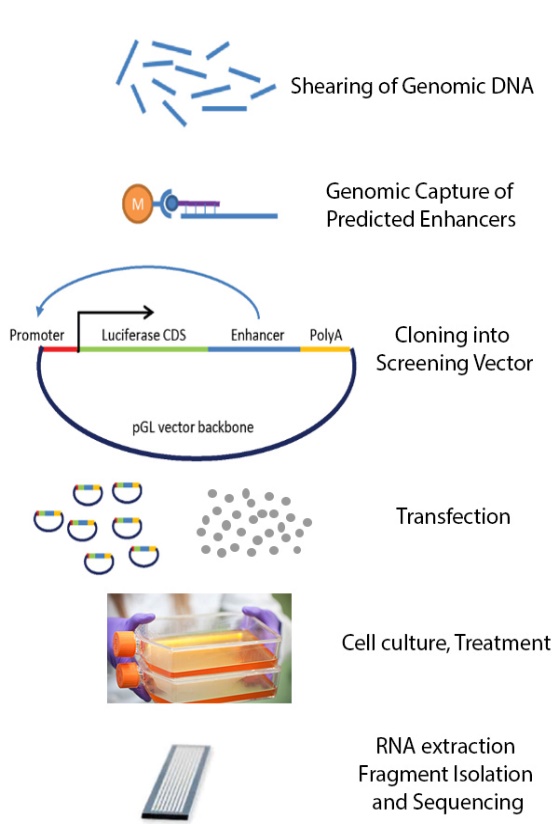
**Aim 3. Systematic genome-wide validation of PsychENCODE regulatory elements**

**3a. Overview**. The results of Aims 1 and 2 will constitute a series of predictions of which regulatory elements are functionally relevant for the expression of neuronal genes involved in normal development and/or neuropsychiatric disorders that are a focus of the PsychENCODE Consortium. Biological validation of these predictions requires testing the regulatory potential of these regions, including allele-specific quantification.

**3b. Validating enhancers on a genome-wide scale**. ***Preliminary:*** Over the last several years, the White Lab has led efforts to functionally validate regulatory elements (as identified by the ENCODE and modENCODE projects) based on predictions made by the Gerstein and Weng Labs [1-4] (Zhang et al., submitted). Most recently, our labs have worked together with Drs. Geschwind and Liu on data production and analysis for the PsychENCODE Consortium[5]. Also, we have recently developed a protocol for testing enhancers genome-wide, based on the **STARR-seq** (Self-Transcribing Active Regulatory Region sequencing) methodology originally developed in *Drosophila* by the Stark Lab and used by others on limited regions of mammalian genomes [6, 7]. STARR-seq involves the insertion of putative enhancers into the transcript, instead of upstream of promoters in the reporter vector. The enhancer sequence effectively acts as a barcode in high-throughput sequencing (Fig 2). More specifically, genomic DNA is sheared and end-repaired (Fig 2A), optionally captured if targeted regions are to be screened (as oppose to the entire genome – see Fig 2B), and subsequently cloned into screening vectors containing a promoter, which then expresses a reporter transcript (Fig 2C). The enhancers are cloned into the 3’ end of the transcript, whereby the reporter transcript will contain the enhancer sequence. This pool of screening vectors is transfected into cells (Fig 2D) and then cultured under appropriate conditions for recovery, growth, and/or differentiation (Fig 2E). mRNA is purified and reverse transcribed, followed by uniform amplification of the inserts, and then sequenced using high-throughput sequencing (Fig 2F). Abundant copies of the reporter transcripts that contain specific enhancers can identify enhancers that up-regulate transcription. STARR-seq removes the need for expensive array synthesis of enhancers while capturing natural variation for quantitative analysis of enhancer function.



**Figure 2**

A

B

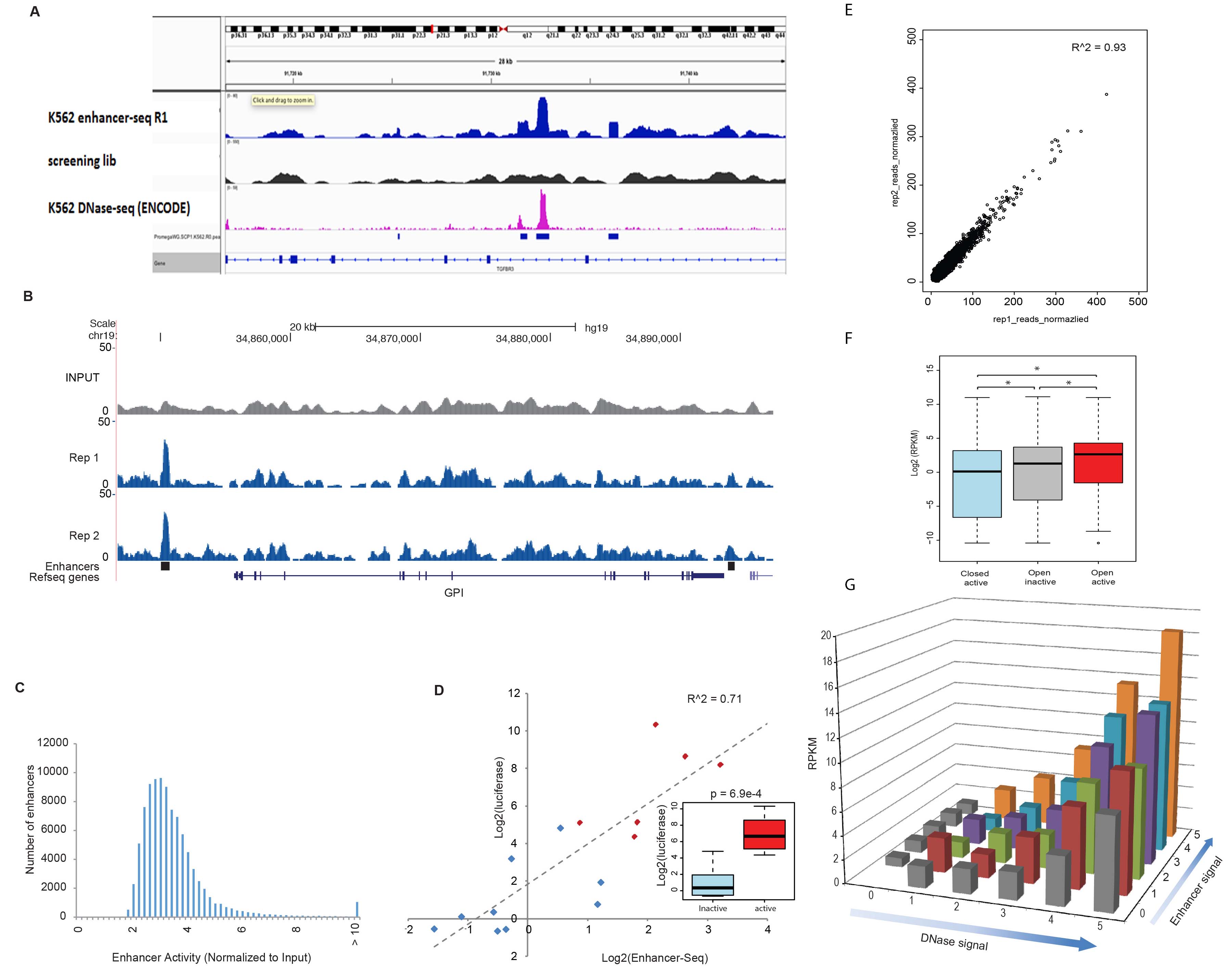
C

D

E

F

Prior to our optimizations, there have been major limitations to scaling this methodology to the whole human genome. For example, screening the *Drosophila melanogaster* genome required transfecting between 0.5 and 1 billion S2 cells. This makes the direct application of STARR-seq technique to the human genome very difficult and expensive, because the human genome is 20 times larger than the fly genome. Our optimizations of STARR-seq in human cells modifies and builds upon the episomal plasmid library approach, expanding its capabilities. We have overcome the major challenges for scaling STARR-seq up to the entire human genome (namely, the required library complexity, large-scale transfection of cells, and inherent inaccuracy of the assay due to PCR duplicates during the sequencing step introduces significant challenges). First, by optimizing multiple parameters in the candidate element cloning step, we have increased complexity while introducing molecular barcodes that allow for PCR duplicate elimination, resulting in a screening library that covers 2.65 Gb of the human genome. Our typical libraries now have more than 50 fragments covering each base pair, given ~250 million post-filtering fragments. This represents a comprehensive screening library, and allows us to effectively screen genomic fragments with enhancer activity in downstream experiments. Second, using industrial-scale transfection protocols, the White Lab has devised a robust technique to screen either the entire human genome or a fraction of the genome that has been captured using oligonucleotide probes (**CapStarr-seq**). We are producing whole genome STARR-seq datasets at a coverage ofmore than 10X per expressed base pair, with 200-300 million paired end 100bp reads, or significantly fewer reads for capture STARR-seq. Third, accuracy has been improved by using single molecular barcodes during the RNA preparation step, along with 160 or more index primers for sequencing, thus allowing for more accurate quantification and elimination of PCR duplicates without removing unique RNA fragments associated with *bone fide* transcriptional activity. Example results are shown in Fig 3. The end result of our optimizations and improvements is that active enhancer regions of the genome are easily identified (Fig 3A, C), the results are highly reproducible (Fig 3B, E), and consistent with traditional reporter assays (Fig 3D), and sequences with enhancer activity overlap with open chromatin marks (Fig 3A, F, G) as well as with RNA expression levels from nearby genes (Fig 3G).

****

**Figure 3. Whole genome [[dc2all: should this be written as “whole genome” or “whole-genome”?]] STARR-seq (A)** Genome browser screen shot shows consistency between the K562 human-STARR-seq signal (1st row) and DNase-seq signal (3rd row). **(B)** Genomic snapshot displaying the *GPI* locus region, as detected by WG-STARR-seq. There is a strong enhancer region approximately 10-kb upstream of *GPI* and another, weaker enhancer regions in the 3’UTR region. Each blue track signifies a normalized enhancer signal of each biological replicate. The gray track represents the normalized input library. **(C)** WG-STARR-seq shows a wide range of enhancer signal strength distributions for all detected enhancers. The median fold change observed was 3.08, with a dynamic range between 1.33 and 119.12. **(D)** The enhancer activity of 6 strong and 9 weak enhancers were validated using traditional luciferase assays in biological triplicates. A strong correlation was observed between luciferase signal and WG-STARR-seq enhancer activity, providing validation of the technique. **(E)** Normalized reads from sequencing were used for reproducibility plots between biological replicates. **(F)** Comparison of expression levels between genes (as measured by RPKM) nearby different groups of enhancers. Statistical significance was calculated using Wilcox Sum Rank test (\*p = 2.2e-16). **(G)** Plot comparing expression level of nearby genes in relation to both DNase I signal and enhancer activity. Both DNase I signal and enhancer signals are binned into 6 separate groups according to DNase I signal and enhancer signal rank (0 – 5), respectively.

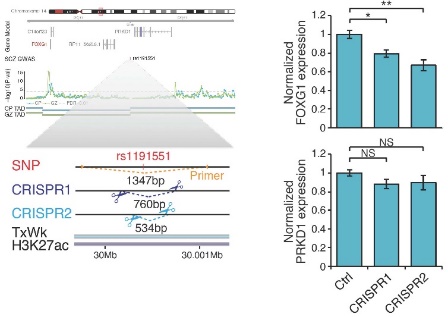
***Plan:***In years 1 and 2 of the project, we will focus our whole genome and STARR-seq efforts on validating candidate enhancers from the PsychENCODE Project (Aims 1 and 2) using SH-SY5Y neuroblastoma cell lines (since billions of cells can easily be grown and transfected, each replicate requires approximately 1 billion cultured cells). While not ultimately an ideal model for testing the disease and developmental relevance of predicted enhancers in a nervous system-specific manner, SH-SY5Y cells share neuronal molecular characteristics with the primary human neuronal precursor cells (phNPCs), and they match *in vivo* fetal brain development to a substantial extent once differentiated in RA and BDNF (see Fig 1, Aim 2 and [8]). They are a more suitable cell line than a non-neural cell when performing unbiased high-throughput screening of the entire human genome for enhancer activity. The data from these experiments will then be used to improve and refine the predictions made in Aim 2.

In Years 2-4 we will turn our focus to phNPCs, and we will use the CapStarr-seq method with specific sets of predictions from Aim 2. Although not genome-wide, CapStarr-seq allows tens of thousands of DNA elements to be tested per experiment. Multiple rounds will be performed as prediction algorithms are iteratively refined in Aims 1 and 2, taking into account the STARR-seq data. For whole genome STARR-seq and for CapStarr-seq, we will perform 3 biological replicates. We will use the phNPCs because they can be grown in 96 well plates, and we have optimized conditions with over 70% transfection efficiency [9-11]. We will analyze at least 2 time points, T0 (48 after being placed in differentiation media, consisting primarily of neural progenitors) and T6 (6 weeks, to capture development and maturation of different neuronal lineages and asytrocytes)[8, 9]. Cell types to be used in Aims 3 and 4 are shown in the Table below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Cell Name** | **Tissue of Origin** | **Cell Type** | **Obtained from** | **Media/Culture Conditions** |
| D8R49 | Fetal Cortex | phNPC [8] | Primary | Neurobasal + BIT9500 +GF (Proliferation) Neurobasal +B27 + GF (Differentiation) |
| 2242.1 | Skin Fibroblast | iPSC [12, 13] | Pasca Lab | E8 Medium + E8 Supp (Proliferation) |
| 2242.1 | iPSC | Forebrain Spheroid [14] | Pasca Lab | E6 Medium + E6 Supp + GF (Differentiation) |
| SH-SY5Y | Bone Marrow Neuroblastoma | Adherent [8] | ATCC | DMEM + 10% FBS |

**STARR-seq analysis:** We will use established STARR-seq data processing pipelines developed by the Gerstein and White groups, including the definitions of signal profiles defined by PeakSeq [15] and MUSIC [16] to generate a relaxed set of regions that show significant enrichment. These regions will be highly sensitive but will contain many false positives. Therefore, we will use the large compendium of existing functional genomics datasets from the PsychENCODE, ENCODE and RMEC projects, utilizing peaks from histone marks and transcription factors to build *a priori* probability estimates of the locations of regulatory regions. We will use the activating marks and transcription factors that associate with enhancers (H3K4me1, H3K27ac, H3K9ac, P300, ATAC-seq, DNase/FAIRE) to build these probabilities. We will also utilize transcription factor binding motif and sequence conservation data as variants in the *a priori* location estimates. We will then combine the whole genome STARR-seq results with these probabilities in a Bayesian framework, and we will train generalized linear models for scoring the candidate relaxed list of regions that we identified from STARR-seq. The sorted list of regions will be used for further validation in the CRISPR mutational assays (detailed below) and the single cell transgenic reporter assays (described in Aim 4).

**3c. Validating enhancers using CRISPR genomic editing. *Preliminary:*** A major caveat of the STARR-seq method is that all assays are done on transfected plasmids that lack the genomic context of the loci from which they have been derived. The advantage of this method is that it allows us to systematically test thousands of genomic regions for regulatory potential. However, for many loci it can be only a modest indicator of the actual regulatory function within the native genomic context. Based on previous experience, we expect that data collected will narrow the regions of interest from tens of thousands of candidate enhancer elements down to hundreds or thousands of partially validated enhancer elements. We will prioritize lists of functionally-validated enhancer elements based on the results from Aim 2. We will then test these regulatory elements using CRISPR-mediated genome editing to determine which regulatory elements show differential function when mutated. Fig 4 shows a recent experiment from the Geschwind Lab [17] using two independent CRISPR constructs to demonstrate the effect on *FOXG1* expression (measured by qRT-PCR) of a specific enhancer associated with SCZ risk that they also showed has allele-specific variants (see Won et al. *Nature* 538:523 2016 for more details[17]). We aim to test 200 such candidates associated with normal developmental or disease states, based on the STARR-seq results and predictions from Aim 2.



**Figure 4**

**Plan:** Using phNPCs, we will functionally validate approximately 200 candidate enhancer regions by endogenously mutating them using the CRISPR/Cas9 targeting system. By utilizing CRISPR technology, we are able to edit the genome using CRISPR and CRISPR-associated (Cas) genes that have been exploited to achieve site-specific DNA recognition and cleavage [18]. In this fashion, not only are we interrogating our target enhancers in their endogenous chromatin context, but we will also be able to obtain a clearer picture of which gene(s) the regulatory element may control. We will generate loss-of-function mutations in putative enhancers using a 96-well plate format and use a qRT-PCR (quantitative reverse transcriptase PCR) assay of nearby gene transcripts to generate quantitative transcriptional read outs. Combining the STARR-seq and CRISPR enhancer mutational data procured in Aim 3, along with the analytical framework to prioritize disease and neuronal subtype/developmental-specific enhancers for each cell type, will enable us to select and target the most disease-relevant enhancers for further testing and evaluation in Aim 4.

It is worth noting that, collectively, these experiments will be performed in parallel with similar experiments being performed on the ENCODE cell lines and on pancreatic tumor cells and organoids that rely on performance of high-throughput gene editing using CRISPR-Cas9 to create a deletion (ENCODE grants UM1 HG009442 and UM1 HG009426), and will thus leverage considerable infrastructure and expertise as the need for any troubleshooting arises. Specific steps that may require optimization include transfection/electroporation efficiency, iteration of library building based on updated predictions from Aim 2, and optimizing timing of harvesting transfected phNPCs after differentiation into neuronal lineages. However, we have used these cells in many studies and regularly obtained over 70% transfection efficiency [8-11].

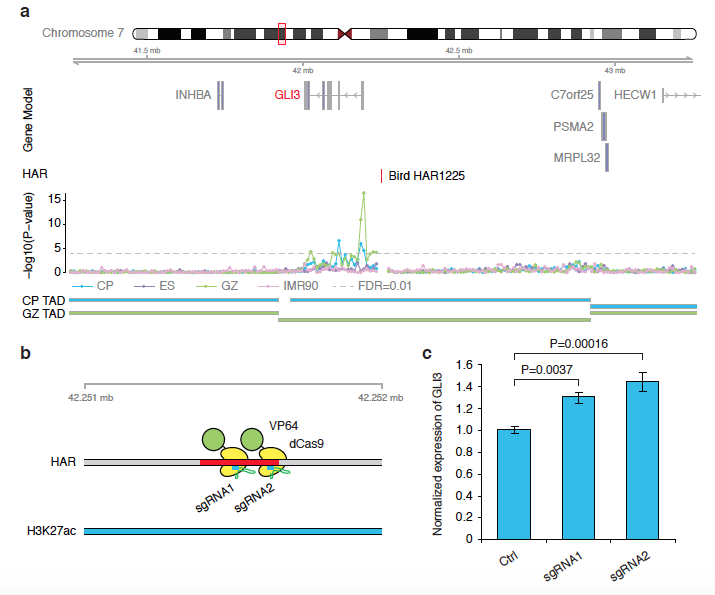
**Aim 4. Deeper biological validation of PsychENCODE regulatory elements**

**4a. Overview.** Geschwind and colleagues have developed novel primary neural progenitor cultures that model brain development *in vitro* (phNPCs) and 3D organoid cultures based on iPS cells differentiated into forebrain cortical spheroids (hFS) that recapitulate all of the major cell classes of the developing brain. Savas Tay has been able to grow such neuronal organoids in microfluidic chips that allow the rapid testing of large numbers of conditions [19]. Choosing from validated enhancers from Aim 3, we will synthesize 100 validated enhancers with polymorphisms predicted to affect function between alleles, and we will test these in phNPC cells differentiated into various lineages and under different conditions in plate-based assays and microfluidic chips. Additionally for 10-15 enhancers that the criteria of differential alleic expression associated with disease, we will transfect them into phNPCs, followed by neuronal differentiation, and we will perform Drop-Seq based single cell sequencing to associate the functioan of these enhancers with particular cell types and developmental stages.

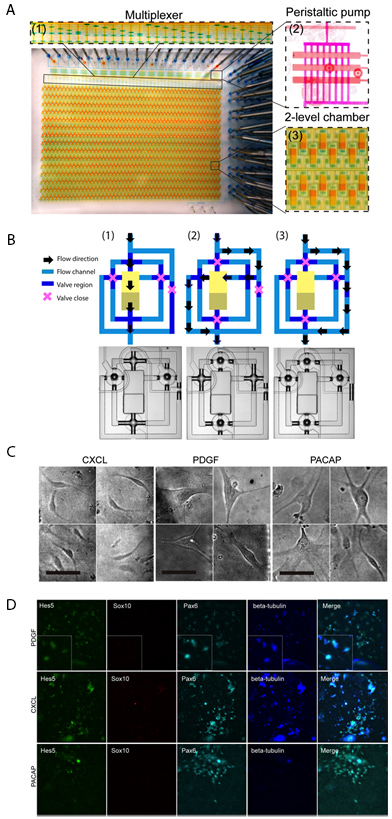
**4b. *In vitro* modeling. *Preliminary*:** Human neural stem cells and primary human neuronal progenitors (phNPCs) circumvent a major challenge facing our understanding of human brain function by providing us with access to living tissue representing different human nervous system cell types, developmental stages and diseases. We have developed both 2D and 3D culture systems for modeling human brain development, disease risk variants and synaptic maturation *in vitro* [8, 12, 13, 20], which provide an unprecedented opportunity to experimentally validate the predicted regulatory relationships. We established a high-throughput quantitative framework to compare differentiation in culture to *in vivo* fetal development and demonstrated extensive overlap of our cultures to *in vivo* brain by standard immunocytochemical methods and comprehensive analysis of gene expression [8, 12, 13, 20]**. The clear matching of transcriptomic patterns in our human *in vitro* models to *in vivo* developmental trajectories and cell types provide confidence that these systems provide a valid platform to assess gene regulatory networks** [8, 12, 13, 20]**.** phNPCs generate neurons with stereotypical morphologies similar to what is observed *in vivo*, first forming bipolar migrating cells, followed by axonogenesis and increases in dendritic arborization and exhibit synaptic activity [8], which also permits more advanced cell biological and physiological characterization. We further validated remarkably similar expression of both classical neuroanatomical markers and transcriptomically defined regional markers between differentiated phNPCs and human fetal cortex cells [8]. Recently, we have also developed even more mature and synaptically active 3D cortical forebrain spheroids (hFS; [12]), and are implementing a highly advanced, novel 3D culture system to model ASD risk genes developed by our collaborator, Sergiu Pasca (see letter of collaboration). This includes all major component cell types that can be identified by scRNAseq [14] (see Fig 1 in Aim 1). In addition to single cell transcriptomic analyses, we have used methylation data to assess the epigenetic maturation (epigenetic clock)[21], which demonstrates clear maturation of our 3D cultures. We will use the 2D cultures based on phNPCs for first line, higher throughput validation, followed by the 3D hFS [12, 14], which also provide more mature cultures, where later acting, postnatal putative cortical enhancers can be tested.

**Plan.** Choosing from validated enhancers from Aim 3, we will synthesize 100 validated enhancers with polymorphisms predicted to affect function between alleles. These will be transfected into engineered phNPC cells (see below) and assayed using an automated microfluidic culture system and associated integrated platform that we have developed for dynamic stimulation, cell manipulation, and time-lapse microscopy (Fig 5). This system allows multi-mode cell culture (single cell, 2-D monolayer and in 3-D organoids) and dynamic stimulation across 1,500 individually addressable cell culture units for high-throughput quantitative studies on mammalian cells (Fig 5A). Each of the 1,500 culture chambers can be programmed to receive a different set of reagents (Fig 5B). Coupled with custom software for chip control and computational data processing, the system can perform programmed delivery of thousands of formulated fluids to any designate on-chip culture unit, while monitoring and analyzing corresponding cellular responses via live cell microscopy. We have thus far used this system to investigate dynamic signaling in the differentiation of Neural Stem Cells (NSCs). Our experiments using primary embryonic (NSCs) and neuronal organoids demonstrated that NSCs proliferation, differentiation and lineage programming can be efficiently assessed at the single cell level via tracking the expression level of self-renewal (Hes5) and differentiation (Dcx) markers in response to dynamic growth factor inputs (Fig 5 C&D). Using this microfluidics system, we will assess the functional differences between alleles for each enhancer tested in our GFP reporter constructs transfected into phNPC cells. This will give us an unprecedented opportunity to test the output of a large number of neuronal enhancers during development and differentiation. This also permits delivery of patterning molecules (e.g. wnt, shh, bmp, smad inhibitors, and RA) to assess the relationships between these enhancers and the signals involved in regionalization and maturation.

**4c. Validation of target gene expression effects.** While reporter assays are powerful for testing sufficiency of an enhancer sequence to drive expression via a minimal promoter, we also wish to test the ability of a subset of 10-15 enhancer sequences to affect nearby gene expression. For this purpose we have engineered fluoresecent reporter phNPC-based lines of specific progenitor classes or postmitotic neuron classes using the CRISPR/Cas9 technologythat can be used to screen for quantitative changes in cell fate or class, as well as purify specific cell lineages for profiling. **We will use CRISPR/Cas9-mediated transcriptional activation (dCas9-VP64) and enhancer deletion (e.g [17]) to validate its activity, as we have previously shown for long range distal brain enhancers (as identified using Hi-C data) using the same system that we propose to use here [17].** We have developed VP64 transcriptional activation to validate a human accelerated region that we predicted was a distal enhancer for Gli-3, a forebrain patterning gene (Fig 6), supporting the feasibility of this sub-aim.



**Figure 6: CRISPR/Cas9-mediated transcriptional activation of a human accelerated enhancer to functionally validate a target gene. (A) Hi-C** Interaction map of a HAR/enhancer that is predicted to interact with *GLI3*. **(B)** Targeted binding sites for two guide RNAs (gRNAs). This HAR is located in a predicted active enhancer (H3K27ac) in fetal brain. **(C)** Targeting dCas9-VP64 to the HAR results in a 30-40% increase in the expression level of *GLI3* in phNPCs differentiated for 4 weeks. Two sets of gRNAs targeting different regions were cloned into EF1a-dCas9-VP64-2A-GFP-sgRNA. An empty vector without any inserted gRNA was used as control. Viruses were generated by co-transfection of CRISPR vectors with pVSVg and psPAX2 in HEK293 cells. Primary human neural progenitor cells (phNPC) were infected with viruses (empty vectors, gRNA1, gRNA2) on the day of split and differentiated as previously described. After 2.5 weeks of differentiation, cells that are infected (GFP+) were sorted by FACS. *GLI3* expression levels were measured by qRT-PCR (LightCycler 480 SYBR Green I Master, Roche).



**Figure 5.** Microfluidic culture system for high-throughput, dynamical analysis of neuronal cell models. **(A)** This microfluidic system performs automated cell culture processes such as cell seeding, stimulation with growth factors, time-lapse imaging and cell tracking, and cell retrieval. An on-chip multiplexer measures several fluids containing signaling molecules or drugs, and mixes them at predetermined ratios, creating complex chemical inputs. A peristaltic pump delivers these inputs to 1,500 independent cell culture chambers for dynamical cell stimulation. Each of the 1,500 chambers can be programmed to receive a different chemical stimulus. The system automatically tracks individual cells, 2-D populations or organoids via time-lapse microscopy. Cells can be immune-stained during or at the end of the experiments, and image processing reveals protein expression and morphology information at the single cell level, allowing quantitative analysis. **(B)** Schematic drawings (top row) and optical images (bottom row) of three distinct flow modes. (1) Fluid is directed to flow over the culture chamber directly (cell loading and retrieval mode); (2) Fluid is guided through the buffering region from the side (stimulation mode); (3) Fluid can be directed to bypass the chamber unit to avoid cross-contamination or perform other fluid manipulation. **(C)** Bright field images of neuronal stem cells (NSCs) cultured on chip in media containing (from left to right) 1000 ng/ml CXCL, 50 ng/mL PDGF and 100 nM PACAP. The scale bars are 50 μm in all images. **(D)** Immunostaining images of NPCs exposed to (top row to bottom) 50 ng/mL PDGF, 1000 ng/ml CXCL and 100 nM PACAP. Markers used for determining NSCs differentiation states are Hes5, Sox10, Pax6 and beta-tubulin. Insertions in the top row are selected NSCs cells with distinct morphology.

**4d. Enhancer single cell Drop-seq for mapping enhancer activity to specific neuronal lineages.** We have optimized plasmid delivery via electroporation using the Nucleofector II device (Lonza) to achieve ~70% efficiency with minimal toxicity. 10 to 15 isolated engineered reporter phNPCs lines will be mixed with our standard non-recombined cultures and differentiated for 2 to 10 weeks to model major stages of cortical neurogenesis. We will harvest cells for single cell isolation, performing each experiment in quadruplicate, and pool cells to obtain 6000 for Drop-seq (which we show in Aim 2; Fig 1 is sufficient to identify and profile all major cell classes *in vitro* and *in vivo* in developing brain) to assess changes in cell fate or changes in transcription due to enhancer activation in specific cell classes sorted based on reporter activity. We can identify cell classes derived from each distinct progenitor type and infer lineages by analyzing cells generated at each time point. Here, because we have cell class definitions from *in vivo* developing brain (e.g., Aims 1 and 2), we can perform a supervised analysis. We will quantify the diversity of cells generated in each progenitor class and characterize the lineages to those classes, as well as any enhancer activities that do not affect cell proliferation or fate (cell type composition), but that affect other cellular processes by single cell profiling.

**Pitfalls.** 2D cultures may not recapitulate all of the cell types and regulatory events, as well as our 3D systems. In addition, for enhancers that are acting post-natally, the 3D hFS system may be preferred due to its ability to achieve a more mature state matching post-natal development. So as an alternative, we will use 3D cultures to test a subset of those enhancers that passed through screens in Aim 2, but do not show activity in Aim 3, even though the more time-consuming culturing of 3D FS is limiting relative to the microfluidic and 2D methods. These 3D hFS can be transfected using the methods described above. We can also engraft progenitor reporter lines into the germinative layers of 3D cultures, and use immunocytochemistry or FISH for markers of the identified descendants of those progenitors and sort them by the reporter expression (following culturing for transcriptomic analysis). Given its high throughput for assessment of enhancer activity and effects on cell type transcriptomes, we will rely primarily on transcriptomics. Neuronal morphology, location, and connectivity are intimately related with function over a wide set of inhibitory and excitatory neurons. In future work, we aim to characterize the morphology, localization, and connectivity that is regulated by enhancers in molecularly defined cell classes.

**Elements Unique to This Site: The** University of Chicago (White, Tay) and UCLA (Geschwind) together provide the entire experimental component of this proposal, including the biological models and samples, single cell sequencing, genome scale enhancer validation (STARR-seq), CRISPR engineered enhancer characterization (using mutation and targeted VP64), and microfluidic-based single cell and organoid-based quantitative analyses of enhancer reporter assays. Specifically, for **Aim 1 and 2,** the Geschwind Lab (UCLA) will provide Hi-C and ATAC-seq maps from fetal human tissue to aid in assigning distal regulatory elements to genes and derive single cell gene expression profiles from DroNc-Seq on frozen human brain from controls and ASD across 3 broad developmental periods for building control and disease-relevant regulatory networks at the single cell level. For **Aim 3,** the White Lab will work closely with the Geschwind Lab to instantiate their well-validated phNPC protocols into the genome-wide STARR-seq and CapStarr-seq enhancer assays. For **Aim 4,** the Geschwind Lab will work closely with Tay and White Labs to perform functional validation experiments using well-characterized *in vitro* systems, hpNPC and hFS, and Crispr/CAS9-mediated deletion or CAS9-VP64-mediated activation of enhancers. They will leverage engineered reporter lines carrying lineage markers, as well as Drop-seq to characterize the functional outcomes of enhancer activation or repression. Finally, The University of Chicago will also issue a small subcontract to cover the salaries of a postdoc and Dr. Liu (at The University of Illinois, Chicago). Dr. Liu is a participant in the broader PsychENCODE Consortium and is an expert in brain eQTL analysis, and will therefore participate in **Aim 2**.

1. Boyle, A.P., et al., *Comparative analysis of regulatory information and circuits across distant species.* Nature, 2014. **512**(7515): p. 453-6.

2. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

3. Kittler, R., et al., *A comprehensive nuclear receptor network for breast cancer cells.* Cell Rep, 2013. **3**(2): p. 538-51.

4. Negre, N., et al., *A cis-regulatory map of the Drosophila genome.* Nature, 2011. **471**(7339): p. 527-31.

5. Psych, E.C., et al., *The PsychENCODE project.* Nat Neurosci, 2015. **18**(12): p. 1707-12.

6. Arnold, C.D., et al., *Genome-wide quantitative enhancer activity maps identified by STARR-seq.* Science, 2013. **339**(6123): p. 1074-7.

7. Vanhille, L., et al., *High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq.* Nat Commun, 2015. **6**: p. 6905.

8. Stein, J.L., et al., *A quantitative framework to evaluate modeling of cortical development by neural stem cells.* Neuron, 2014. **83**(1): p. 69-86.

9. Konopka, G., et al., *Modeling the functional genomics of autism using human neurons.* Mol Psychiatry, 2012. **17**(2): p. 202-14.

10. Rosen, E.Y., et al., *Functional genomic analyses identify pathways dysregulated by progranulin deficiency, implicating Wnt signaling.* Neuron, 2011. **71**(6): p. 1030-42.

11. Wexler, E.M., et al., *Genome-wide analysis of a Wnt1-regulated transcriptional network implicates neurodegenerative pathways.* Sci Signal, 2011. **4**(193): p. ra65.

12. Pasca, A.M., et al., *Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture.* Nat Methods, 2015. **12**(7): p. 671-8.

13. Pasca, S.P., et al., *Using iPSC-derived neurons to uncover cellular phenotypes associated with Timothy syndrome.* Nat Med, 2011. **17**(12): p. 1657-62.

14. Birey, F., et al., *Assembly of functionally integrated human forebrain spheroids.* Nature, 2017. **545**(7652): p. 54-59.

15. Rozowsky, J., et al., *PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls.* Nat Biotechnol, 2009. **27**(1): p. 66-75.

16. Harmanci, A., J. Rozowsky, and M. Gerstein, *MUSIC: identification of enriched regions in ChIP-Seq experiments using a mappability-corrected multiscale signal processing framework.* Genome Biol, 2014. **15**(10): p. 474.

17. Won, H., et al., *Chromosome conformation elucidates regulatory relationships in developing human brain.* Nature, 2016. **538**(7626): p. 523-527.

18. Mali, P., et al., *RNA-guided human genome engineering via Cas9.* Science, 2013. **339**(6121): p. 823-6.

19. Zhang, C., et al., *Universal Microfluidic System for Analysis and Control of Cell Dynamics.* bioRxiv, 2017. [**http://www.biorxiv.org/content/early/2017/06/28/157057**](http://www.biorxiv.org/content/early/2017/06/28/157057).

20. Martinez, R.A., et al., *Genome engineering of isogenic human ES cells to model autism disorders.* Nucleic Acids Res, 2015. **43**(10): p. e65.

21. Horvath, S., *DNA methylation age of human tissues and cell types.* Genome Biol, 2013. **14**(10): p. R115.