In Situ Perturb-Seq of Transcriptomes and RNA Neural Recordings

Cipriano William Romero
B.S. Electrical Engineering
California Institute of Technology (2013)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of
Master of Science
in
Electrical Engineering and Computer Science
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2021

©2021 Cipriano William Romero. All rights reserved.
The author hereby grants MIT permission to reproduce and distribute paper and electronic copies of this thesis document in whole or in part.

Author ....
Department of Electrical Engineering and Computer Science
May 20, 2021

Certified by..
Edward S. Boyden
Y. Eva Tan Professor in Neurotechnology
Thesis Supervisor

Accepted by..
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

In Situ Perturb-Seq of Transcriptomes and RNA Neural Recordings

by

Cipriano William Romero

Submitted to the Department of Electrical Engineering and Computer Science on May 20, 2021 in partial fulfillment of the requirements for the degree of Master of Science in Electrical Engineering and Computer Science

Abstract

In this work, we explore the intersection of in situ sequencing, neural recording, and CRISPR screens. An intracellular technology is outlined for encoding neural activity in the form of RNA, theoretically enabling single-cell resolution recording of whole-brain activity. This neural recording system can be coupled with perturb-seq in order to observe high-throughput genetic perturbations of neurons with both temporal and transcriptomic information. Untargeted expansion sequencing (ExSeq) can be used to generate a high-resolution spatiotemporal dataset that includes single guide RNAs (sgRNAs), neural activity, and transcriptomics. Targeted ExSeq, with the inclusion of no-gap padlock probes and SplintR ligase, can be applied to enhance the detection of sgRNA barcodes and targeted transcripts. In vitro and in vivo experimental pipelines are proposed for the fusion of these technologies, in this theoretical thesis.

Thesis Supervisor: Edward S. Boyden

Title: Y. Eva Tan Professor in Neurotechnology

Acknowledgments

First and foremost, I would like to thank Ed Boyden for his meaningful mentorship and support. His profound passion for science and engineering inspires the Synthetic Neurobiology Group (SynthNeuro) to think outside of the box and reimagine what is possible. SynthNeuro is an incredible community that strives to create paradigm shifts in the world of neurotechnology and beyond.

A special thanks to Anu Sinha, Fei Chen, Keith Tyo, Paul Reginato, and Sam Rodriques for brainstorming ideas with me and providing valuable feedback.

I would like to express my gratitude towards Hock Tan, Lisa Yang, and Y. Eva Tan for their significant generosity and philanthropy.

Leslie Kolodziejski and Janet Fischer have helped me navigate my EECS graduate school experience, and I am grateful for their guidance.

Finally, I would like to express appreciation towards my family and friends for their support and encouragement.

Table of Contents

1. Introduction......5

2. Perturb-Seq......6

3. RNA Neural Recordings................,......7

   3.1 System Overview......7

   3.2 In Vitro Development and Optimization......10

   3.3 In Vivo Development and Optimization......12

4. Expansion Sequencing......15

   4.1 Combining Ex Situ NGS with ExSeq for Neural Recordings......16

5. Untargeted Multiplexed ExSeq......16

6. Perturb-Seq with Targeted ExSeq......18

7. Experimental Workflows......20

   7.1 In Vitro Screening Strategies......20

   7.2 In Vivo Screening Strategies......22

8. Conclusion......23

List of Figures......26

References......27

1 Introduction

Ever since Jennifer Doudna and Emmanuelle Charpentier's landmark 2012 paper on CRISPR technology[1], the scientific community has entered a renaissance of genetic editing possibilities[2]. Since its introduction, CRISPR has quickly superseded previously developed genome editing methods, including transcription activator-like effector nucleases (TALENs) and zinc finger nucleases (ZFNs)[3]. Applications have ranged from permanently correcting disease mutations in mice[4], to enhancing oil and meal quality of pennycress plants[5]. In short, CRISPR has revolutionized biotechnology.

The human genome is far from being fully characterized and understood[6,7], and CRISPR screens are accelerating this field of study through high-throughput interrogations of the genome[8,9]. Researchers have demonstrated knockout, activation, and inhibition screens with CRISPR-Cas9[10,11]. There is a multitude of applications, including drug development[12], cellular pathway investigation[13], and disease research[14]. Significant advancements in the DNA sequencing industry have led to gradual cost reductions[15], thereby increasing accessibility to CRISPR screens.

The transcriptome reflects a snapshot of the transcripts in a given cell, and technologies such as microarrays and RNA-Seq have been able to capture ex situ transcriptomics[16]. Although ex situ methods are high-throughput, they fail to provide a spatial context[16–18]. In situ methods preserve spatial information, and they have revealed insights in brain connectomics[18], point mutations in breast cancer[19], and more[20].

Arguably, the most challenging organ to research is the brain. A mouse brain has approximately 7.5×10^7 neurons, action potentials on the millisecond scale, and considerable absorption and scattering of electromagnetic waves[21]. There have been many efforts to record brain activity with electrodes[22],

fMRI[23], fluorescent probes[24], and other technologies. Currently, there is not a recording technology that has shown whole-brain activity of a mammal at single-cell resolution[21].

2 Perturb-Seq

Perturb-seq is a high-throughput CRISPR screening technology that enables pooled transductions of sgRNAs with single-cell RNA-seq readouts[25]. A pooled sgRNA library is synthesized and cloned into a plasmid with a sorting feature (antibiotic resistance or fluorescence). After cloning, the pooled plasmids are packaged into viruses. Target cells are then transduced with the pooled viruses. Cell selection then occurs (via antibiotics or FACS), and the remaining cells are used for single-cell RNA-Seq[25–27].

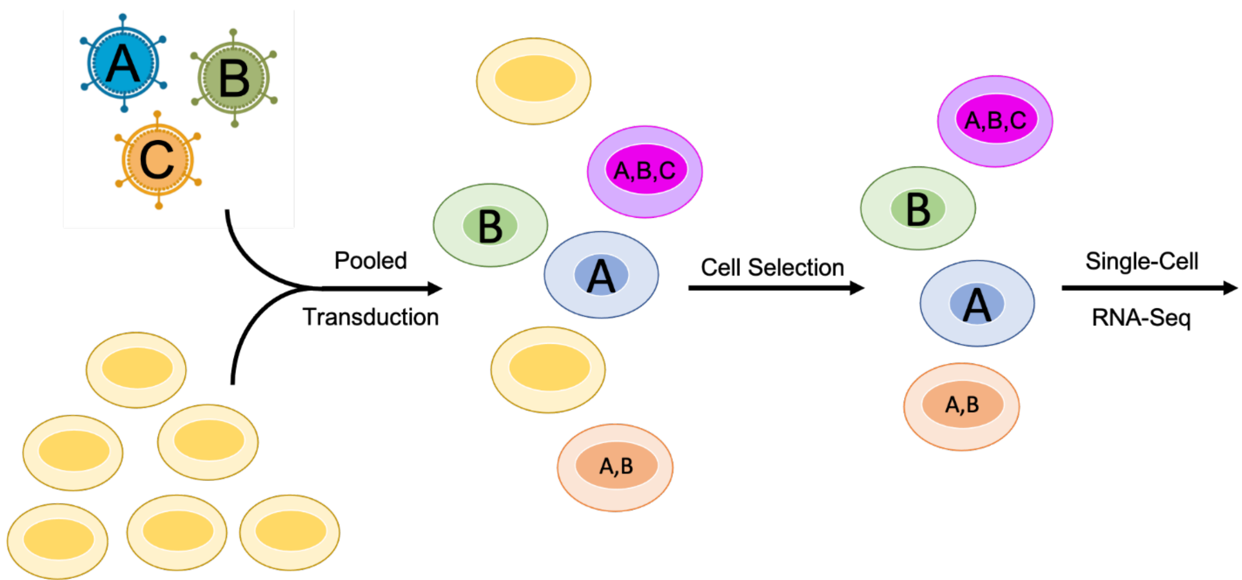

Figure 1: Ex situ perturb-seq workflow. Cas9-expressing cells (yellow) are infected with pooled viruses that contain sgRNAs (A, B, or C). Consequently, cells receive one sgRNA (A, B, or C), multiple sgRNAs, or no sgRNAs. Cells are then selected (via puromycin, FACS, etc.) to remove cells that did not receive any sgRNAs. Finally, single-cell RNA-seq is performed to study the pooled perturbations.

There are different methods for detecting which sgRNA(s) are present in a given cell. Guide barcode (GBC) sequences can be placed in the perturb-seq vector, giving each transduced cell a sgRNA along with its corresponding GBC. Working with the single-cell RNA-seq data, researchers can use detected GBC sequences to match sequenced transcriptomes with gene perturbations[25]. However, this method faces the challenge of sgRNA-GBC uncoupling; lentiviral transduction of co-packaged sgRNAs and GBCs can result in mismatches between sgRNAs and cells[26]. Another approach is CROP-seq. With the CROP-seq vector, both polyadenylated (non-editing) and editing sgRNA sequences are formed in order to enable direct detection of sgRNAs via single-cell RNA-seq[28]. Perturb-seq has successfully been demonstrated in vitro[25] as well as in vivo[29].

3 RNA Neural Recordings

3.1 System Overview

A new approach to overcoming the challenge of large-scale single-cell neural recordings would be to encode neural activity in the form of nucleotides and leverage the latest sequencing technologies to recover the extensive spatiotemporal dataset[21]. In vitro molecular recording of stimuli in the form of DNA has been demonstrated to potentially function within the temporal scale of minutes[30]. Additionally, a method that records the age of RNAs via an adenosine editing strategy has been shown to work with a timescale of hours[31]. Here, we will describe a proposed RNA-based neural recording technology that detects intracellular calcium influxes.

This proposed system uses an RNA template with MS2 sites and fused proteins to primarily record action potentials as uridines and resting states as adenines. The MS2 coat proteins (MCPs) stochastically bind to MS2 binding sites[32] on the RNA template. These MCPs are fused to either poly-A polymerase or

M13 peptides[33]. In a resting state with intracellular low calcium levels, these poly-A polymerases polyadenylate the RNA recording, thereby marking the passage of time in a resting state. After an action potential-induced calcium influx[34], calmodulin (CaM) that is fused to terminal U transferase (TuT)[35] binds to M13 peptides, just like CaM and M13 bind in GCaMP systems[36]. TuTs that are bound to the CaM-M13s proceed to polyuridylate the RNA recording. After intracellular calcium levels return to a resting state, CaM unbinds from M13s, and only polyadenylation continues to occur. Although there is simultaneous stochastic polyuridylation and polyadenylation during high calcium events, decoding algorithms can be developed in order to temporally stich the RNA recordings together.

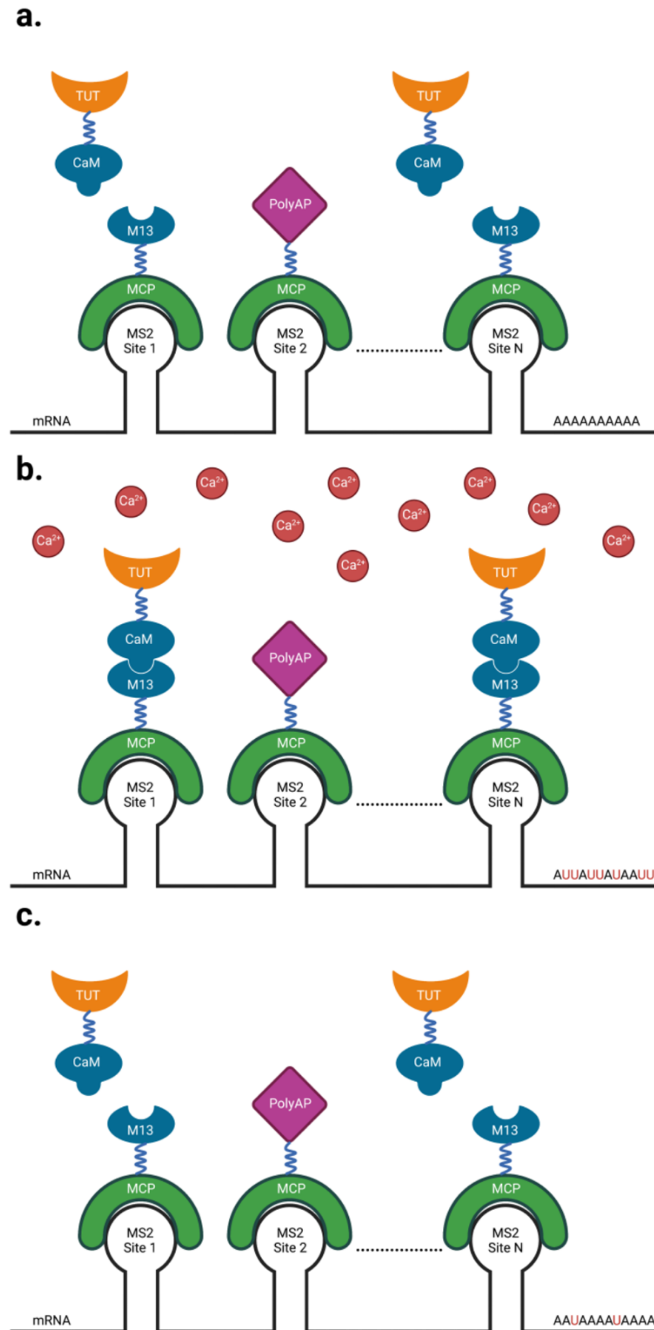

Figure 2: An RNA-based neural recording system. In the absence of relatively high calcium levels, the poly-A polymerase-fused MCPs and M13-fused MCPs are stochastically bound to MS2 sites on an mRNA. Polyadenlyation occurs due to the poly-A polymerases (a). Following an action potential-induced intracellular calcium influx, calcium ions bind to CaM. This promotes the formation of a CaM-M13 complex, which is fused to both TuT and a fraction of the MCPs on MS2 sites. Consequently, polyuridylation occurs in addition to polyadenylation (b). After intracellular calcium levels return to equilibrium, CaM unbinds from M13. Polyadenlyation continues (c).

3.2 In Vitro Development and Optimization

After successfully demonstrating calcium sensing, we will want to improve the temporal resolution and run length of our RNA recordings. A possible enhancement will be increasing polymerase speeds; faster polyuridylation and polyadenylation will increase the temporal resolution. However, since polyadenylation is calcium-independent, it will be crucial to have the polyuridylation speed be relatively close to the polyadenylation speed for decoding purposes. Another key improvement will be having longer well-preserved RNA neural recordings, which will enable longer recording time-scales. Additionally, having the optimal concentrations for template RNAs and polymerase complexes will give us a large number of recording sequences to facilitate decoding, optimize diffusion of system components, and minimize interference with neuronal functionality. Having sufficient, well-diffused component concentrations (template RNAs, poly-A polymerases, poly-U polymerases, and poly-U binding protein complexes) at different intracellular locations will be crucial for in situ studies.

We will perform high-throughput screening of both wildtype and mutagenized[37,38] polymerases fused to calcium sensors for optimizing polymerase speeds, run lengths, recording accuracies, and local component concentrations. There will be additional screens that vary the linker lengths (between polymerases, fluorescent proteins, and binding proteins) and linker compositions to improve protein folding and functionality. By increasing the number of MS2 binding sites, we can increase the probability of adenylation and uridylation (uridylation only occurs with high Ca^{2+}) within a given time window, as seen in Figure 3. We should observe nonlinear scaling of polyadenylation and polyuridylation with the number of MS2 binding sites that eventually saturates (Figure 2). Consequently, the temporal resolution should increase with number of MS2 binding sites. We will also explore different RNA preservation techniques: immediate inactivation of RNAses, flash-freezing cells at -80°C[39], etc.

There are several aforementioned variables to optimize, and there will be pros and cons for each optimized variable. We will create several generalized protocols for users that enable them to select among optimizations for longer run lengths, time scales, etc. Depending on the specifics of the desired recording (time length, time-resolution, recording accuracy, etc.), we will need to create customized options for users.

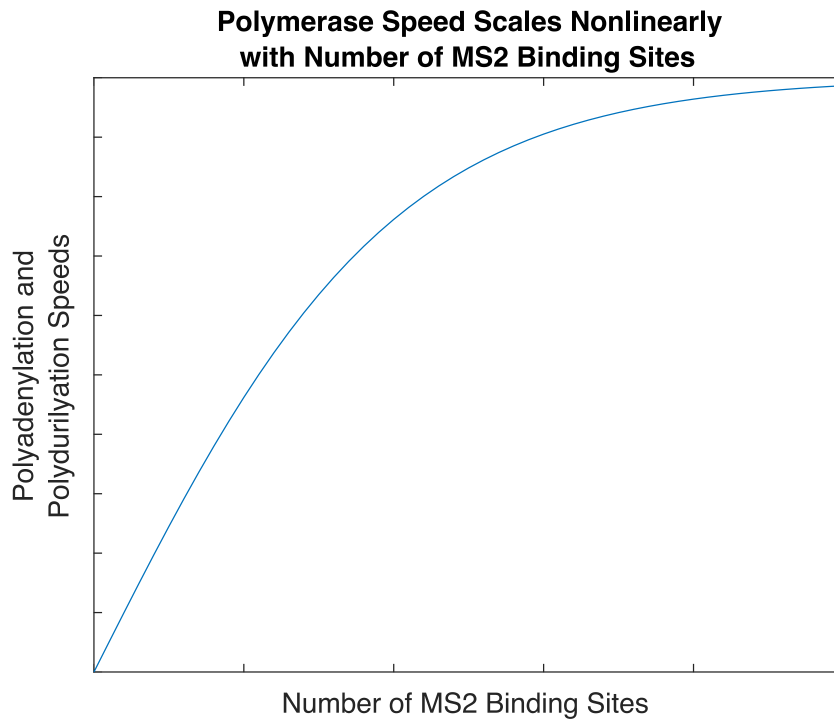

Figure 3: Expected trend for polyadenylation and polyuridylation speeds from varying the number of MS2 binding sites on template mRNAs. Adding more MS2 binding sites increases the number of polymerases that can become bound to the template mRNA, thereby increasing the frequency of neural recording. This speed is eventually expected to saturate as more and more MS2 sites are present.

3.3 In Vivo Development and Optimization

We want to demonstrate that RNA neural recording is a revolutionary tool that can be applied to a wide variety of organisms. It will be important to characterize how RNA recording systems function in organisms with unique neural properties. For example, graded potentials play a relatively large role in C. elegans[40]. Also, we will need to develop protocols for organism euthanasia and data recovery. These protocols will be critical for enabling a decoding "stopwatch" that allows the user to have a reference time for when the recordings occurred.

First, we will virally transfect our RNA recording systems to targeted neuron(s) in vivo. We then have three options: stimulating neuron(s) connected to or containing RNA recording systems with reliable stimulation methods (optogenetics[41], electrodes, etc.), passively allowing the organism time to have natural activation of well-characterized neuronal pathways, or actively causing the organism to perform behaviors with well-characterized neuronal pathways. As soon as the desired recordings are completed, the organism must be euthanized immediately. The time of euthanasia will serve as a reference time for the RNA recordings, so immediate euthanasia is critical. Next, the RNA can be extracted from the transfected cells (via lysis or other methods) and then sequenced with RNA-seq. We will also use expansion microscopy to perform in situ RNA sequencing in order to study connectomics and observe neural activity at different intracellular locations. Finally, we will use decoding algorithms to reconstruct intracellular Ca^{2+} concentrations vs. time, as seen in Figure 4. These algorithms will work backwards in time (starting with the time of euthanasia) to stitch together RNA recordings, and they will account for polyadenylation and polyuridylation speeds. The reconstructed data will be compared with the expected results from stimulation or neural activity in well-characterized pathways to verify the RNA

recording's in vivo performance. There will be comparative studies to determine how well different RNA recording systems work in different organisms and brain regions.

When we deal with in vivo studies of the brain, we must account for the challenge of dealing with highly varied spatiotemporal scales. We will start with simple neurons and neural pathways, and then we will gradually record larger, more complex systems. Depending on the number of recorded neurons and RNA recording concentrations, there could be nontrivial amounts of data to process. Also, different organisms have varying neuronal properties: NTP concentrations[42], ability to translate transfected genes, etc. These different properties will require specific, optimized recording systems to be deployed to each organism. Additionally, the algorithms will have to self-adapt or be modified to fit each organism.

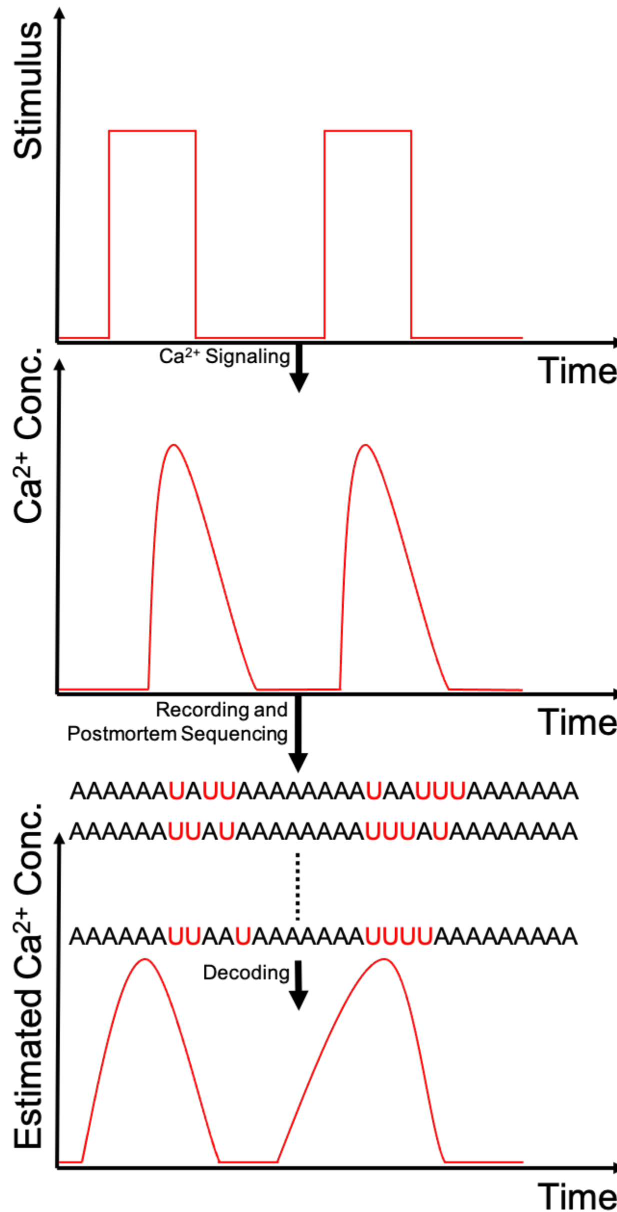

Figure 4: An example workflow of in vivo verification experiment for RNA neural recordings. Stimuli (from optogenetics, electrodes, etc.) are applied to neurons, resulting in calcium signaling. The RNA neural recording system creates temporal sequences of adenine without high calcium levels. With high calcium levels, temporal sequences are extended with both adenine and uridine. After the recordings are sequenced (via ExSeq or other methods), decoding algorithms can be used to reconstruct neural activity.

4 Expansion Sequencing

Fluorescent in situ sequencing (FISSEQ) is an in situ sequencing method that can be applied to mRNAs in order to keep cells intact and preserve their spatial information[43]. After chemically fixing the tissue, in situ reverse transcription (RT) is applied to the mRNA, and the remaining mRNA is degraded. Next, the cDNA fragments from RT are circularized, and the cDNA is modified with primary amines and then crosslinked. Afterwards, the cDNAs are amplified with rolling circle amplification (RCA). The amplicon products of RCA are then cross-linked to form ball-like clusters of replica DNA. Sequencing by Oligonucleotide Ligation and Detection (SOLiD) to sequence the amplicons, and the sequence locations are also recorded[17,20,44].

Expansion microscopy (ExM) is a super-resolution imaging tool that enables the user to overcome the diffraction limits of traditional microscopy[45]. There is an application of ExM for FISSEQ that is called expansion sequencing (ExSeq)[18]. ExSeq is used to facilitate enzyme transport via tissue expansion, as well as enhance the imaging resolution for sequencing. ExSeq was developed to chemically fix tissue, then covalently link in situ mRNA to hydrophilic monomers. The monomers are then cross-linked to form an acrylate/acrylamide gel. This is followed by a proteolysis treatment that degrades proteins and homogenizes the sample's mechanical properties. The sample is then placed in water, and the gel isotropically expands along with the mRNA. Next, the RT, cDNA circularization, and RCA steps from FISSEQ are applied to create well-spaced, highly-dense cDNA amplicons. Custom software is then used to sequence the amplicons, and the spatial coordinates of the sequences are stored.

4.1 Combining Ex Situ NGS with ExSeq for Recordings

One approach that we will pursue uses reverse transcription with next-generation sequencing (NGS)[46] in order to obtain the temporal information that is stored on the 3' mRNA ends, and we will use ExSeq for retrieving the spatial information. This can be accomplished with thinly sliced tissue (e.g. 5μm thick) that has been transfected with neural recording systems that contain cellular barcodes in their mRNA templates. Alternating serial tissue slices will be used for ExSeq and NGS; one slice will be used for NGS, the next slice will be used for ExSeq, the following slice will be used for NGS, etc. By combining the NGS temporal data from the 3' template ends with the ExSeq-determined corresponding cellular locations, we will create a powerful spatiotemporal recording system. Additionally, we will read out the endogenous, non-recording transcripts in order to determine the cellular identity and state of each recording site.

5 Untargeted Multiplexed ExSeq

Here, we describe how we can combine all three of the main technologies of this document (perturb-seq, RNA neural recordings, and ExSeq) to obtain an unprecedented spatiotemporal dataset that reveals how pooled CRISPR perturbations affect neural activity and transcriptomics. By transducing Cas9-expressing cells with perturb-seq vectors (promoting the expression of GBC or direct sgRNA barcodes) and RNA neural recording systems, we can create pooled (single or multi) gene perturbations and record their impacts on neural functioning and the transcriptome.

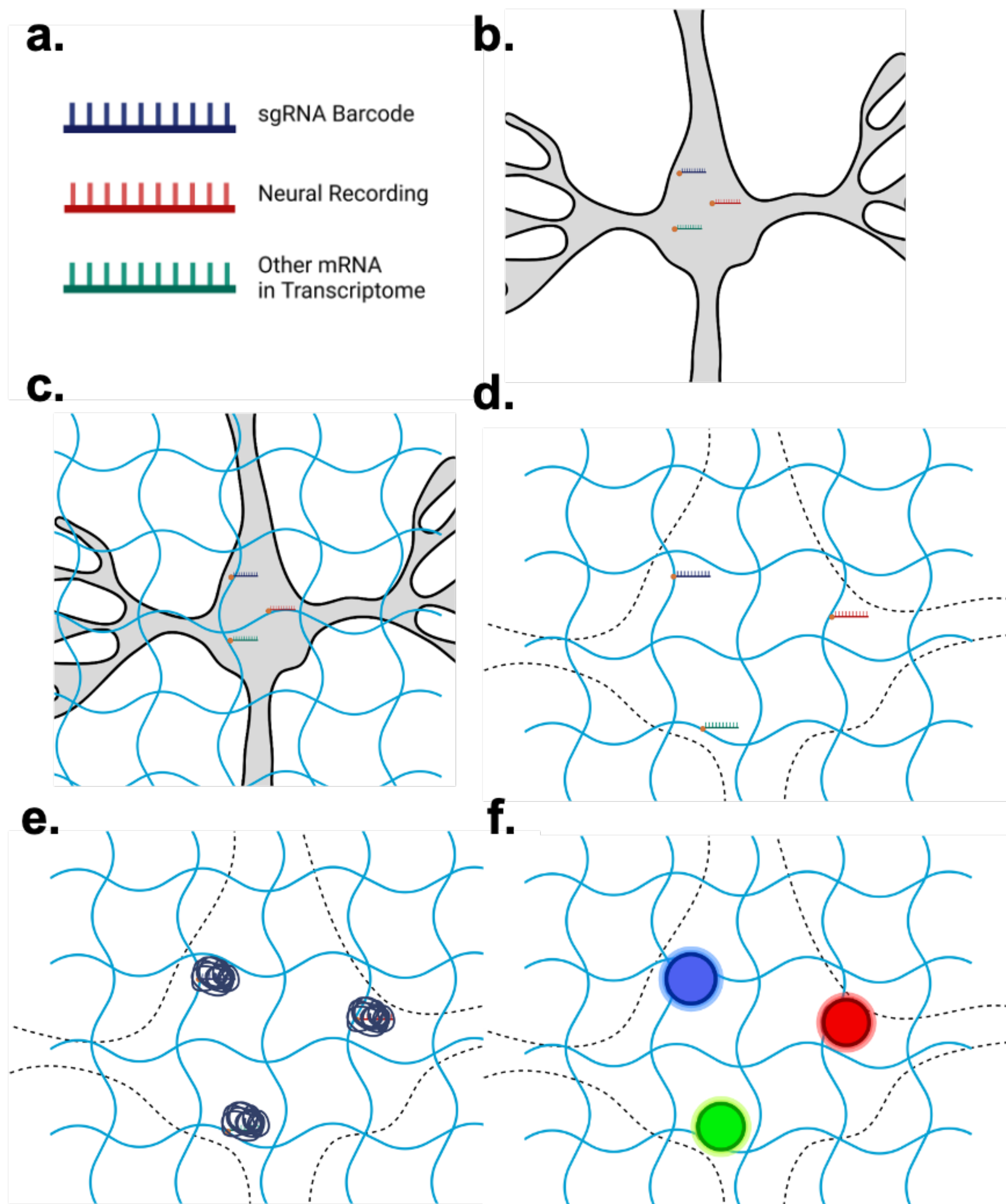

Figure 5: Untargeted ExSeq of sgRNAs, neural recordings, and other mRNAs (a) in the transcriptome. The tissue is fixed and RNA is anchored with reagent LabelX (b). After fixation, the tissue is embedded into a gel that contains hydrophilic monomers (c). Water is added after digestion, causing the gel network to isotropically expand (d). FISSEQ is used to create cDNA amplicons (e). Sequencing chemistry and imaging is performed to sequence the sgRNAs, neural recordings, and other mRNAs in situ (f).

6 Perturb-Seq with Targeted ExSeq

Due to the stochastic nature of the proposed RNA neural recording system, it is not feasible to design efficient padlock probes for neural recordings. However, targeted ExSeq can still be applied to sgRNA barcodes (GBC or sgRNA sequences) as well as transcripts of interest. Researchers have done targeted in situ sequencing of sgRNAs with gap-filling padlock probes[47]. However, they only included optical phenotype and sgRNA readouts without additional sequencing of the transcriptome. Additionally, their protocol required a reverse transcription step before padlock gap-filling[47].

Padlock gap-filling can be challenging; double-stranded DNA ligase can create over-extensions or insufficient extensions of padlock probes during gap-filling[48]. No-gap padlock probes offer significantly higher sensitivity with approximately 30% RNA detection[49]. Although no-gap padlock probes do not store the sequence of the targeted RNA section (that could have been gap-filled), they are ideal for perturb-seq because the sgRNA library is predefined. Synthesizing an sgRNA library along with a corresponding no-gap padlock probe library would offer higher sensitivity than the gap-filling approach. Additionally, a library of non-sgRNA transcripts and its corresponding no-gap padlock probe library could be synthesized to observe selected fractions of the transcriptome with high sensitivity.

Reverse transcription of sgRNAs and other transcripts into cDNA is only necessary for double-stranded DNA ligase[48]. Reverse transcription can theoretically be skipped for no-gap padlock probes with SplintR Ligase. SplintR Ligase has the ability to ligate single-stranded DNA that is splinted by complementary RNA[50], so it theoretically can circularize no-gap padlock probes.

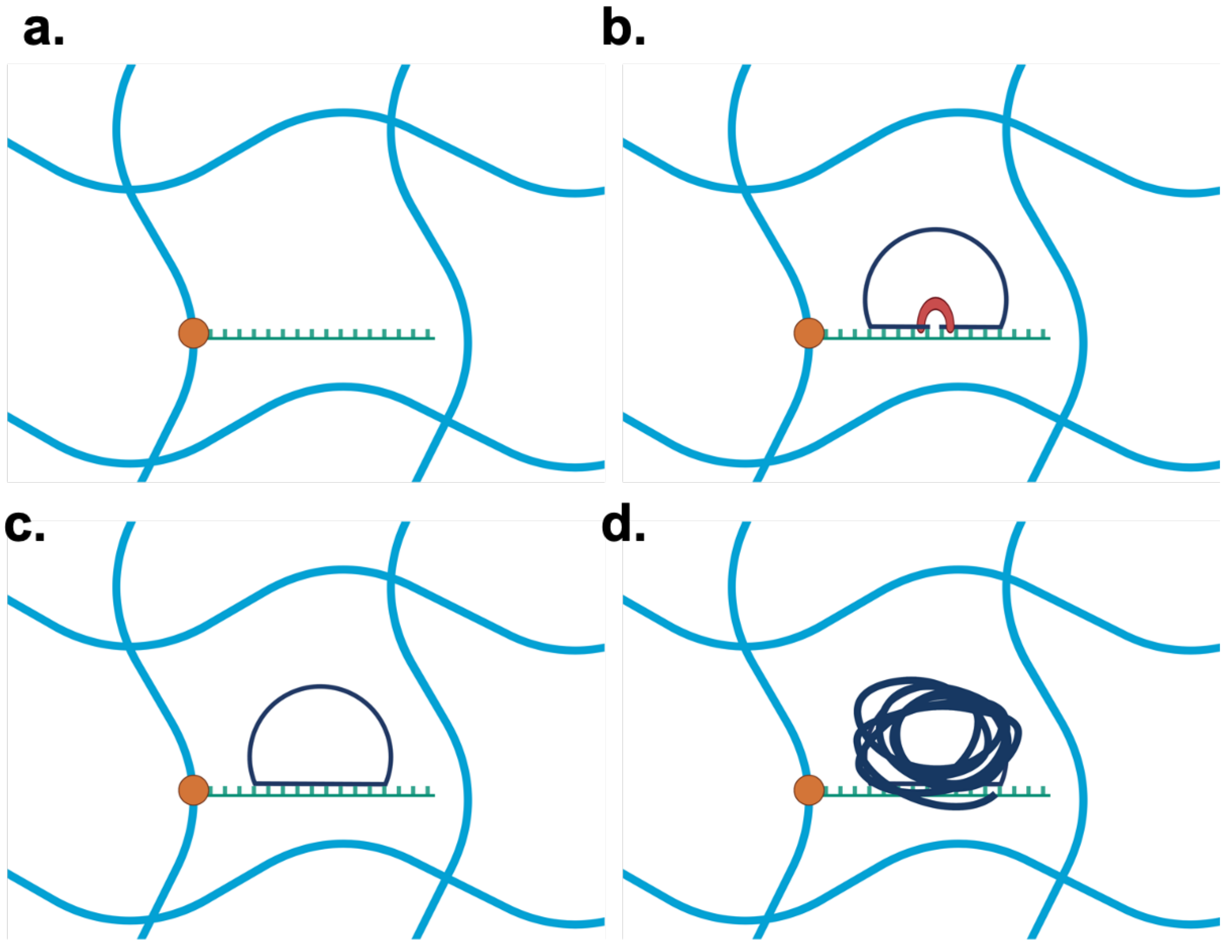

Figure 6: Targeted ExSeq of sgRNAs and other mRNAs in the trancriptome via a no-gap padlock probe approach. RNA that is anchored the gel of expanded tissue. A no-padlock probe hybridizes to the mRNA, and SplintR Ligase (red) is applied (b). SplintR Ligase circularizes the DNA padlock probe that is hybridized to mRNA (c). FISSEQ is applied to create cDNA amplicons of the padlock probe (d).

7 Experimental Workflows

7.1 In Vitro Screening Strategies

After all of these in situ tools have been successfully integrated and demonstrated, researchers can use them to perform high-throughput CRISPR screens in vitro for a variety of studies. One area of interest is screening for gene perturbations that shift in situ transcriptomics of disease model cells towards transcriptomics of apparently healthy cells. Disease model cells could include immortalized cell lines, iPSC-derived cell lines, and others[51]. Alternatively, diseased cell cultures could directly come from primary cell donors. Pooled perturbations of both diseased and healthy cells with in situ sequencing outputs is a powerful assay for identifying key perturbations that could counteract or exacerbate disease phenotypes.

Researchers can select genes of interest, as well as the type of perturbation (knockout, activation, inhibition, etc.). Pooled sgRNA libraries can then be designed to maximize chances of successful Cas9-targeting for each gene, as well as suitable controls (intergenic sgRNAs, non-targeting sgRNAs, etc.)[52,53]. Depending on gene selection and cell type, it may be advisable to have multiple sgRNAs per gene in order to avoid misinterpreting an unsuccessful sgRNA. Although whole-genome perturb-seq is possible, the currently high costs of single-cell RNA-seq[54] prohibit many labs from pursuing whole-genome CRISPR screens with single-cell RNA-seq data. TAP-seq is an option for conducting large-scale sgRNA library experiments by targeting a subset of the transcriptome for sequencing[55].

Following sgRNA library cloning into perturb-seq vectors, vectors can be packaged in lentiviruses and titering can be adjusted to target a multiplicity of infection (MOI). Relatively high MOI experiments result in a higher fraction of cells with multi-gene perturbations, whereas relatively low MOI

experiments result in a higher fraction of cells with single-gene perturbations[25]. After pooled lentiviral transduction and cell selection have occurred, targeted ExSeq[18] can be applied with no-gap padlock probes in order to observe the resulting changes to the transcriptome in situ with enhanced sensitivity.

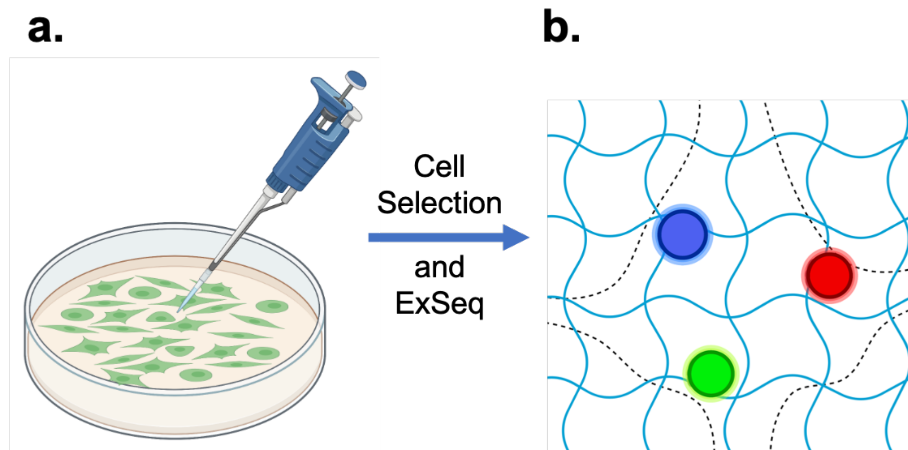

Figure 7: High-throughput in vitro CRISPR screens with in situ sequencing. Cas9 cell cultures can be infected with pooled viruses that contain sgRNAs and neural recording systems (a). ExSeq can then be applied to sequence the sgRNAs, neural recordings, and other transcripts in situ (b).

The in situ dataset can then be used to draw comparisons between perturbed cells in both healthy and diseased states, as well as unperturbed controls. There are opportunities to develop new software pipelines to analyze this data, especially with machine learning[56]. The analysis and conclusions from a round of perturb-seq experiments can inform future experiments. For example, researchers may determine that they want higher MOIs with fewer gene targets in subsequent studies. That way, they can investigate only the genes that showed statistical promise and explore combinatorics without prohibitive cost increases.

7.2 In Vivo Screening Strategies

Although in vitro screens are higher-throughput than in vivo work, they do not enable researchers to study perturbations at the organism-level[29]. In vivo work is especially important for investigating neuroscience, where complex connectomics is essential to understanding the brain. The human brain is estimated to contain 58.9 x 10^{12} synapses in the frontal cortex alone[57]. Depending on the application, in vitro perturb-seq can be used to identify gene candidates for lower-throughput in vivo perturb-seq experiments.

In vivo perturb-seq is where the combination of RNA neural recording and ExSeq technologies can have great impact. For example, an experiment pipeline for studying disease in mice is outlined in Figure 8. Healthy mice in addition to mouse models for brain disease[58,59] can be used for this assay. Pooled transduction of lentiviral perturb-seq vectors along with RNA recording systems can be applied to a brain region that is relevant to the disease. After mice recover from this surgery, they can participate in a behavioral experiment that is used to demonstrate disease phenotypes (e.g. the Morris water maze test for Alzheimer's disease model mice[60]). Following a behavioral experiment, the mouse is immediately euthanized, and relevant brain tissue slices are prepared. Untargeted ExSeq[18] is then applied to gather spatiotemporal data. Just like the previously described in vitro pipeline, in vivo results can also inform future experiments. Different sgRNA libraries and combinatorics can be explored. New behavioral experiments can be incorporated. ExSeq can be applied to other brain regions.

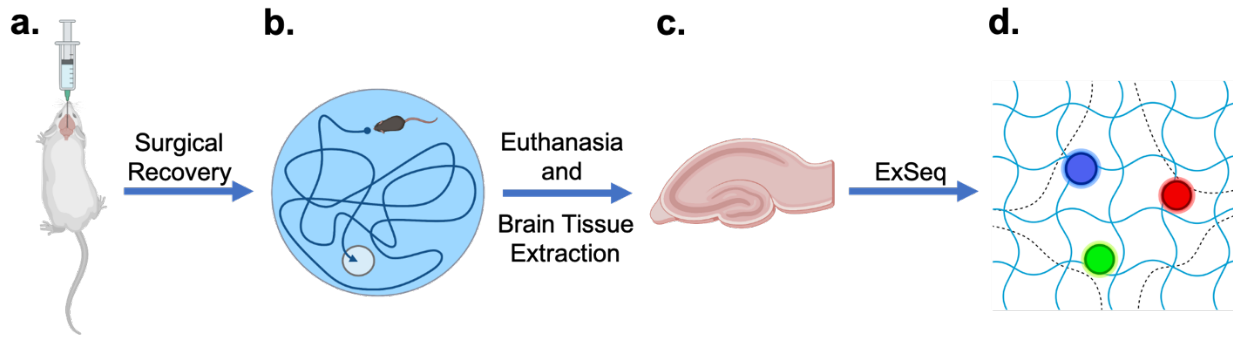

Figure 8: High-throughput in vivo CRISPR screens with in situ sequencing. Transgenic Cas9 animal brains can be transduced with pooled viruses that contain sgRNAs and neural recording systems. If the animals are not transgenic with Cas9, then Cas9 vectors can also be virally transduced as well (a). Animals with the pooled sgRNAs and neural recording systems, as well as control groups, can perform behavioral experiments (b). Immediately after the behavioral activity, the animals are euthanized. Tissue slices are recovered postmortem (c). ExSeq can then be applied to sequence the sgRNAs, neural recordings, and other transcripts in situ (d).

8 Conclusion

The intersection of CRISPR screens, in situ sequencing, and molecular neural recording could revolutionize neurotechnology in the coming years. Although this work focusses on these three technologies, it could have also been expanded to include others (e.g. optogenetics[41]). The incredible complexity of biological systems, ranging from the human genome to the brain, requires the use of integrative tools to fully capture and understand biological phenomena. Technological advancements, both incremental and monumental, always build upon each other, and neurotechnology is no exception. As more and more tools are developed for neuroscience, it is critical for neuroscientists to identify applications that can leverage the integration of these powerful methods.

Although many current CRISPR screening methods rely on ex situ data collection, in situ perturb-seq will offer an unprecedented look at the perturbed transcriptome. Although ex situ single-cell sequencing data can provide valuable insights, it treats the transcriptome as being spatially

homogenous, which it clearly is not[18]. By coupling ex situ and in situ sequencing technologies (e.g. alternating tissue slices), then we can benefit from the relatively long read-lengths of ex situ approaches, as well as the high spatial resolutions of tissue-expanding in situ methods. This fusion could lead to accelerated drug development, novel gene therapies, basic science discoveries, and more.

RNA neural recording is currently a theoretical approach, but it is important to develop due to its ability to provide single-cell whole-brain recordings. Additionally, its application of being multiplexed with perturbed transcriptomics makes it a powerful option for future in vivo perturb-seq experiments. Although current efforts towards molecular neural recording have only been able to demonstrate timescales of minutes to hours[30,31], these timescales could be improved with additional protein-engineering, RNA template design, and decoding methodology. Due to the proposed euthanasia requirement for sequencing the recordings, this technology will not be available for storing in vivo neural activity of humans. However, this technology could prove to be highly impactful for future animal studies.

Although perturb-seq is a revolutionary screening method, the currently high costs of single-cell RNA-seq are preventing faster dissemination and widespread usage. Currently, researchers are faced with experimental optimization and trade-offs for a fixed budget. If perturb-seq is performed with more sgRNAs, then larger sequencing depths will be necessary to adequately process the data. The larger sequencing costs could prevent the researcher from investigating more cell cultures (for in vitro experiments), with the assumption of a fixed budget. Fortunately, as the costs of sequencing continue to decrease, the technology will become more financially accessible to researchers. In the future, it will be feasible for many labs to conduct whole-genome perturb-seq experiments with extensive combinatorics. Additionally, the rapid development of machine learning approaches to genetics[56] will help these labs extract deeper insights from experimental results.

At the time of writing, it has been less than a decade[1] since the enormous potential of CRISPR was discovered. The rapid development of CRISPR enzymes[61,62], screening methods[28,29,47,63], sgRNA design tools[53,64], and more are paving the way for profound gene perturbation experiments. The integration of the CRISPR revolution with in situ sequencing, molecular neural recording, and other tools could dramatically alter the field of neuroscience. These technological advancements will eventually result in the improvement of human health, as well as a greater understanding of biology.

List of Figures

Figure 1. Ex situ perturb-seq workflow……………………..……………………………………………………….…………..6

Figure 2. RNA-based neural recording system….……….……………………………………………………………….9

Figure 3. Number of MS2 sites alters temporal resolution..…………..…………………………………………11

Figure 4. In vivo RNA recording validation experiment workflow..………..…………….………………….14

Figure 5. Untargeted multiplexed ExSeq workflow…………….……………………………………………………17

Figure 6. Targeted ExSeq with no-gap padlock probe..……….…………………………………………………19

Figure 7. Integrative in vitro experiments.……………………..……………..……………………………………21

Figure 8. Integrative in vivo experiments……….……………..………………………………………………………23

Bibliography

1. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, (2012).

2. Knott, G. J.; Doudna, J. A. CRISPR-Cas guides the future of genetic engineering. *Science* **361**, (2018).

3. Barrangou, R. Cas9 targeting and the CRISPR revolution. *Science* **344**, (2014).

4. Yin, H.; Xue, W.; Chen, S.; Bogorad, R. L.; Benedetti, E.; Grompe, M.; Koteliansky, V.; Sharp, P. A.; Jacks, T.; Anderson, D. G. Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nature Biotechnology* **32**, (2014).

5. Gao, C. The future of CRISPR technologies in agriculture. *Nature Reviews Molecular Cell Biology* **19**, (2018).

6. Collins, F. S.; Morgan, M.; Patrinos, A. The Human Genome Project: Lessons from large-scale biology. *Science* **300**, (2003).

7. Galperin, M. Y.; Koonin, E. V. From complete genome sequence to “complete” understanding? *Trends in Biotechnology* **28**, (2010).

8. Bodapati, S.; Daley, T. P.; Lin, X.; Zou, J.; Qi, L. S. A benchmark of algorithms for the analysis of pooled CRISPR screens. *Genome Biology* **21**, (2020).

9. Lin, X.; Chemparathy, A.; La Russa, M.; Daley, T.; Qi, L. S. Computational Methods for Analysis of Large-Scale CRISPR Screens. *Annual Review of Biomedical Data Science* **3**, (2020).

10. Joung, J.; Konermann, S.; Gootenberg, J. S.; Abudayyeh, O. O.; Platt, R. J.; Brigham, M. D.; Sanjana, N. E.; Zhang, F. Genome-scale CRISPR-Cas9 knockout and transcriptional activation screening. *Nature Protocols* **12**, (2017).

11. Kampmann, M. CRISPRi and CRISPRa Screens in Mammalian Cells for Precision Biology and Medicine. *ACS Chemical Biology* **13**, (2018).

12. Shalem, O.; Sanjana, N. E.; Hartenian, E.; Shi, X.; Scott, D. A.; Mikkelsen, T. S.; Heckl, D.; Ebert, B. L.; Root, D. E.; Doench, J. G.; Zhang, F. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, (2014).

13. Birsoy, K.; Wang, T.; Chen, W. W.; Freinkman, E.; Abu-Remaileh, M.; Sabatini, D. M. An Essential Role of the Mitochondrial Electron Transport Chain in Cell Proliferation Is to Enable Aspartate Synthesis. *Cell* **162**, (2015).

14. Chen, S.; Sanjana, N. E.; Zheng, K.; Shalem, O.; Lee, K.; Shi, X.; Scott, D. A.; Song, J.; Pan, J. Q.; Weissleder, R.; Lee, H.; Zhang, F.; Sharp, P. A. Genome-wide CRISPR screen in a mouse model of

tumor growth and metastasis. *Cell* **160**, (2015).

15. Shendure, J.; Balasubramanian, S.; Church, G. M.; Gilbert, W.; Rogers, J.; Schloss, J. A.; Waterston, R. H. DNA sequencing at 40: Past, present and future. *Nature* **550**, (2017).

16. Lowe, R.; Shirley, N.; Bleackley, M.; Dolan, S.; Shafee, T. Transcriptomics technologies. *PLoS Computational Biology* **13**, (2017).

17. Lee, J. H.; Daugharthy, E. R.; Scheiman, J.; Kalhor, R.; Ferrante, T. C.; Terry, R.; Turczyk, B. M.; Yang, J. L.; Lee, H. S.; Aach, J.; Zhang, K.; Church, G. M. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nature protocols* **10**, 442–458 (2015). [PMID:25675209]

18. Alon, S.; Goodwin, D. R.; Sinha, A.; Wassie, A. T.; Chen, F.; Daugharthy, E. R.; Bando, Y.; Kajita, A.; Xue, A. G.; Marrett, K.; Prior, R.; Cui, Y.; Payne, A. C.; Yao, C. C.; Suk, H. J.; Wang, R.; Yu, C. C.; Tillberg, P.; Reginato, P.; Pak, N.; Liu, S.; Punthambaker, S.; Iyer, E. P. R.; Kohman, R. E.; Miller, J. A.; Lein, E. S.; Lako, A.; Cullen, N.; Rodig, S.; Helvie, K.; Abravanel, D. L.; Wagle, N.; Johnson, B. E.; Klughammer, J.; Slyper, M.; Waldman, J.; Jané-Valbuena, J.; Rozenblatt-Rosen, O.; Regev, A.; Church, G. M.; Marblestone, A. H.; Boyden, E. S. Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. *Science* **371**, (2021).

19. Ke, R.; Mignardi, M.; Pacureanu, A.; Svedlund, J.; Botling, J.; Wählby, C.; Nilsson, M. In situ sequencing for RNA analysis in preserved tissue and cells. *Nature Methods* **10**, (2013).

20. Lee, J. H.; Daugharthy, E. R.; Scheiman, J.; Kalhor, R.; Yang, J. L.; Ferrante, T. C.; Terry, R.; Jeanty, S. S. F.; Li, C.; Amamoto, R.; Peters, D. T.; Turczyk, B. M.; Marblestone, A. H.; Inverso, S. A.; Bernard, A.; Mali, P.; Rios, X.; Aach, J.; Church, G. M. Highly Multiplexed Subcellular RNA Sequencing in Situ. *Science* **343**, 1360–1363 (2014). [PMID:24578530]

21. Marblestone, A. H.; Zamft, B. M.; Maguire, Y. G.; Shapiro, M. G.; Cybulski, T. R.; Glaser, J. I.; Amodei, D.; Benjamin Stranges, P.; Kalhor, R.; Dalrymple, D. A.; Seo, D.; Alon, E.; Maharbiz, M. M.; Carmena, J. M.; Rabaey, J. M.; Boyden, E. S.; Church, G. M.; Kording, K. P. Physical principles for scalable neural recording. *Frontiers in Computational Neuroscience* No. OCT (2013).

22. Stevenson, I. H.; Kording, K. P. In *Nature Neuroscience*; 2011; Vol. 14

23. Shapiro, M. G.; Westmeyer, G. G.; Romero, P. A.; Szablowski, J. O.; Küster, B.; Shah, A.; Otey, C. R.; Langer, R.; Arnold, F. H.; Jasanoff, A. Directed evolution of a magnetic resonance imaging contrast agent for noninvasive imaging of dopamine. *Nature Biotechnology* **28**, (2010).

24. Barnett, L.; Platisa, J.; Popovic, M.; Pieribone, V. A.; Hughes, T. A Fluorescent, Genetically-Encoded Voltage Probe Capable of Resolving Action Potentials. *PLoS ONE* **7**, (2012).

25. Dixit, A.; Parnas, O.; Li, B.; Chen, J.; Fulco, C. P.; Jerby-Arnon, L.; Marjanovic, N. D.; Dionne, D.; Burks, T.; Raychowdhury, R.; Adamson, B.; Norman, T. M.; Lander, E. S.; Weissman, J. S.; Friedman, N.; Regev, A. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853–1866 (2016).

26. Replogle, J. M.; Xu, A.; Norman, T. M.; Meer, E. J.; Terry, J. M.; Riordan, D.; Srinivas, N.; Mikkelsen, T. S.; Weissman, J. S.; Adamson, B. *bioRxiv*. 2018

27. Adamson, B.; Norman, T. M.; Jost, M.; Weissman, J. S. *bioRxiv*. 2018

28. Datlinger, P.; Rendeiro, A. F.; Schmidl, C.; Krausgruber, T.; Traxler, P.; Klughammer, J.; Schuster, L. C.; Kuchler, A.; Alpar, D.; Bock, C. Pooled CRISPR screening with single-cell transcriptome readout. *Nature Methods* **14**, (2017).

29. Jin, X.; Simmons, S. K.; Guo, A.; Shetty, A. S.; Ko, M.; Nguyen, L.; Jokhi, V.; Robinson, E.; Oyler, P.; Curry, N.; Deangeli, G.; Lodato, S.; Levin, J. Z.; Regev, A.; Zhang, F.; Arlotta, P. In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science* **370**, (2020).

30. Bhan, N. J.; Strutz, J.; Glaser, J.; Kalhor, R.; Boyden, E.; Church, G.; Kording, K.; Tyo, K. E. J. *bioRxiv*. 2019

31. Rodriques, S. G.; Chen, L. M.; Liu, S.; Zhong, E. D.; Scherrer, J. R.; Boyden, E. S.; Chen, F. *bioRxiv*. 2020

32. Tutucci, E.; Vera, M.; Biswas, J.; Garcia, J.; Parker, R.; Singer, R. H. An improved MS2 system for accurate reporting of the mRNA life cycle. *Nature Methods* **15**, (2018).

33. Yagi, K.; Yazawa, M.; Ikura, M.; Hikichi, K. Interaction between calmodulin and target proteins. *Advances in experimental medicine and biology* **255**, (1989).

34. Clapham, D. E. Calcium Signaling Review. *Cell* **80**, 259–268 (1995).

35. Aphasizhev, R.; Sbicego, S.; Peris, M.; Jang, S. H.; Aphasizheva, I.; Simpson, A. M.; Rivlin, A.; Simpson, L. Trypanosome mitochondrial 3′ terminal uridylyl transferase (TUTase): The key enzyme in U-insertion/deletion RNA editing. *Cell* **108**, (2002).

36. Tian, L.; Hires, S. A.; Mao, T.; Huber, D.; Chiappe, M. E.; Chalasani, S. H.; Petreanu, L.; Akerboom, J.; McKinney, S. A.; Schreiter, E. R.; Bargmann, C. I.; Jayaraman, V.; Svoboda, K.; Looger, L. L. Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nature Methods* **6**, (2009).

37. Arnold, F. H. Protein engineering for unusual environments. *Current Opinion in Biotechnology* **4**, (1993).

38. Liu, Q.; Xun, G.; Feng, Y. The state-of-the-art strategies of protein engineering for enzyme stabilization. *Biotechnology Advances* **37**, (2019).

39. Milani, P.; Escalante-Chong, R.; Shelley, B. C.; Patel-Murray, N. L.; Xin, X.; Adam, M.; Mandefro, B.; Sareen, D.; Svendsen, C. N.; Fraenkel, E. Cell freezing protocol suitable for ATAC-Seq on motor neurons derived from human induced pluripotent stem cells. *Scientific Reports* **6**, (2016).

40. Liu, Q.; Hollopeter, G.; Jorgensen, E. M. Graded synaptic transmission at the Caenorhabditis

elegans neuromuscular junction. *Proceedings of the National Academy of Sciences of the United States of America* **106**, (2009).

41. Boyden, E. S. Optogenetics and the future of neuroscience. *Nature Neuroscience* **18**, (2015).

42. Nedialkov, Y. A.; Gong, X. Q.; Hovde, S. L.; Yamaguchi, Y.; Handa, H.; Geiger, J. H.; Yan, H.; Burton, Z. F. NTP-driven translocation by human RNA polymerase II. *Journal of Biological Chemistry* **278**, (2003).

43. Tuma, R. S.; Beaudet, M. P.; Jin, X.; Jones, L. J.; Cheung, C. Y.; Yue, S.; Singer, V. L. Characterization of SYBR Gold nucleic acid gel stain: a dye optimized for use with 300-nm ultraviolet transilluminators. *Analytical biochemistry* **268**, 278–288 (1999). [PMID:10075818]

44. Avital, G.; Hashimshony, T.; Yanai, I. Seeing is believing: new methods for in situ single-cell transcriptomics. *Genome Biology* **15**, 110 (2014). [PMID:25000927]

45. Chen, F.; Tillberg, P. W.; Boyden, E. S. Expansion microscopy. *Science* **347**, 543–548 (2015). [PMID:25592419]

46. Mardis, E. R. Next-Generation DNA Sequencing Methods. *Annual Review of Genomics and Human Genetics* **9**, 387–402 (2008). [PMID:18576944]

47. Feldman, D.; Singh, A.; Schmid-Burgk, J. L.; Carlson, R. J.; Mezger, A.; Garrity, A. J.; Zhang, F.; Blainey, P. C. Optical Pooled Screens in Human Cells. *Cell* **179**, (2019).

48. Chen, X.; Sun, Y. C.; Church, G. M.; Lee, J. H.; Zador, A. M. Efficient in situ barcode sequencing using padlock probe-based BaristaSeq. *Nucleic Acids Research* **46**, (2018).

49. Larsson, C.; Grundberg, I.; Söderberg, O.; Nilsson, M. In situ detection and genotyping of individual mRNA molecules. *Nature Methods* **7**, (2010).

50. Jin, J.; Vaud, S.; Zhelkovsky, A. M.; Posfai, J.; McReynolds, L. A. Sensitive and specific miRNA detection method using SplintR Ligase. *Nucleic Acids Research* **44**, (2016).

51. Slanzi, A.; Iannoto, G.; Rossi, B.; Zenaro, E.; Constantin, G. In vitro Models of Neurodegenerative Diseases. *Frontiers in Cell and Developmental Biology* **8**, (2020).

52. Hanna, R. E.; Doench, J. G. Design and analysis of CRISPR–Cas experiments. *Nature Biotechnology* **38**, (2020).

53. Brazelton, V. A.; Zarecor, S.; Wright, D. A.; Wang, Y.; Liu, J.; Chen, K.; Yang, B.; Lawrence-Dill, C. J. A quick guide to CRISPR sgRNA design tools. *GM crops & food* **6**, (2015).

54. Ding, J.; Adiconis, X.; Simmons, S. K.; Kowalczyk, M. S.; Hession, C. C.; Marjanovic, N. D.; Hughes, T. K.; Wadsworth, M. H.; Burks, T.; Nguyen, L. T.; Kwon, J. Y. H.; Barak, B.; Ge, W.; Kedaigle, A. J.; Carroll, S.; Li, S.; Hacohen, N.; Rozenblatt-Rosen, O.; Shalek, A. K.; Villani, A. C.; Regev, A.; Levin, J. Z. Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. *Nature*

Biotechnology **38**, (2020).

55. Schraivogel, D.; Gschwind, A. R.; Milbank, J. H.; Leonce, D. R.; Jakob, P.; Mathur, L.; Korbel, J. O.; Merten, C. A.; Velten, L.; Steinmetz, L. M. Targeted Perturb-seq enables genome-scale genetic screens in single cells. *Nature Methods* **17**, (2020).

56. Libbrecht, M. W.; Noble, W. S. Machine learning applications in genetics and genomics. *Nature Reviews Genetics* **16**, (2015).

57. Tang, Y.; Nyengaard, J. R.; De Groot, D. M. G.; Gundersen, H. J. G. Total regional and global number of synapses in the human brain neocortex. *Synapse* **41**, (2001).

58. Jackson-Lewis, V.; Przedborski, S. Protocol for the MPTP mouse model of Parkinson's disease. *Nature Protocols* **2**, (2007).

59. Jankowsky, J. L.; Zheng, H. Practical considerations for choosing a mouse model of Alzheimer's disease. *Molecular Neurodegeneration* **12**, (2017).

60. Bromley-Brits, K.; Deng, Y.; Song, W. Morris Water Maze test for learning and memory deficits in Alzheimer's disease model mice. *Journal of Visualized Experiments* No. 53 (2011).

61. Chatterjee, P.; Lee, J.; Nip, L.; Koseki, S. R. T.; Tysinger, E.; Sontheimer, E. J.; Jacobson, J. M.; Jakimo, N. A Cas9 with PAM recognition for adenine dinucleotides. *Nature Communications* **11**, (2020).

62. Ma, D.; Xu, Z.; Zhang, Z.; Chen, X.; Zeng, X.; Zhang, Y.; Deng, T.; Ren, M.; Sun, Z.; Jiang, R.; Xie, Z. Engineer chimeric Cas9 to expand PAM recognition based on evolutionary information. *Nature Communications* **10**, (2019).

63. Adamson, B.; Norman, T. M.; Jost, M.; Cho, M. Y.; Nuñez, J. K.; Chen, Y.; Villalta, J. E.; Gilbert, L. A.; Horlbeck, M. A.; Hein, M. Y.; Pak, R. A.; Gray, A. N.; Gross, C. A.; Dixit, A.; Parnas, O.; Regev, A.; Weissman, J. S. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867–1882 (2016).

64. Doench, J. G.; Fusi, N.; Sullender, M.; Hegde, M.; Vaimberg, E. W.; Donovan, K. F.; Smith, I.; Tothova, Z.; Wilen, C.; Orchard, R.; Virgin, H. W.; Listgarten, J.; Root, D. E. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology* **34**, (2016).