequences *in vivo*

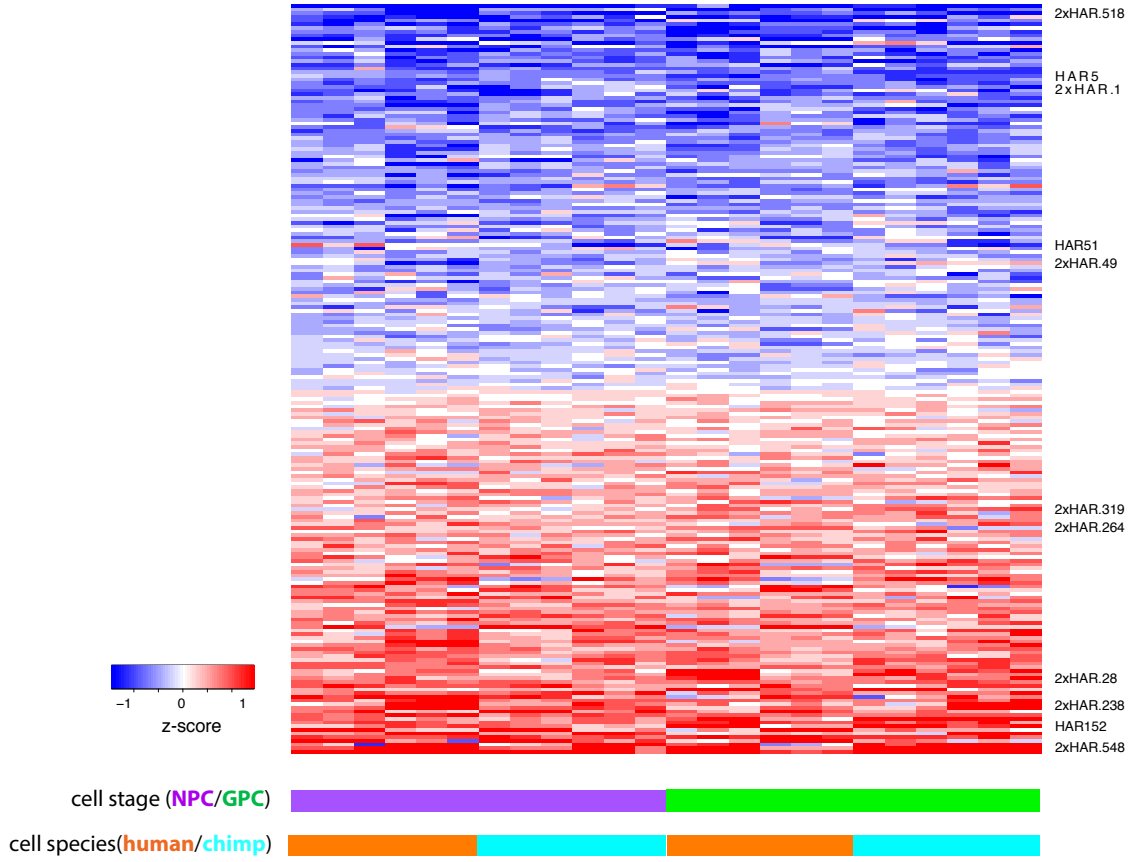


Representative HAR embryos showing enhancer patterns for the human (hg19) and chimpanzee (pt4) sequences for HAR152 at E10.5 and 2xHAR.548 at E13.5. Expression patterns for both the human and chimpanzee sequences for HAR512 were consistent in the brain stem. Expression patterns for 2xHAR.548 showed some suggested species-specific enhancer domains. The human sequence for 2xHAR.548 showed more consistent expression patterns in the mid- and hindbrain, whereas the expression pattern for the chimpanzee sequence was more consistent in the eye and neural tube. Nearby neurodevelopment-associated gene names are written in parenthesis next to the HAR ID.

Table 1. Top most differentially active HAR enhancers in primate NPCs and GPCs.

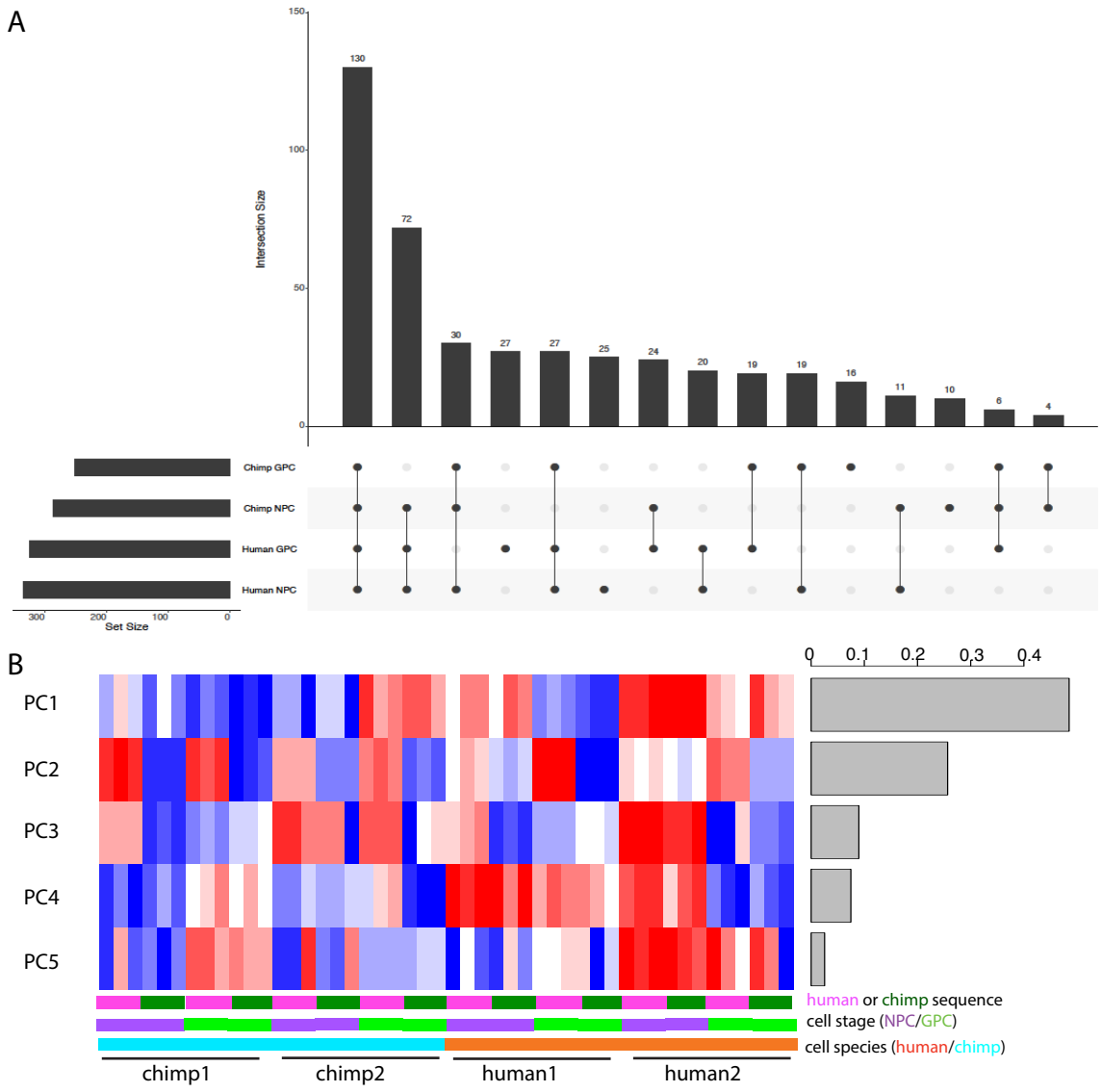
HAR	Abs(Fold-Change)	Adjusted P-Value	Species with greater enhancer activity	Nearby Genes within the same TAD
2xHAR.518	0.615539666	3.19E-08	Chimp	NRXN3, DIO2-AS1, LOC105370586, DIO2
2xHAR.548	0.575888017	1.78E-09	Human	FOXP1
2xHAR.401	0.350382349	1.90E-05	Chimp	RPAIN, USP6, RABEP1, LOC100130950, ZNF594
2xHAR.10	0.323575902	2.71E-10	Human	PAX8, PSD4
2xHAR.534	0.290302649	3.65E-08	Chimp	MIR873, LINGO2, MIR876
HAR59	0.259526542	2.75E-09	Chimp	GPC6
2xHAR.432	0.259190297	5.79E-06	Chimp	LOC284294, LINC00305, LINC01538
HAR152	0.252445701	5.15E-05	Human	NEUROG2, ALPK1
2xHAR.499	0.249619689	5.15E-05	Human	VEPH1, SHOX2
2xHAR.412	0.237701832	1.13E-15	Human	HOXD8

Figure 8. 204 HARs show species-specific enhancer activity across cell types.



The relative activity of human and chimpanzee HAR enhancers are depicted as z-scores for 204 differentially active HARs (adjusted p-value<0.01 across all samples). The three HARs that demonstrated significant cell species effects were 2xHAR.518, HAR51, and 2xHAR.264. The seven HARs that were significantly impacted by cell stage effects were 2xHAR.319, HAR5, 2xHAR.28, 2xHAR.238, 2xHAR.1, and 2xHAR.49.

Figure 9. HAR MPRA data shows that the impact of *cis* versus *trans* features.



(A) HAR enhancer activity is separated by four conditions - human NPCs, human GPCs, chimp NPCs, chimp GPCs. The number of HARs that exhibit activity in any given set of conditions is depicted by the histogram and the intersection of conditions are represented by the connected dots below the histogram. Most HARs that have enhancer activity are active in all four conditions. **(B)** Patterns in HAR enhancer activity separate out on the first five principal components. This heatmap depicts z-scores for the first five principal components that capture the majority of variation in enhancer activity for human and chimpanzee HAR sequences across all replicates, cell species, and cell stages. Annotations for sequence origin, cell stage, and cell species are all color-coded below the heatmap. The barplot on the right represents the proportion of variance accounted for per principal component.

Table 2. Species-specific TF motifs created by fixed differences in HAR enhancers

Transcription factors with more motifs in human	
TF	EGR1
	EGR2
	EGR3
	EGR4
	FOXP1
	HES7
	Hoxb5
	KLF5
	NFYB
	PLAG1
	PROP1
	SP1
	SP2
Transcription factors with more motifs in chimpanzee	
TF	EWSR1-FLI1
	Gfi1b
	HINFP
	HOXA13
	IRF1
	JDP2
	JUN
	JUND
	KLF16
	MEF2A
	NFAT5
	NFE2
	POU2F2
	Pou2f3
	POU4F1
	Pou5f1::Sox2
	PRDM1
	SP8
	ZNF263
	ZNF740

Table 3. GWAS SNPs within a 2MB window of HAR brain enhancers.

HAR	GWAS SNP(s)	Associated trait
2xHAR.270	rs8321, rs2021722	ASD,ADHD,BIP,MDD,SCZ
HAR48	rs12576775	ASD,ADHD,BIP,MDD,SCZ
2xHAR.511	rs11191454	ASD,ADHD,BIP,MDD,SCZ
2xHAR.396	rs7004633	ASD,ADHD,BIP,MDD,SCZ
2xHAR.273	rs10513249	BIP
2xHAR.85	rs6558872, rs10503253	ASD,ADHD,BIP,MDD,SCZ
2xHAR.149	rs4721295, rs10275045	SCZ,BIP
HAR3;2xHAR.26	rs4721295, rs10275045	SCZ,BIP
2xHAR.43	rs6703335	SCZ,BIP
2xHAR.274	rs6703335	SCZ,BIP

4 Summary

During my PhD studies, I was interested in the role of regulatory elements like enhancers in drug response and in human evolution, specifically in the context of uniquely human patterns in neurodevelopment. Previous work in the Ahituv lab demonstrated that antibiotics such as rifampin can activate drug-dependent regulatory regions in human hepatocytes. This motivated the project described in Chapter 2, which aimed to characterize the effect of using cell culture antibiotics (such as penicillin-streptomycin (PenStrep), and gentamicin) on gene expression and regulatory patterns in common human cell lines. Through RNA-sequencing and ChIP-sequencing, I was able to identify genome-wide changes in gene expression (209 DE genes) and regulation (9,514 DE H3K27ac peaks) in HepG2 cells due to penicillin-streptomycin at a standard cell culture concentration. I performed most of the wet lab work involved in performing the cell culture, RNA isolation, chromatin immunoprecipitation, and library prep. I used bioinformatics tools to perform alignment of reads from the RNA-seq and ChIP-seq data as well as some of the differential gene expression analysis. In Chapter 3, I designed and optimized the massively parallel reporter assay to test all human accelerated regions for brain enhancer activity in neural and glial progenitor cells I differentiated from human and chimpanzee iPSCs. With the help of my many collaborators, I was able to identify and quantify the regulatory activity of 306 novel brain enhancer HARs and 204 HARs that exhibited species-specific function. Through a joint effort with members of the Pollard group, we were able to quantify the relative impact of fixed differences, cell species, and cell stage on HAR enhancer activity. Our results show that *cis* features drive a greater proportion of variation in HAR enhancer activity than *trans* factors (cell species and stage). The overlap between TFBS motifs and fixed differences in the differentially active HAR enhancers further elucidates the potential mechanisms underlying uniquely human patterns in early neurodevelopment.

References

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

Arnold, C. D., Gerlach, D., Stelzer, C., Boryń, Ł. M., Rath, M., & Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science, 339(6123), 1074-1077.

Araujo, J., Breuer, P., Dieringer, S., Krauss, S., Dorn, S., Zimmermann, K., ... & Evert, B. O. (2011). FOXO4-dependent upregulation of superoxide dismutase-2 in response to oxidative stress is impaired in spinocerebellar ataxia type 3. Human molecular genetics, 20(15), 2928-2941.

Bar, J., Hasim, M. S., Baghai, T., Niknejad, N., Perkins, T. J., Stewart, D. J., ... & Dimitroulakos, J. (2016). Induction of Activating Transcription Factor 3 Is Associated with Cisplatin Responsiveness in Non–Small Cell Lung Carcinoma Cells. Neoplasia, 18(9), 525-535.

Barban, N., Jansen, R., De Vlaming, R., Vaez, A., Mandemakers, J. J., Tropf, F. C., ... & Tragante, V. (2016). Genome-wide analysis identifies 12 loci influencing human reproductive behavior.

Benko, S., Fantes, J. A., Amiel, J., Kleinjan, D. J., Thomas, S., Ramsay, J., ... & McBride, D. (2009). Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. Nature genetics, 41(3), 359-364.

Berry, F. B., Skarie, J. M., Mirzayans, F., Fortin, Y., Hudson, T. J., Raymond, V., ... & Walter, M. A. (2007). FOXC1 is required for cell viability and resistance to oxidative stress in the eye through the transcriptional regulation of FOXO1A. Human molecular genetics, 17(4), 490-505.

Capra, J. A., Erwin, G. D., McKinsey, G., Rubenstein, J. L., & Pollard, K. S. (2013). Many human accelerated regions are developmental enhancers. Phil. Trans. R. Soc. B, 368(1632), 20130025.

Chen, G. K., Marjoram, P., & Wall, J. D. (2009). Fast and flexible simulation of DNA sequence data. Genome research, 19(1), 136-142.

Clément, M., Basatemur, G., Masters, L., Baker, L., Bruneval, P., Iwawaki, T., ... & Mallat, Z. (2016). Necrotic Cell Sensor Clec4e Promotes a Pro-Atherogenic Macrophage Phenotype Through Activation of the Unfolded Protein Response. Circulation, Circulationaha-116.

Conway, J. R., Lex, A., & Gehlenborg, N. (2017). UpSetR: An R Package For The Visualization Of Intersecting Sets And Their Properties. bioRxiv, 120600.

Cretekos, C. J., Wang, Y., Green, E. D., Martin, J. F., Rasweiler, J. J., Behringer, R. R., & NISC Comparative Sequencing Program. (2008). Regulatory divergence modifies limb length between mammals. Genes & development, 22(2), 141-151.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., ... & Boyer, L. A. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proceedings of the National Academy of Sciences, 107(50), 21931-21936.

Crittenden, J. R., Heidersbach, A., & McManus, M. T. (2007). Lentiviral strategies for RNAi knockdown of neuronal genes. Current Protocols in Neuroscience, 5-26.

Denk, F., Ramer, L. M., Erskine, E. L., Nassar, M. A., Bogdanov, Y., Signore, M., ... & Ramer, M. S. (2015). Tamoxifen induces cellular stress in the nervous system by inhibiting cholesterol synthesis. Acta neuropathologica communications, 3(1), 74.

Di Masi, A., De Marinis, E., Ascenzi, P., & Marino, M. (2009). Nuclear receptors CAR and PXR: Molecular, functional, and biomedical aspects. Molecular aspects of medicine, 30(5), 297-343.

Dickel, D. E., Zhu, Y., Nord, A. S., Wylie, J. N., Akiyama, J. A., Afzal, V., ... & Visel, A. (2014). Function-based identification of mammalian enhancers using site-specific integration. Nature methods, 11(5), 566-571.

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

Edson, R. S., & Keys, T. F. (1983, February). The aminoglycosides. Streptomycin, kanamycin, gentamicin, tobramycin, amikacin, netilmicin, sisomicin. In Mayo Clinic proceedings (Vol. 58, No. 2, pp. 99-102).

ENCODE Project Consortium. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature, 489(7414), 57.

Euskirchen, G. M., Auerbach, R. K., Davidov, E., Gianoulis, T. A., Zhong, G., Rozowsky, J., ... & Snyder, M. (2011). Diverse roles and interactions of the SWI/SNF chromatin remodeling complex revealed using global approaches. PLoS genetics, 7(3), e1002008.

Farh, K. K. H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W. J., Beik, S., ... & Hatan, M. (2015). Genetic and Epigenetic Fine-Mapping of Causal Autoimmune Disease Variants. Nature, 518(7539), 337.

Faure, A. J., Schmidt, D., Watt, S., Schwalie, P. C., Wilson, M. D., Xu, H., ... & Flicek, P. (2012). Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. Genome research, 22(11), 2163-2175.

Fukasawa, K., Park, G., Iezaki, T., Horie, T., Kanayama, T., Ozaki, K., ... & Kitajima, S. (2016). ATF3 controls proliferation of osteoclast precursor and bone remodeling. Scientific reports, 6, 30918.

Gauthier, J., Siddiqui, T. J., Huashan, P., Yokomaku, D., Hamdan, F. F., Champagne, N., ... & Fathalli, F. (2011). Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. Human genetics, 130(4), 563-573.

Germovsek, E., Barker, C. I., & Sharland, M. (2016). What do I need to know about aminoglycoside antibiotics?. Archives of Disease in Childhood-Education and Practice, edpract-2015.

Goldstein, A. L., & Johnson, P. R. (1982). Primary culture of adipoblasts from obese and lean Zucker rat adipose tissue. Metabolism, 31(6), 601-607.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... & Hansen, N. F. (2010). A draft sequence of the Neandertal genome. science, 328(5979), 710-722.

Hailey, D. W., Esterberg, R., Linbo, T. H., Rubel, E. W., & Raible, D. W. (2017). Fluorescent aminoglycosides reveal intracellular trafficking routes in mechanosensory hair cells. The Journal of clinical investigation, 127(2), 472.

Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B., & King, M. C. (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. Science, 1684-1689.

Hansen, D. V., Lui, J. H., Parker, P. R., & Kriegstein, A. R. (2010). Neurogenic radial glia in the outer subventricular zone of human neocortex. Nature, 464(7288), 554-561.

Hiatt, J. B., Patwardhan, R. P., Turner, E. H., Lee, C., & Shendure, J. (2010). Parallel, tag-directed assembly of locally derived short sequence reads. Nature methods, 7(2), 119-122.

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proceedings of the National Academy of Sciences, 106(23), 9362-9367.

Hneino, M., Blirando, K., Buard, V., Tarlet, G., Benderitter, M., Hoodless, P., ... & Milliat, F. (2012). The TG-interacting factor TGIF1 regulates stress-induced proinflammatory phenotype of endothelial cells. Journal of Biological Chemistry, 287(46), 38913-38921.

Hu, X., Zhang, J., Jin, C., Mi, W., Wang, F., Ma, W., ... & Du, B. (2013). Association study of NRXN3 polymorphisms with schizophrenia and risperidone-induced bodyweight gain in Chinese Han population. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 43, 197-202.

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols, 4(1), 44.

Jang, M. K., & Jung, M. H. (2015). ATF3 inhibits PPARγ-stimulated transactivation in adipocyte cells. Biochemical and biophysical research communications, 456(1), 80-85.

Kim, J. Y., Park, K. J., Kim, G. H., Jeong, E. A., Lee, D. Y., Lee, S. S., ... & Kim, W. H. (2013). In vivo activating transcription factor 3 silencing ameliorates the AMPK compensatory effects for ER stress-mediated β-cell dysfunction during the progression of type-2 diabetes. Cellular signalling, 25(12), 2348-2361.

Kim, K. H., Park, B., Rhee, D. K., & Pyo, S. (2015). Acrylamide induces senescence in macrophages through a process involving ATF3, ROS, p38/JNK, and a telomerase-independent pathway. Chemical research in toxicology, 28(1), 71-86.

King, M. C., & Wilson, A. C. (1975). Evolution at two levels in humans and chimpanzees.

Knapska, E., & Kaczmarek, L. (2004). A gene for neuronal plasticity in the mammalian brain: Zif268/Egr-1/NGFI-A/Krox-24/TIS8/ZENK?. Progress in neurobiology, 74(4), 183-211.

Koh, I. U., Lim, J. H., Joe, M. K., Kim, W. H., Jung, M. H., Yoon, J. B., & Song, J. (2010). AdipoR2 is transcriptionally regulated by ER stress‐inducible ATF3 in HepG2 human hepatocyte cells. The FEBS journal, 277(10), 2304-2317.

Kothary, R., Clapoff, S., Brown, A., Campbell, R., Peterson, A., & Rossant, J. (1988). A transgene containing lacZ inserted into the dystonia locus is expressed in neural tube. Nature, 335(6189), 435-437.

Kuo, J. R., Shang, H. S., Ho, C. T., Lai, K. G., Liu, T. Z., Chen, Y. J., & Chiou, J. F. (2016). Cisplatin-induced regulation of signal transduction pathways and transcription factors in p53-mutated subclone variants of hepatoma cells: Potential application for therapeutic targeting. Oncology letters, 12(5), 3723-3730.

Kwon, J. W., Kwon, H. K., Shin, H. J., Choi, Y. M., Anwar, M. A., & Choi, S. (2015). Activating transcription factor 3 represses inflammatory responses by binding to the p65 subunit of NF-κB. Scientific reports, 5.

Labzin, L. I., Schmidt, S. V., Masters, S. L., Beyer, M., Krebs, W., Klee, K., ... & De Nardo, D. (2015). ATF3 is a key regulator of macrophage IFN responses. The Journal of Immunology, 195(9), 4446-4455.

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., ... & Chen, Y. (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome research, 22(9), 1813-1831.

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

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., ... & de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Human molecular genetics, 12(14), 1725-1735.

Li, Q., Barkess, G., & Qian, H. (2006). Chromatin looping and the probability of transcription. Trends in Genetics, 22(4), 197-202.

Liu, C., Luan, S., OuYang, H., Huang, Z., Wu, S., Ma, C., ... & Xin, W. (2016). Upregulation of CCL2 via ATF3/c-Jun interaction mediated the Bortezomib-induced peripheral neuropathy. Brain, behavior, and immunity, 53, 96-104.

Liu, Y. F., Wei, J. Y., Shi, M. H., Jiang, H., & Zhou, J. (2016). Glucocorticoid Induces Hepatic Steatosis by Inhibiting Activating Transcription Factor 3 (ATF3)/S100A9 Protein Signaling in Granulocytic Myeloid-derived Suppressor Cells. Journal of Biological Chemistry, 291(41), 21771-21785.

Liu, C., Yao, J., Mercola, D., & Adamson, E. (2000). The transcription factor EGR-1 directly transactivates the fibronectin gene and enhances attachment of human glioblastoma cell line U251. Journal of Biological Chemistry, 275(27), 20315-20323.

Llobet, L., Montoya, J., López-Gallardo, E., & Ruiz-Pesini, E. (2015). Side effects of culture media antibiotics on cell differentiation. Tissue Engineering Part C: Methods, 21(11), 1143-1147.

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.

Lu, Y., Li, T., Qureshi, H. Y., Han, D., & Paudel, H. K. (2011). Early growth response 1 (Egr-1) regulates phosphorylation of microtubule-associated protein tau in mammalian brain. Journal of Biological Chemistry, 286(23), 20569-20581.

Luizon, M. R., Eckalbar, W. L., Wang, Y., Jones, S. L., Smith, R. P., Laurance, M., ... & Zhou, Y. (2016). Genomic characterization of metformin hepatic response. PLoS genetics, 12(11), e1006449.

Martinex‐Liarte, J. H., Solano, F., & Lozano, J. A. (1995). Effect of Penicillin‐Streptomycin and Other Antibiotics on Melanogenic Parameters in Cultured B16/F10 Melanoma Cells. Pigment Cell & Melanoma Research, 8(2), 83-88.

Mathis, J. M., Simmons, D. M., He, X., Swanson, L. W., & Rosenfeld, M. G. (1992). Brain 4: a novel mammalian POU domain transcription factor exhibiting restricted brain-specific expression. The EMBO journal, 11(7), 2551.

Maurano, M. T., Humbert, R., Rynes, E., Thurman, R. E., Haugen, E., Wang, H., ... & Shafer, A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science, 337(6099), 1190-1195.

McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., ... & Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nature biotechnology, 28(5), 495-501.

McLean, C. Y., Reno, P. L., Pollen, A. A., Bassan, A. I., Capellini, T. D., Guenther, C., ... & Wenger, A. M. (2011). Human-specific loss of regulatory DNA and the evolution of human-specific traits. Nature, 471(7337), 216.

Mikkelsen, T. S., Hillier, L. W., Eichler, E. E., & Zody, M. C. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. Nature, 437(7055), 69.

Mucsi, Z., Chass, G. A., Ábrányi-Balogh, P., Jójárt, B., Fang, D. C., Ramirez-Cuesta, A. J., ... & Csizmadia, I. G. (2013). Penicillin's catalytic mechanism revealed by inelastic

neutrons and quantum chemical theory. Physical Chemistry Chemical Physics, 15(47), 20447-20455.

Ning, Z., McLellan, A. S., Ball, M., Wynne, F., O'neill, C., Mills, W., ... & O'keeffe, G. (2015). Regulation of SPRY3 by X chromosome and PAR2-linked promoters in an autism susceptibility region. Human molecular genetics, 24(18), 5126-5141.

Noctor, S. C., Flint, A. C., Weissman, T. A., Dammerman, R. S., & Kriegstein, A. R. (2001). Neurons derived from radial glial cells establish radial units in neocortex. Nature, 409(6821), 714.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., ... & Oskarsson, S. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. Nature, 533(7604), 539.

Ongen, H., Andersen, C. L., Bramsen, J. B., Oster, B., Rasmussen, M. H., Ferreira, P. G., ... & Padioleau, I. (2014). Putative cis-regulatory drivers in colorectal cancer. Nature, 512(7512), 87.

Patwardhan, R. P., Hiatt, J. B., Witten, D. M., Kim, M. J., Smith, R. P., May, D., ... & Ahituv, N. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. Nature biotechnology, 30(3), 265-270.

Peter, A., Weigert, C., Staiger, H., Machicao, F., Schick, F., Machann, J., ... & Schleicher, E. (2009). Individual stearoyl-coa desaturase 1 expression modulates endoplasmic reticulum stress and inflammation in human myotubes and is associated with skeletal muscle lipid storage and insulin sensitivity in vivo. Diabetes, 58(8), 1757-1765.

Pollard, K. S., Salama, S. R., King, B., Kern, A. D., Dreszer, T., Katzman, S., ... & Rosenbloom, K. R. (2006). Forces shaping the fastest evolving regions in the human genome. PLoS genetics, 2(10), e168.

Pollard, K. S., Salama, S. R., Lambert, N., Lambot, M. A., Coppens, S., Pedersen, J. S., ... & Kern, A. D. (2006). An RNA gene expressed during cortical development evolved rapidly in humans. Nature, 443(7108), 167.

Prabhakar, S., Visel, A., Akiyama, J. A., Shoukry, M., Lewis, K. D., Holt, A., ... & Pennacchio, L. A. (2008). Human-specific gain of function in a developmental enhancer. Science, 321(5894), 1346-1350.

Ptashne, M. (1986). Gene regulation by proteins acting nearby and at a distance. Nature, 322(6081), 697-701.

Qi, L., Saberi, M., Zmuda, E., Wang, Y., Altarejos, J., Zhang, X., ... & Olefsky, J. (2009). Adipocyte CREB promotes insulin resistance in obesity. Cell metabolism, 9(3), 277-286.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A., & Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature, 470(7333), 279.

Ruderfer, D. M., Fanous, A. H., Ripke, S., McQuillin, A., Amdur, R. L., Gejman, P. V., ... & Kelsoe, J. R. (2014). Polygenic dissection of diagnosis and clinical dimensions of bipolar disorder and schizophrenia. Molecular psychiatry, 19(9), 1017-1024.

Rujescu, D., Ingason, A., Cichon, S., Pietiläinen, O. P., Barnes, M. R., Toulopoulou, T., ... & Murray, R. (2008). Disruption of the neurexin 1 gene is associated with schizophrenia. Human molecular genetics, 18(5), 988-996.

Ryu, A. H., Eckalbar, W. L., Kreimer, A., Yosef, N., & Ahituv, N. (2017). Use antibiotics in cell culture with caution: genome-wide identification of antibiotic-induced changes in gene expression and regulation. Scientific Reports.

Saitsu, H., Sonoda, M., Higashijima, T., Shirozu, H., Masuda, H., Tohyama, J., ... & Miyatake, S. (2016). Somatic mutations in GLI3 and OFD1 involved in sonic hedgehog signaling cause hypothalamic hamartoma. Annals of clinical and translational neurology, 3(5), 356-365.

Schmidt, D., Wilson, M. D., Ballester, B., Schwalie, P. C., Brown, G. D., Marshall, A., ... & Talianidis, I. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. Science, 328(5981), 1036-1040.

Schwarze, P. E., Solheim, A. E., & Seglen, P. O. (1982). Amino acid and energy requirements for rat hepatocytes in primary culture. In Vitro Cellular & Developmental Biology-Plant, 18(1), 43-54.

Shao, Z., Zhang, Y., Yuan, G. C., Orkin, S. H., & Waxman, D. J. (2012). MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. Genome biology, 13(3), R16.

Sharma, D., Cukras, A. R., Rogers, E. J., Southworth, D. R., & Green, R. (2007). Mutational analysis of S12 protein and implications for the accuracy of decoding by the ribosome. Journal of molecular biology, 374(4), 1065-1076.

Shen, J., Lin, H., Li, G., Jin, R. A., Shi, L., Chen, M., ... & Cai, X. (2016). TR4 nuclear receptor enhances the cisplatin chemo-sensitivity via altering the ATF3 expression to better suppress HCC cell growth. Oncotarget, 7(22), 32088.

Smart, I. H., Dehay, C., Giroud, P., Berland, M., & Kennedy, H. (2002). Unique morphological features of the proliferative zones and postmitotic compartments of the neural epithelium giving rise to striate and extrastriate cortex in the monkey. Cerebral cortex, 12(1), 37-53.

Smemo, S., Campos, L. C., Moskowitz, I. P., Krieger, J. E., Pereira, A. C., & Nobrega, M. A. (2012). Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. Human molecular genetics, 21(14), 3255-3263.

Smith, R. P., Eckalbar, W. L., Morrissey, K. M., Luizon, M. R., Hoffmann, T. J., Sun, X., ... & Liu, Y. (2014). Genome-wide discovery of drug-dependent human liver regulatory elements. PLoS genetics, 10(10), e1004648.

Smith, R. P., Taher, L., Patwardhan, R. P., Kim, M. J., Inoue, F., Shendure, J., ... & Ahituv, N. (2013). Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. Nature genetics, 45(9), 1021-1028.

Sun, R., Jiang, B., Qi, H., Zhang, X., Yang, J., Duan, J., ... & Li, G. (2015). SOX4 contributes to the progression of cervical cancer and the resistance to the chemotherapeutic drug through ABCG2. Cell death & disease, 6(11), e1990.

Tiscornia, G., Singer, O., & Verma, I. M. (2006). Production and purification of lentiviral vectors. Nature protocols, 1(1), 241.

Tothova, Z., Kollipara, R., Huntly, B. J., Lee, B. H., Castrillon, D. H., Cullen, D. E., ... & Armstrong, S. A. (2007). FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. Cell, 128(2), 325-339.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., ... & Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology, 28(5), 511-515.

Tsai, W. W., Matsumura, S., Liu, W., Phillips, N. G., Sonntag, T., Hao, E., ... & Montminy, M. (2015). ATF3 mediates inhibitory effects of ethanol on hepatic gluconeogenesis. Proceedings of the National Academy of Sciences, 112(9), 2699-2704.

VanderMeer, J. E., & Ahituv, N. (2011). cis‐regulatory mutations are a genetic cause of human limb malformations. Developmental Dynamics, 240(5), 920-930.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., ... & Afzal, V. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature, 457(7231), 854.

Weng, S., Zhou, L., Deng, Q., Wang, J., Yu, Y., Zhu, J., & Yuan, Y. (2016). Niclosamide induced cell apoptosis via upregulation of ATF3 and activation of PERK in Hepatocellular carcinoma cells. BMC gastroenterology, 16(1), 25.

Whitmore, M. M., Iparraguirre, A., Kubelka, L., Weninger, W., Hai, T., & Williams, B. R. (2007). Negative regulation of TLR-signaling pathways by activating transcription factor-3. The Journal of Immunology, 179(6), 3622-3630.

Wu, J. Q., Habegger, L., Noisa, P., Szekely, A., Qiu, C., Hutchison, S., ... & Cui, W. (2010). Dynamic transcriptomes during neural differentiation of human embryonic stem cells revealed by short, long, and paired-end sequencing. Proceedings of the National Academy of Sciences, 107(11), 5254-5259.

Zhang, Y., Gan, B., Liu, D., & Paik, J. H. (2011). FoxO family members in cancer. Cancer biology & therapy, 12(4), 253-259.

Zur Nieden, N. I., Kempka, G., & Ahr, H. J. (2004). Molecular multiple endpoint embryonic stem cell test—a possible approach to test for the teratogenic potential of compounds. Toxicology and applied pharmacology, 194(3), 257-269.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature Ann Hane Ryu Date 09/08/2017