Functional genomic and transcriptomic tools for spatial and dynamic phenotypes

by

Hong Anh Anna Le

B.S., Biological Chemistry, B.A., Chemistry

University of Chicago, 2017

Submitted to the Department of Biological Engineering in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological Engineering

at the

Massachusetts Institute of Technology

September 2023

Authored by: Hong Anh Anna Le
Department of Biological Engineering
July 27, 2023

Certified by: Paul C. Blainey
Associate Professor of Biological Engineering
Thesis Supervisor

Accepted by: Katharina Ribbeck
Andrew (1956) and Erna Viterbi Professor
Chair of Graduate Program, Department of Biological Engineering

Functional genomic and transcriptomic tools for spatial and dynamic phenotypes

by

Hong Anh Anna Le

Submitted to the Department of Biological Engineering on July 27, 2023, in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological Engineering

Abstract

Biology is driven by complex cellular processes that require precise regulation in time and in space. However, the genetic and molecular factors underlying these behaviors are difficult to study in their native contexts and, as a result, are often not well understood. Although next-generation sequencing and image-based methods have enabled high-throughput profiling of cell states, there is still a need for technologies that systematically probe and measure complex behaviors, including cell non-autonomous and dynamic phenotypes.

In this thesis, we present the development of functional genomic and synthetic biology tools to address this challenge. We first applied optical pooled screening to quantify cell-cell interactions in mixed cultures with primary neurons and reveal functional interaction partners of synaptogenic cell adhesion molecules. Using these screens, we identified differential modulators of excitatory and inhibitory synapse formation, implicating diverse cellular pathways in this process. To increase the throughput of these optical pooled screens, we also built a fluidics platform for automated *in situ* sequencing. Finally, we leveraged retroviral polyproteins to package cellular RNAs for non-destructive measurements, enabling longitudinal recording of transcriptional states in living cells. Together, this work establishes scalable tools to measure and understand spatial and dynamic cellular phenotypes.

Thesis Supervisor: Paul C. Blainey
Title: Associate Professor of Biological Engineering, Massachusetts Institute of Technology

Acknowledgments

My time in graduate school has been one of the most challenging but rewarding experiences of my life. I am immensely grateful to all my mentors, friends, and family who have supported me through this PhD journey.

First and foremost, I would like to thank my advisor, Paul Blainey. Your faith in me always pushed me to go after new ideas, no matter how challenging. Your rigor and kindness have made me a better scientist and collaborator, and your optimism and excitement, even for the smallest of victories, always motivated me through moments of self-doubt. You allowed me to be open and honest with you and uplifted me during tough times in ways that I needed most.

I am also indebted to my committee members, Doug Lauffenburger and Beth Stevens, who were always ready to engage in fruitful discussions, provided helpful feedback that pushed forward the direction of my work, and encouraged me at every step. Thomas Biederer has been a wonderful collaborator and mentor as well, without whom my thesis work would not be possible. You have shown me endless generosity with your time, always answering my questions and guiding me through next steps with such a positive attitude and enthusiasm that left me invigorated each time.

The Broad and MIT have been fantastic training grounds, filled with amazing scientists and all-around kindest staff that made it possible for me to do my research. I want to thank Karen Perez de Arce, who selflessly taught me everything I know about culturing neurons. I also had the privilege of working closely with two research associates, Tridib Biswas and Bryce Kirby – thank you for your hard work, creative ideas, and positive energy. You always made tackling even the hardest of challenges fun and exciting. I am also grateful for everyone I met during my time as a BE Communication Lab fellow – coaching you and with you has made me a better teacher and communicator.

I am incredibly thankful for everyone in the Blainey lab. I count myself lucky to have been able to work with such talented lab members who, despite being pulled in so many different directions, were always willing to go above and beyond to answer questions, offer advice, and occasionally take breaks outside of lab. To Team Lasagna, past and present, even though we all worked on our independent projects, I could always lean on you to get technical advice or commiserate over screening struggles. Thank you to Miguel Reyes, Avtar Singh, Josh Elacqua, Josh Peters, Luke Funk, and Jake Qiu, for the game nights, the potlucks, the hiking adventures, and everything in between. To Becca, thank you for sticking by my side through all the ups and downs; I could always talk to you about anything and everything, and I couldn’t imagine sharing this experience with anyone else. To Emily, our friendship is one of the best things the pandemic has given me; thank you for being such a caring, selfless friend, always. Finally, to Mo, I’m so very glad that my first mentor in this lab turned into a life-long friend. I continue to learn from you to this day, from your scientific rigor and creativity but, more importantly, from your kindness and generosity you show those around you.

Thank you to my friends, especially Aliya Moreira, Priyanka Mehta, Chuck Nguyen, and Brooke Huisman, for always being there for me, no matter how far apart or busy we are.

To my parents, thank you for all the sacrifices you have made for me to have these opportunities to follow my passions and be where I am today. Thank you for your boundless love and support, and your constant reminders for me to eat well.

Thank you to Hanh, my favorite (and only) sister. I will never be able to express how grateful I am to you for raising me since high school and being there for me at every life stage. To Sy, thank you for your curious questions, your dorky jokes, and your words of encouragement; I am so glad to have you as my big brother. To Sara and Liyana, I'm in awe of the amazing little girls you have turned out to be – watching you grow up has brought me so much joy and kept me grounded throughout the years.

Finally, to my husband, Peter – how lucky am I to have shared this journey with you. I could not have done it without you by my side, lending me reagents when I ran out, delivering snacks during my late nights in lab, and cheering me up after every failed experiment. You have been my best friend, my advocate, my safe place, my never-ending source of joy. Thank you for your love, patience, and support.

Table of Contents

List of Figures

List of Tables

List of Supplementary Figures

Chapter 1. Introduction

1.1 Advances in functional genomic approaches

1.1.1 Overview of forward genetic screening

The genome encodes a variety of RNA and protein products, whose expression levels, regulatory networks, and interactions are often cell state- or cell type-specific. Given this complexity, characterizing the functions of all genes remains a long-standing challenge in human biology. Forward genetics is one approach to define gene function by analyzing the phenotype that results from a specific induction of a change in the genotype. Historically, these mutagenesis studies were performed in model organisms, such as *C. elegans*, *Drosophila*, and yeast, as they provided convenient systems in which to introduce mutations and observe their consequences at the molecular, cellular, or even organismal level.

Technological advances, in both genome sequencing and editing, have enabled targeted manipulation of the human genome and transcriptome, and thus, functional genomic screening in human cells. In these genetic screens, many perturbations are applied across a constant genetic background to causally link genotypes with corresponding phenotypes at scale. The development of synthetic reagents for mammalian expression of small interfering RNAs (siRNAs) or short hairpin RNAs (shRNAs), which leverage RNA interference pathways to knock down endogenous transcripts, have facilitated early large-scale loss-of-function screens in human models[1–3]. However, these tools, though programmable, suffer from high off-target effects, limiting the interpretability of RNAi screen results.

More recently, CRISPR-based technologies have provided a more efficient and robust means of precisely modulating gene activity and are now commonly used in genetic screens. Most loss-of-function screens rely on single-guide RNAs (sgRNAs) to direct Cas9 or Cas12a endonucleases to specific loci, through recognition of a programmable 20-nucleotide spacer sequence complementary to a target genomic site. These ribonucleoprotein complexes then induce a double-strand break (DSB), which, upon repair by error-prone end joining pathways, results in stochastic insertion and deletion mutations that typically frameshift the targeted gene[4]. Alternatively, instead of inducing gene knockouts, CRISPR interference (CRISPRi) can repress gene expression by using a dead Cas9 (dCas9) without nuclease activity fused to effector proteins to mediate transcriptional silencing[5]. This approach is especially well-suited for screens investigating essential genes or non-coding elements, or using models that cannot tolerate the cytotoxicity caused by DSBs[6]. The versatility and efficacy of CRISPR-based tools have thus advanced the scale, feasibility, and complexity of functional genomic screens in human cells.

1.1.2 Multi-dimensional readouts for pooled genetic screens

In pooled screens, genetically encoded perturbations are introduced into a cell population as a pool, in contrast to arrayed screens, in which cells with each perturbation are maintained separately throughout the screening process. Pooled screens therefore are more easily scalable and allow for more robust comparisons between perturbations within a sample but require a method to associate genotypes with observed phenotypes. Approaches for demultiplexing the perturbations within a pooled cell library vary depending on the phenotypic readout of a screen.

In enrichment screens, cells are selected for a predefined phenotype of interest and subjected to next-generation sequencing (NGS) of perturbation barcodes to determine enrichment scores, based on their abundances in the pre- and post-selection populations (Figure 1.1). Enrichment screens have successfully identified genes that confer a cellular fitness effect or drug resistance[3,7,8]. They may also be applied to phenotypes that can be selected by the expression of fluorescent reporters or markers to allow cell isolation via fluorescence-activated cell sorting (FACS). These enrichment-based methods thus necessarily project complex cellular behaviors to a single-dimensional space for selection, although recent advances have sought to expand the types of phenotypes accessible to enrichment screening[9]. Nevertheless, because the perturbation barcodes are detected via bulk NGS, enrichment scores reflect an average of a cell population.

Thus, molecular profiling methods for pooled cell libraries have been developed to simultaneously acquire single-cell genotype identity and phenotype measurements, using single-cell sequencing (e.g. scRNA-seq or scATAC-seq) or mass spectrometry approaches (e.g. CyTOF)[10–16]. The captured phenotypes are high-dimensional, with thousands to tens of thousands of features extracted for each individual cell, but they focus primarily on the abundance of molecules based on sequencing or proteomic readouts. (Figure 1.1). More complex phenotypes, including cell morphology, localization of RNAs, proteins, and organelles, dynamic processes, and cell-cell interactions, therefore cannot be directly assayed using such profiling tools.

In contrast, image-based approaches can measure these phenotypes using techniques such as immunofluorescence (IF) staining and fluorescent *in situ* hybridization (FISH)[17–20]. In addition to fixed-cell imaging, live-cell methods, e.g. using fluorescent reporters, can capture changes in cell states and dynamic processes. In the past, microscopy-based assays have been widely used in arrayed screens addressing a range of biological questions. More recently, several groups have developed methods to detect perturbation barcodes *in situ* and perform pooled image-based screens[16,21–25]. In this thesis, we will focus on optical pooled screening, a method leveraging *in situ* sequencing by synthesis (SBS) to enable large-scale pooled screens of image-based phenotypes in human cells[23].

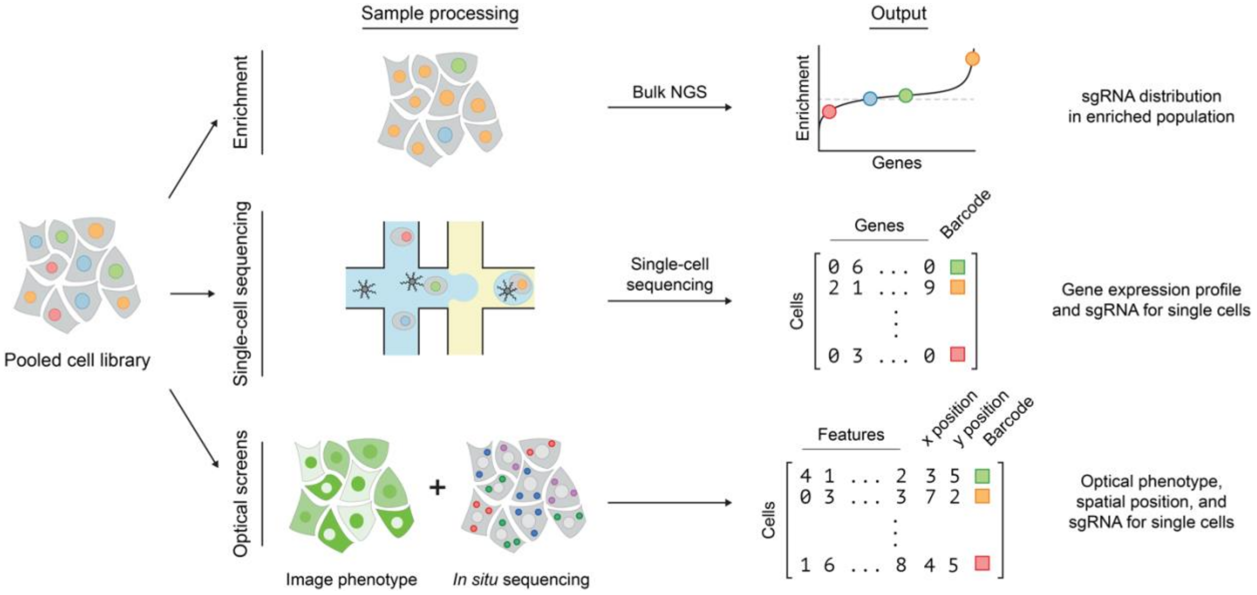


Figure 1.1 | Approaches for pooled genetic screening.

In pooled screening, a population of cells is subjected to a library of genetic perturbations, such as guide RNAs for CRISPR screens. Enrichment, single-cell profiling, and optical-based assays are three common approaches for phenotypic readout. Enrichment-based screens determine population-level changes in perturbation abundance by bulk next-generation sequencing (NGS) following an applied selection. Single-cell profiling and optical screens do not require an enrichment step and instead rely on information-rich phenotypic measurements. Single-cell assays pair perturbation barcodes to a cell phenotype at single-cell resolution, such as cell transcriptome for single-cell RNA sequencing-based screens. Through *in situ* sequencing, optical pooled screens pair image-based phenotypes with perturbation barcodes, also at single-cell resolution. Figure adapted from Feldman et al. 2022[26].

1.1.3 Optical pooled screening for complex phenotypes

Optical pooled screening technology couples image-based phenotypic assays with *in situ* sequencing to read out the genetic perturbation in each single cell[27]. Perturbation barcodes (e.g. sgRNA sequences expressed using the CROPseq vector) are detected via *in situ* amplification, involving reverse transcription, padlock-based gap filling, and rolling circle amplification, followed by iterative cycles of *in situ* SBS, thereby preserving spatial information of the genetic perturbation and phenotype (Figure 1.1). Optimizations to the detection efficiency and scalability of optical pooled screening have rendered the platform tractable for screening thousands of perturbations, up to genome-wide libraries, in tens of millions of cells to gain novel biological insights, including defining functions of essential genes and identifying regulators in immune signaling pathways[23,28,29].

Sample processing bottlenecks, however, still pose substantial limits to the throughput of this technology. Currently, *in situ* sequencing requires significant amounts of time and labor: completion of all cycles of SBS for a screen can take several weeks. As a result, optical pooled screening is not yet suitable for larger-scale screens, particularly in cell models less tractable than cancer cell lines. In Chapter 3, we developed a custom fluidics platform for automated *in situ* SBS, which increases the effective throughput of optical pooled screens by at least two-fold. Using our automation system, we can complete sequencing of a full 6-well plate in ~3 days with minimal hands-on time. This platform will lower the practical barrier to performing larger-scale and more complex optical pooled screens and accelerate the deployment of the technology to the broader research community.

Leveraging the ability of imaging techniques to measure cell-cell interactions, in Chapter 2, we demonstrated for the first time the application of optical pooled screening to capture cell non-autonomous phenotypes. We performed a CRISPR knockout screen of interaction partners of a synaptogenic cell adhesion molecule using a co-culture of non-neuronal cells and primary neurons, quantifying intercellular interactions in more than 1.8 million cells in an artificial synapse formation assay. This work established optical pooled screening as one of the few high-throughput methods compatible with screening for cell non-autonomous phenotypes and a highly generalizable strategy to address this challenge.

1.2 Mechanisms of synaptogenesis

1.2.1 Structure and function of neuronal synapses

The human nervous system consists of hundreds of billions of neurons organized into interconnected neural circuits. Neurons within these networks communicate using electrical and chemical signals, via asymmetric intercellular junctions. These so-called synapses are macromolecular structures comprised of pre- and postsynaptic specializations and a synaptic cleft, which mediate cell-cell adhesion and intercellular signaling.

Synapses transmit signals between neurons in one of two ways. More direct and rapid communication occurs at electrical synapses, which act as conductive gap junctions for ionic currents[30,31]. Most neurons, however, rely on chemical synapses. These convert the information relayed in an action potential into a chemical signal, through the release of neurotransmitters from presynaptic neurons, and then back to an electrical impulse in postsynaptic cells[31,32]. Specifically, once an action potential reaches the presynaptic terminal, where neurotransmitters are packaged into vesicles, the voltage-gated Ca^{2+} channels are opened. The resulting influx of Ca^{2+} triggers the exocytotic fusion of these vesicles and the release of neurotransmitters into the synaptic cleft within milliseconds[33]. These neurotransmitters then bind postsynaptic receptors, which in turn induce a change in the membrane potential of the target neuron.

Structural and biochemical studies have revealed the molecular players involved in these processes. The presynaptic active zone contains proteins required for neurotransmitter release (vesicle trafficking, packaging, and recycling), while the postsynaptic density expresses neurotransmitter receptors and other molecules for downstream signal transduction[34–36]. In addition, the structural integrity and organization of these components are maintained by a host of scaffolding proteins. Although all synapses have these general features, they vary in their specific protein compositions to accommodate diverse functional properties, such as different types of neurotransmitters or strengths of synaptic transmission.

Through this network of proteins, synapses control the flow of information between neurons. The precise patterns of their assembly and the proper regulation of synaptic stability dictate the connectivity of neural networks and, thus, are critical to the functions of the nervous system. Indeed, proper synapse formation during early development is fundamental to learning, memory, and cognition during adulthood. Decades of research have deepened our understanding of the molecular mechanisms by which signals are transduced from a pre- to a postsynaptic neuron. However, how these synapses are formed during development and maintained through adulthood in a spatiotemporal manner is still largely unknown. Elucidating the mechanisms by which neurons initiate formation of synaptic structures is key to addressing how this process is regulated, how it contributes to synaptic function, and how it malfunctions in the context of neuropsychiatric and neurodegenerative diseases.

1.2.2 Synaptogenic cell adhesion molecules and their functional interactions

Synaptic junctions contain cell-adhesion molecules (CAMs), which facilitate cell adhesion and intercellular signaling by engaging in homophilic or heterophilic interactions across the synaptic cleft[37]. Trans-synaptic complexes formed by CAMs initiate contact between a pre- and a postsynaptic site and mediate the downstream bidirectional organization of the active zone and the postsynaptic density[38]. Structural components of a synapse are recruited through a variety of mechanisms. Scaffolding molecules and synaptic vesicles are trafficked to the active zone via vesicular transport, while the assembly of postsynaptic structures, such as PSD-95, occurs mostly by the gradual accumulation of proteins[39]. The recruitment of PSD-95 is followed by that of neurotransmitter receptors, as the synapse matures.

Indeed, CAMs acts as critical synaptic organizers that induce and regulate synapse formation. An increasing number of CAM protein families have been characterized, including neurexins (NRXN), neuroligins (NLGN), leucine-rich repeat proteins (LRRTM), ephrins, cadherins, and Ig-domain proteins such as SynCAMs (CADM). The bidirectional signaling induced by these complexes enables rapid recruitment of proteins necessary for synaptic development, while the specificity of these interactions contributes to the spatiotemporal regulation of synapse assembly.

Though some of these have been shown to induce synaptogenesis, how they direct pre- and postsynaptic organization in neighboring cells and what factors regulate their expression levels and functions are not well understood. The identification of many different CAMs – with shared and unique binding partners – suggests that synaptogenesis is mediated by diverse but likely redundant pathways[40]. This raises questions regarding the role of genetic interactions in the induction and regulation of synapse formation. Do these molecules act independently? If not, how do they coordinate their functions?

Recent studies have begun addressing these questions by identifying cooperative effects in synaptogenesis. For example, CDH2 facilitates target cell recognition for NLGN1 to induce synapse formation, while MDGA1 inhibits the synaptogenic activity of NLGN2 by blocking its interactions with neurexin[41,42]. These findings suggest that multiple cell-adhesion signals are coordinated to induce synapse formation. This cooperativity between synaptic proteins likely contributes to the specificity and tight regulation of synaptic differentiation. With so many CAMs and other molecules present in the synaptic proteome, however, the number of possible combinations presents a challenge for systematically assessing which genes interact.

1.2.3 Genetic approaches to identify synaptic organizers

As already discussed, genetic manipulations in cell models and model organisms are powerful tools for causally linking a genotype with its functional phenotype. Model organisms, such as *C. elegans*, fruit flies, zebrafish, or mice, have been widely used to study synapse biology. Both loss-of-function and gain-of-function approaches have been applied to generate mutant and transgenic organisms in which to observe resulting changes in synaptic patterns, structural defects during development, or behaviors. In some models, such methods can be applied at a higher throughput to systematically uncover pre- and postsynaptic organizers. Screens performed in *C. elegans*, *Drosophila*, and zebrafish have implicated genes in signaling pathways regulating synapse formation, organization, and growth[43–50]. However, although model organisms provide critical insights into these fundamental processes, they cannot fully recapitulate mammalian biology, due to functional differences in homologs, isoforms, and binding partners[51].

In vitro models can thus provide a more relevant context for elucidating mechanisms of synapse formation. Primary cells (e.g. from mice, rats, or humans) and iPSC-derived neurons are amenable to high-throughput genetic screening. Indeed, ORF overexpression screens in co-cultures and shRNA knockdown screens in primary neurons have revealed many regulators of synaptogenesis, including the LRRTM family and LRP6[52–55].

1.2.4 Image-based phenotypes in synapse formation assays

Image-based profiling in primary neuronal cultures, upon genetic perturbation, has helped identify synaptic CAMs and investigate their functional mechanisms. However, isolating the effects of a particular molecule in genetically perturbed neurons remains challenging. Neuronal cultures are heterogenous, and the expression of other proteins may compensate for any changes in the molecule of interest. These features often result in smaller dynamic ranges of observed phenotypes and obscure the effects of subtle modulators[56].

To this end, artificial synapse formation (ASF) assays have been developed to investigate the role of synaptic molecules in a more targeted manner. In these assays, neurons are co-cultured with non-neuronal cells, such as HEK293T (293T), COS, or PC12 cells, that overexpress a protein of interest[57]. This protein can then trigger the heterologous synapse formation between neighboring neurons and a non-neuronal cell. These synapses are functional and can be analyzed using imaging (e.g. immunostaining of synaptic markers, live-cell staining for vesicle recycling) or electrophysiological (e.g. patch-clamp recording) assays. Due to the diversity and genetic redundancy of the synaptic proteome makes, attributing a functional phenotype to a specific protein in neurons is difficult. In contrast, non-neuronal cells provide an isolated setting to study the effects of overexpressed proteins, eliminating many confounding factors that compensate for or interfere with the activity of a synaptogenic molecule. By simplifying the synaptic context, these co-culture experiments provide a convenient system in which to assess the functional importance of genetic factors in synaptogenesis, and have been used to discover many candidate synaptic organizers, including neuroligins and SynCAM[58–63].

Given the increasing appreciation for the complexity of signals required to induce synapse assembly, beyond the canonical trans-synaptic interactions of CAMs, we sought to develop a scalable approach to probe and map the regulatory network driving synaptogenesis. In Chapter 2, we integrated the ASF assay into the optical pooled screening workflow to enable high-throughput, systematic identification of functional interactions of a prominent postsynaptic organizer, neuroligin-1 (NLGN1). We analyzed the effects of 644 synaptic gene knockouts on NLGN1-induced presynaptic organization in over 1.8 million cells. By leveraging the high-content imaging dataset of both excitatory and inhibitory synapses, we implicated diverse cellular mechanisms, including cell adhesion, cytoskeletal dynamics, and signal transduction, in synapse formation.

1.3 Applications of virus-like particles

Retroviruses are a diverse family of viruses that integrate a copy of their single-stranded RNA genome into the host's DNA. This transfer of genetic material to other cells requires the efficient packaging of nucleic acids into virions. Retroviral assembly is mediated by the structural polyprotein Gag, whose key domains include the matrix (MA), capsid (CA), and nucleocapsid (NC)[64]. Through its NC domain near the C-terminal, Gag binds genomic RNA, which itself is a crucial structural component of viral particles[65]. The myristoylated N-terminal MA domain associates with the plasma membrane, while the CA domain facilitates Gag-Gag interactions, forming the viral protein core[64,66–69]. These mechanisms cooperate to induce the assembly and budding of the virion[70].

Viral structural proteins, like Gag, are sufficient for the self-assembly of virus-like particles (VLPs), rendering these nanostructures an attractive platform for biotechnological applications. One example is the use of VLPs as a potently immunogenic but safer alternative to conventional vaccines by leveraging the ability of VLPs to display antigens on their surface[71]. VLPs can also serve as delivery platforms, given their ability to package a range of cargo, including small molecule drugs, nucleic acids, and proteins. Indeed, several groups have engineered C-terminal Gag fusions to enable delivery of proteins of interest and even ribonucleoprotein complexes of CRISPR-based gene editing systems[72–76]. Gag fusions with bait proteins can also be applied to detect protein-protein interactions[77]. Moreover, since virions naturally package viral genomes, VLPs are well-suited for nucleic acid transfer. In particular, mammalian Gag homologs, such as PEG10, have been repurposed to package specific mRNAs for therapeutic delivery[78]. *In vivo*, Arc, another Gag homolog, has been shown to self-assemble into VLPs and transfer its own mRNA to neighboring neurons[79,80].

These examples therefore highlight the opportunities for leveraging retroviral elements in bioengineering. In Chapter 4, we adapted the murine leukemia virus (MLV) Gag to encapsulate and export cellular RNA, enabling longitudinal recording of transcriptional states in living cells.

Chapter 2. Identifying regulators of synaptogenesis using optical pooled screens

Authors: Anna Le, Thomas Biederer, Paul C. Blainey

2.1 Introduction

Synapses control the flow of information between neurons by mediating cell-cell adhesion and signaling through a complex network of proteins. The dynamic and precise regulation of synaptogenesis is thus crucial to the proper wiring and biological function of the nervous system. A diversity of cell-adhesion molecules (CAMs) is instrumental to the formation and maintenance of neuronal synapses, facilitating the bidirectional organization of components in the active zone and the postsynaptic density through trans-synaptic interactions[37–39,81]. Identifying CAM protein families, including neurexins, neuroligins, leucine-rich repeat proteins, ephrin B receptors, and cadherins, has therefore been critical for advancing our understanding of synapse biology and synaptogenesis in particular.

Genetic screening in model systems has been a powerful approach to identify pre- and postsynaptic organizers. Since isolating the cell-autonomous effects of a particular molecule in genetically perturbed neurons is difficult in neural tissues, artificial synapse formation (ASF) assays were developed to investigate the role of synaptic CAMs in a simplified context[56,57]. In these assays, neurons are co-cultured with non-neuronal cells overexpressing an adhesion molecule of interest, which can then trigger heterologous synapse formation atop the non-neuronal cells. ASF assays have been crucial in the discovery of several CAMs, and have been successfully coupled with image-based readouts and gain-of-function genetic perturbations to uncover novel regulators of synapse assembly[52,53,58–63].

Despite these advances, mechanisms that dictate the specificity of synaptogenic factors are still largely unknown. Recent work has begun to hint at the intricate network of molecules that regulate the ability of CAMs to direct pre- and postsynaptic assembly across synaptic junctions[41,42,54,82–88]. However, these studies have relied on low-throughput, arrayed experiments, which has hindered systematic analysis of the large number of possible interactions between CAMs and other factors in the synaptic proteome. To deepen our understanding of synapse formation, we therefore sought to map the functional interactions between synaptogenic proteins at scale.

To elucidate these combinatorial factors in synapse assembly, we deployed an optical pooled screening (OPS) approach, which enables highly robust and scalable image-based genetic screens of thousands to millions of cells. OPS uses *in situ* sequencing to detect perturbation identities, allowing direct matching of the genetic perturbations to the cellular phenotypes of single cells[23,26,28,29]. Specifically, we coupled *in situ* sequencing with an ASF assay to perform a large-scale screen for modulators of a known synaptic organizer, neuroligin-1 (NLGN1). NLGN1 is a critical postsynaptic cell adhesion protein that interacts with neurexins to regulate synapse assembly and specification[58,89–91]. Prior work demonstrated that NLGN1-

expressing HEK293T cells induce heterologous synapse formation, but not when N-cadherin (*CDH2*) is knocked out[41]. This finding highlights that synaptogenic CAMs do not act on their own, but rather function in cooperation with other proteins. We reasoned a similar loss-of-function approach applied in an unbiased manner at scale could reveal new mechanisms modulating synaptic cell adhesion molecules.

Here, we built and applied a scalable screening platform to illuminate the functional interactions underlying synaptogenesis. In the first demonstration of OPS for non-cell autonomous phenotypes, we integrated optical pooled CRISPR knockout-based screening with the heterologous synapse formation assay to investigate the roles of 644 synaptic genes in NLGN1-induced presynaptic organization. Based on a multi-dimensional analysis of excitatory and inhibitory synapses conducted in almost two million cells, we demonstrated how diverse cellular mechanisms, including cell adhesion, signal transduction, and cytoskeletal dynamics, contribute to synaptogenesis. We envision that OPS will enable routine functional interrogation of molecular interactions driving synapse formation and other neuronal phenotypes, providing an important tool for further exploration of neuronal biology.

2.2 Results

2.2.1 Design of a high-throughput CRISPR knockout screen for synaptogenic regulators

To systematically probe the effects of individual genes on synapse assembly, we established a pooled screening approach for ASF assays in which non-neuronal cells can induce pre- or postsynaptic specializations in co-cultured primary neurons. By narrowing the analysis of complex trans-synaptic signals at neuronal surfaces to those received from non-neuronal cells, this approach can isolate the contributions of specific genetic perturbations. We reasoned that integrating the ASF assay with optical pooled screening would address the limited throughput and heterogeneity of co-cultures, enabling large-scale screens for regulators of synapse formation (Figure 2.1A)[23,57].

Using this screening approach, we sought to identify cellular factors that modulate the induction of synaptogenesis by a prominent synaptic organizer, NLGN1. We first engineered HEK293 cells to contain a stably integrated, doxycycline-inducible Cas9 construct[92]. We then stably overexpressed the full-length coding sequence of wild-type human *NLGN1* in those HEK293 cells from an integrated lentiviral vector (Figure 2.1A). The sequence included an HA-tag following the N-terminal signal sequence to enable surface immunostaining of NLGN1-overexpressing HEK293 cells, while maintaining membrane localization and synaptogenic function of the protein. We chose HEK293 cells as the non-neuronal model for the screen, as they have been previously validated for ASF assays, can be easily engineered to introduce genetic perturbations, and endogenously express a subset of synaptic genes due to their neuronal origin[93,94]. Next, we defined the set of genes to target via CRISPR-mediated knockout. We selected 644 synaptic genes that are endogenously expressed in HEK293 cells by comparing the complete list of synaptic genes in the SynGO database with published RNA and protein expression profiles for HEK293 from the Human Protein Atlas (Methods)[95,96]. We designed and cloned a pooled plasmid library of 2,626 single guide RNAs (sgRNAs), consisting of 4 sgRNAs targeting each gene and 50 non-targeting sgRNAs (Methods). We then lentivirally transduced this sgRNA library at a low multiplicity-of-infection (MOI < 0.1) into NLGN1-overexpressing HEK293 cells with inducible Cas9.

To perform our screen, we induced the expression of Cas9 for 7 days, resulting in a pool of HEK293 cells, each with a single knockout of a gene target. Next, we co-cultured these cells with DIV 8 E18 primary rat hippocampal neurons for 24 hours (Figure 2.1A). To quantify the induction of presynaptic specializations in co-cultured neurons by NLGN1 expressed in HEK293 cells, we fixed and stained the cells for nuclei (DAPI), dendrites (anti-MAP2 antibody), NLGN1 (anti-HA-tag antibody), and presynaptic proteins (anti-VGAT, anti-Bassoon, and anti-VGLUT1 antibodies) (Figure 2.1B). After imaging these synaptic phenotypes, we performed *in situ* amplification and sequencing by synthesis of the sgRNAs (Figure 2.1C)[23,26]. We then extracted 263 geometry- and intensity-based features at the cell and puncta levels from phenotype images and obtained the corresponding sgRNA sequences in each HEK293 cell from the sequencing images (Supplementary Figure 2.1A; Methods). This automated analysis pipeline generated a phenotypic profile with a mapped sgRNA identity for 1,843,247 cells with a median of 2,924 cells per gene target across 4 sgRNAs (Figure 2.1D, Supplementary Figure 2.1B-D).

To confirm whether the pooled screen recovered the expected phenotypes of negative and positive controls, we compared the extent of NLGN1-induced presynaptic specializations in cells containing non-targeting sgRNAs to those with sgRNAs targeting either *NLGN1* or *CDH2* (encoding N-cadherin) (Figure 2.1E). As expected, HEK293 cells that expressed HA-tagged NLGN1 and a non-targeting sgRNA induced substantial synapse formation in co-cultured neurons. In contrast, cells with knockouts of *NLGN1* (evidenced by decreased expression of the HA-tagged protein) or *CDH2* resulted in significantly decreased synapse assembly, as quantified by fewer Bassoon puncta on the surface of these HEK293 cells, relative to those with non-targeting sgRNAs (Figure 2.1F, Supplementary Figure 2.1E). These data indicate that NLGN1 stimulates synaptogenesis and that N-cadherin is required for its activity, consistent with previous studies demonstrating the requirement of those two proteins for heterologous synapse formation[41,58]. Our screening approach thus was able to recapitulate the control phenotypes and quantitatively characterize synapse development. To reduce the effect of well- and plate-level variability in comparisons between gene targets across the 8 6-well plates used in the screen, we normalized the phenotypic profiles for subsequent analyses by calculating the median robust z-scores of features for each cell relative to local non-targeting control cells in the same well or row of a well-plate (Supplementary Figure 2.1F).

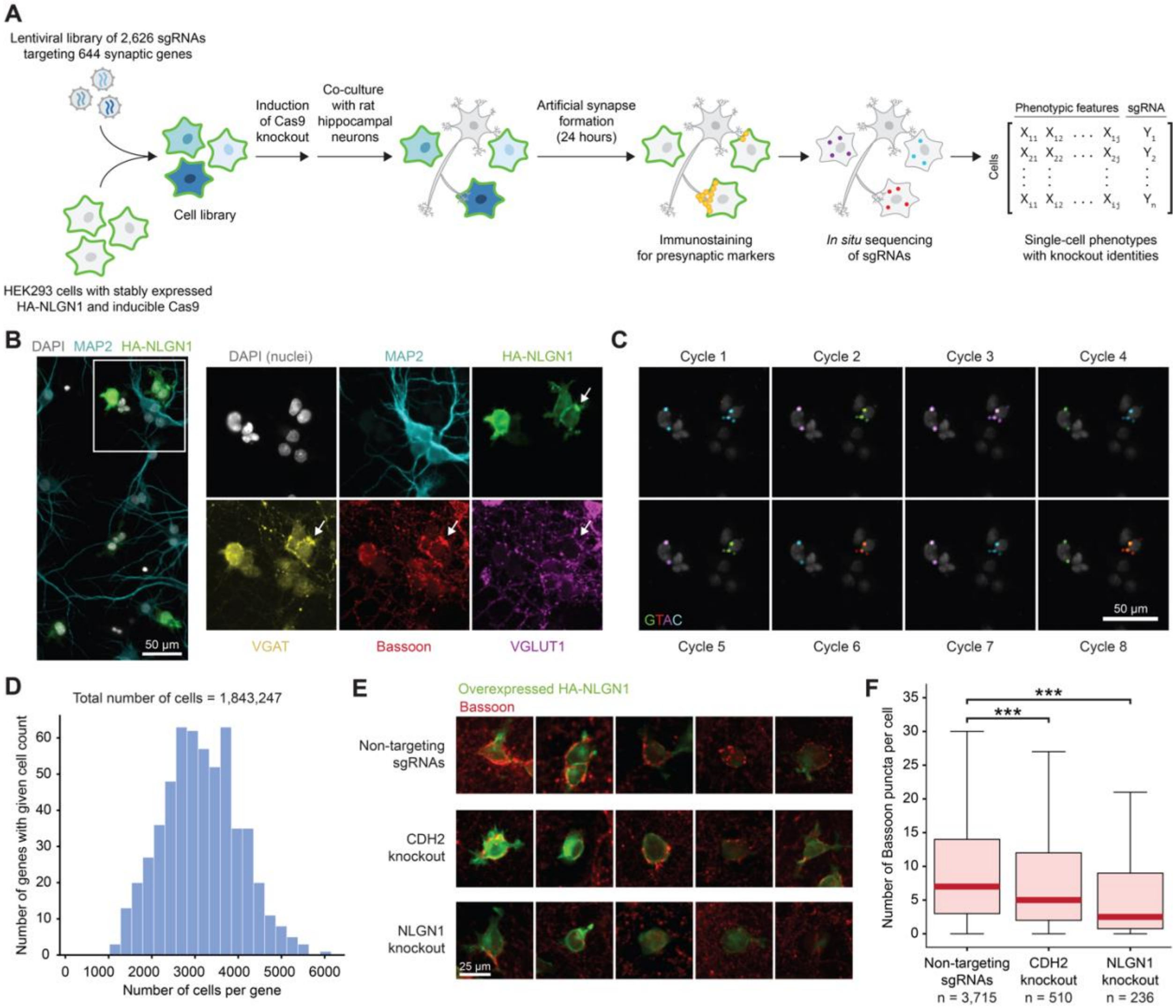


Figure 2.1 | Image-based pooled CRISPR screening enables high-throughput discovery of synaptogenic regulators.

(A) Workflow of optical pooled CRISPR knockout screen. HEK293 cells stably expressing HA-tagged NLGN1 and inducible Cas9 were transduced with a library of sgRNAs. Synapse formation was measured after 24-hour co-culture of HEK293 cells with hippocampal neurons by immunostaining of synaptic markers. The sgRNA identity in each HEK293 cell was then determined by *in situ* sequencing.

(B) Example field-of-view from the screen. Insets show stains used. Arrows indicate presynaptic specialization induced by NLGN1. Scale bar, 50 μm.

(C) Images of the first 8 cycles of *in situ* sequencing (Laplacian-of-Gaussian filtered) for the same field-of-view as in (B). Scale bar, 50 μm.

(D) Histogram of the number of cells analyzed for each highly expressed gene target (n = 1,843,247 cells analyzed in the primary screen).

(E) Example images of Bassoon (red) and HA-NLGN1 (green) staining from the screen show decreased induction of synaptogenesis by HEK293 cells in which *CDH2* or *NLGN1* is knocked out, as compared to the cells with non-targeting sgRNAs. Scale bar, 25 μm.

(F) Quantification of presynaptic specializations in the negative and positive controls. Box plots show the number of Bassoon puncta per HEK293 cell in one well plate (non-targeting sgRNAs, n = 3,715 cells; *CDH2* knockout, n = 510 cells; *NLGN1* knockout, n = 236 cells). Statistical significance measured by two-tailed Mann-Whitney U test comparing each condition to the non-targeting control (***p<0.0001).

2.2.2 High-dimensional phenotypic profiles reveal the role of NLGN1 expression in heterologous synapse formation

We first sought to leverage the multi-dimensional feature profiles extracted from the screen to determine the relationships between synaptic phenotypes of tested perturbations. After removing highly correlated redundant phenotypic parameters, we computed summary scores from the 95 remaining features at the sgRNA level. For each sgRNA and feature, we calculated the difference between the area under the curve (ΔAUC) of the cumulative distribution of the non-targeting control cells and that of the perturbed cells. We then averaged the ΔAUC of all 4 sgRNAs targeting a given gene to obtain a gene-level aggregate feature score. To identify gene targets with similar synaptic phenotypes, we visualized their phenotypic profiles using the PHATE algorithm and clustered them using the Leiden algorithm (Figure 2.2A, Supplementary Figure 2.2A-B; Methods)[97,98].

We next asked whether the identified clusters varied by the effect of gene targets on NLGN1-mediated synapse assembly. We first defined a set of features that best captures the extent of heterologous synapse formation induced by a HEK293 cell. We performed principal component analysis on the gene-level phenotypic profiles, consisting of the 95 selected features, and separated the non-targeting sgRNAs from the NLGN1 and CDH2 positive controls on the first principal component (PC1) (Figure 2.2B). The top loadings of PC1 included intensity-, geometry-, and count-based metrics of the presynaptic markers used in the screen: Bassoon, which labeled all presynaptic sites, as well as VGLUT1 and VGAT, which marked excitatory and inhibitory presynaptic specializations, respectively (Figure 2.2C). Informed by these loadings and commonly used metrics from published ASF assays, we defined a “synaptogenesis score” for each HEK293 cell as the sum of the median robust z-scores from 6 features: the total area occupied by puncta normalized to the cell area (“normalized area”) and the total number of puncta per cell (“puncta count”) for each presynaptic marker. To determine the effect of a given gene knockout on NLGN1-induced synapse development, we computed the mean ΔAUC of the synaptogenesis score from the cumulative distribution of cells perturbed with gene-targeting sgRNAs compared to that of cells perturbed with non-targeting sgRNAs, similarly to our calculation of gene-level summary scores for other features.

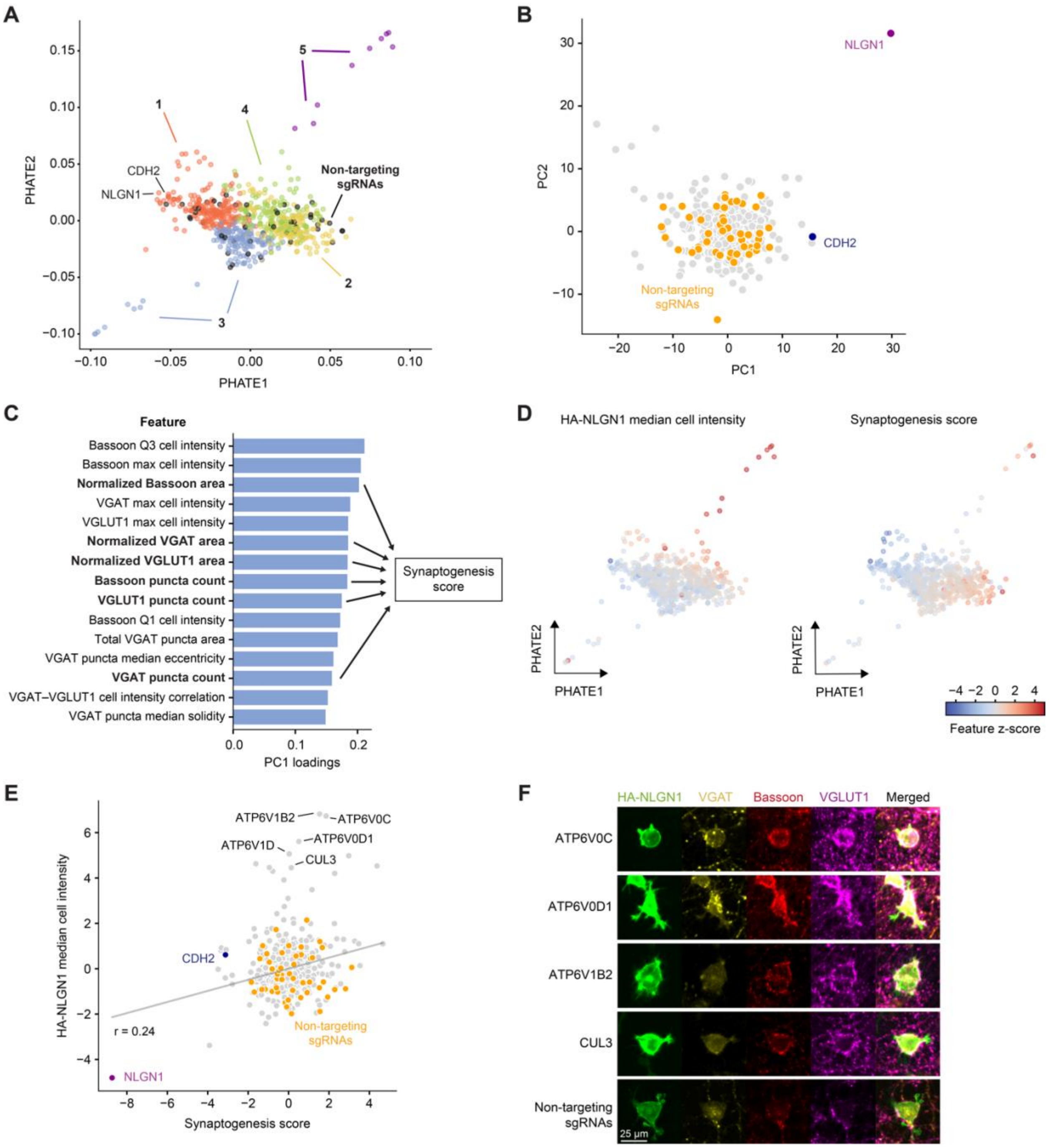


Figure 2.2 | Multi-dimensional phenotypes underlying synaptogenesis.

(A) Two-dimensional PHATE visualization of synaptic phenotype gene profiles. Profiles were aggregated using 95 filtered phenotype parameters for each of the 644 gene knockouts and 50 non-targeting sgRNAs. Colors indicate Leiden clusters, with non-targeting sgRNAs shown in black.

(B) Two-dimensional PCA scores plot of all gene knockouts. PCA was performed on the phenotypic profiles, consisting of 95 filtered features, for 644 gene knockouts and 50 non-targeting sgRNAs. Negative controls are highlighted in orange, and positive controls are shown in navy (CDH2) and purple (NLGN1).

(C) Top 15 loadings of PC1 include puncta area, count, and intensity features for Bassoon, VGLUT1, and VGAT stains. Puncta count and normalized area median robust z-scores for those three markers (bolded) were summed into a synaptogenesis score for each cell. The synaptogenesis scores were aggregated to the gene level by first computing the area under the curve (AUC) difference between cumulative distribution functions of cells containing each gene-targeting sgRNA and those with non-targeting sgRNAs (ΔAUC), followed by taking the mean ΔAUC of all sgRNAs scores targeting the same gene.

(D) Same plot shown in (A) with genes colored by HA-NLGN1 median cell intensity (left) and synaptogenesis score (right).

(E) Scatter plot comparing HA-NLGN1 median cell intensities and synaptogenesis scores for all genes, Pearson correlation coefficient r = 0.24. Non-targeting controls are highlighted in orange, and positive controls are shown in navy (CDH2) and purple (NLGN1).

(F) Illustrative images show gene knockouts labeled in (E) with increased expression of HA-NLGN1 (green) and modestly increased induction of presynaptic specialization (VGAT, yellow; Bassoon, red; VGLUT1, magenta), relative to non-targeting controls. Scale bar, 25 μm.

When we overlaid this resulting synaptogenesis score and the HA-NLGN1 median cell intensity (as a measure of NLGN1 protein expression) onto the PHATE visualization, we found that these two features mostly varied along different axes, defining the similarities within the groupings identified by Leiden clustering (Figure 2.2D, Supplementary Figure 2.2C-E). Cluster 3 included genes whose knockouts generally resulted in lower NLGN1 expression and modestly reduced induction of presynaptic specializations, while clusters 4 and 5 contained genes with opposite effects. Gene knockouts in clusters 1 and 2 exhibited more subtle differences in NLGN1 expression but resulted in larger changes in heterologous synapse formation.

Indeed, we observed a weak correlation between the synaptogenesis score and HA-NLGN1 median cell intensity (r = 0.24) (Figure 2.2E). Several gene knockouts resulted in substantially higher NLGN1 expression with some increase in the induction of presynaptic specializations by NLGN1 (Figure 2.2F). These genes included *CUL3*, a ubiquitin ligase, and multiple components of the vacuolar ATPase (v-ATPase), a complex responsible for acidifying organelles and critical for membrane trafficking and protein degradation[99,100]. The similar phenotypic profiles of v-ATPase subunits further demonstrate the robustness of our approach and highlight the role of protein sorting of membrane proteins in synapse development. Together, this analysis indicates that genes regulating NLGN1 expression may have some effect on heterologous synapse formation, but many perturbations affect synapse development by mechanisms other than altering protein expression of this synaptic organizer.

2.2.3 Synaptogenesis score captures broad modulators of NLGN1

With the synaptogenesis score we established above, we further explored these modulators of NLGN1-induced presynaptic specialization. As expected, knockouts of *NLGN1* and *CDH2* yielded significantly lower synaptogenesis scores relative to non-targeting controls (Figure 2.3A). Other perturbations that resulted in decreased heterologous synapse formation included factors involved in receptor tyrosine kinase signaling: *CRKL* (an adaptor protein that was previously shown to regulate neuromuscular synapse formation and acetylcholine receptor clustering) and *PTEN* (a critical component of the PTEN/PI3K/AKT signaling pathway; deficits in PTEN in development have been demonstrated to impair neuronal growth and synaptic plasticity) (Figure 2.3B-C, Supplementary Figure 2.3A-B)[88,101–104]. We observed that knockout of profilin-1 (*PFN1*) also impaired the assembly of presynaptic sites (Figure 2.3B-C, Supplementary Figure 2.3A-B). PFN1 is a modulator of actin polymerization and dynamics that is critical for spine formation in development and has been implicated in neurodevelopmental and neurodegenerative disorders, including Fragile X syndrome and amyotrophic lateral sclerosis[105,106].

Conversely, we identified several gene knockouts that significantly increased heterologous synapse assembly induced by NLGN1, including *CADM1* (Figure 2.3D-E, Supplementary Figure 2.3C-D). CADM1, also known as SynCAM 1, is an immunoglobulin cell adhesion molecule that, like NLGN1, is clustered at the postsynaptic surface and promotes excitatory synapse formation and modulates synaptic plasticity[59,107,108]. How CADM1 can act as a presumed negative regulator of NLGN1 remains to be determined. Moreover, depletions of NSF and its binding partner NAPA both resulted in an increased synaptogenesis score relative to non-targeting sgRNAs (Figure 2.3D-E, Supplementary Figure 2.3C-D). Although the most notable roles of NSF are facilitating vesicle fusion as a chaperone of the SNARE complex and regulating trafficking of AMPARs and other membrane proteins, NSF may also be involved in other cellular processes that ultimately affect NLGN1 activity, such as cytoskeletal dynamics and intracellular signaling through its interactions with Rab effectors[109,110]. The gene knockout that exhibited the largest increase in the synaptogenesis score was *SFPQ*, a ubiquitous splicing factor (Figure 2.3D-E, Supplementary Figure 2.3C-D). Alternative splicing is relevant for many neuronal functions, and the effect we observed with the *SFPQ* knockout shows a likelihood of proteoform-specific effects on synapse development[111]. Interestingly, knockouts of *NSF*, *NAPA*, and *SFPQ* clustered similarly and exhibited increased HA-NLGN1 cell intensity in addition to increased heterologous synapse formation (Figure 2.2D, Supplementary Figure 2.3E). This finding reaffirms the important role that the regulation of membrane protein expression plays in synapse assembly.

2.2.4 High-content imaging uncovers genes controlling synaptic specificity

Additionally, we noted that the synaptogenesis score describes the overall change in synapse assembly, but does not distinguish their specific types. While NLGN1 is a prominent organizer of excitatory synapses, it can also induce inhibitory specializations. To explore the genes with distinctive effects on excitatory or inhibitory synapses, we first compared the VGAT (for inhibitory) and VGLUT1 (for excitatory synapses) components of the synaptogenesis score and found that they were moderately correlated ($r = 0.49$) (Figure 2.3F). However, several gene targets exhibited more strongly differential effects on excitatory and inhibitory synapses (Figure 2.3G). We observed substantial decreases in the formation of excitatory but not inhibitory synapses in knockouts of *VPS35* (a member of the retromer complex previously shown to affect excitatory but not inhibitory currents in mouse neurons) and *ROR2* (a Wnt5a receptor, which regulates synapse development and potentiates synaptic transmission at excitatory glutamatergic synapses)[112–114]. Conversely, we found that knockouts of two modulators of cell adhesion, *CDC42* (a Rho GTPase and key regulator of cytoskeletal dynamics) and *DAG1* (a dystrophin-associated protein) exhibit stronger effects on inhibitory than excitatory synapse assembly, consistent with previous reports[115–118]. Analyzing the complex information extracted from the rich imaging dataset can thus reveal regulators driving synaptic specificity.

Because NLGN1 primarily induces the assembly of excitatory synapses, we reasoned that changes in inhibitory synapse formation may be rarer and less apparent in the aggregate synaptogenesis score. To determine factors critical for inducing inhibitory presynaptic specializations, we identified genes that altered the total area of VGAT puncta normalized to the area of HEK293 cells (Supplementary Figure 2.3F-G). We observed that depletion of two cytoskeletal regulators, the intracellular cofilin-1 (CFL1) and the cell surface protein dystroglycan (DAG1), resulted in a significantly decreased normalized area of VGAT puncta. Prior work has demonstrated that NLGN1 modulates synapse number and plasticity through actin reorganization mediated by the LIMK1/CFL1 complex[119,120]. DAG1, on the other hand, is more specifically localized to inhibitory GABAergic synapses and acts as an adhesion molecule that binds ligands in the extracellular matrix; intracellularly, DAG1 is anchored to the actin cytoskeleton and interacts with downstream signaling proteins[117,121,122]. Since DAG1 was not a significant hit when we considered only the synaptogenesis score, this analysis highlights the need to examine individual features to capture gene targets with more selective effects.

2.2.5 Identification of novel regulators of NLGN1 synaptogenic activity

To summarize candidate genes that modulate NLGN1-induced synaptogenesis, we therefore calculated statistics of 7 features—the synaptogenesis score and its 6 component parameters—for all perturbations. We identified 102 genes from the screen, which significantly altered the induction of presynaptic specializations (false discovery rate, FDR < 0.1 in at least 2 of these 7 features). Of those, we further narrowed down a list of 39 top scoring genes (by lowering to FDR < 0.05 and setting an absolute feature z-score threshold of 1.5, for at least 2 of the 7 features). These top hits included 24 novel candidate positive regulators (whose knockouts decreased synapse formation) and 15 candidate negative regulators of NLGN1-driven synapse assembly (Figure 2.3H).

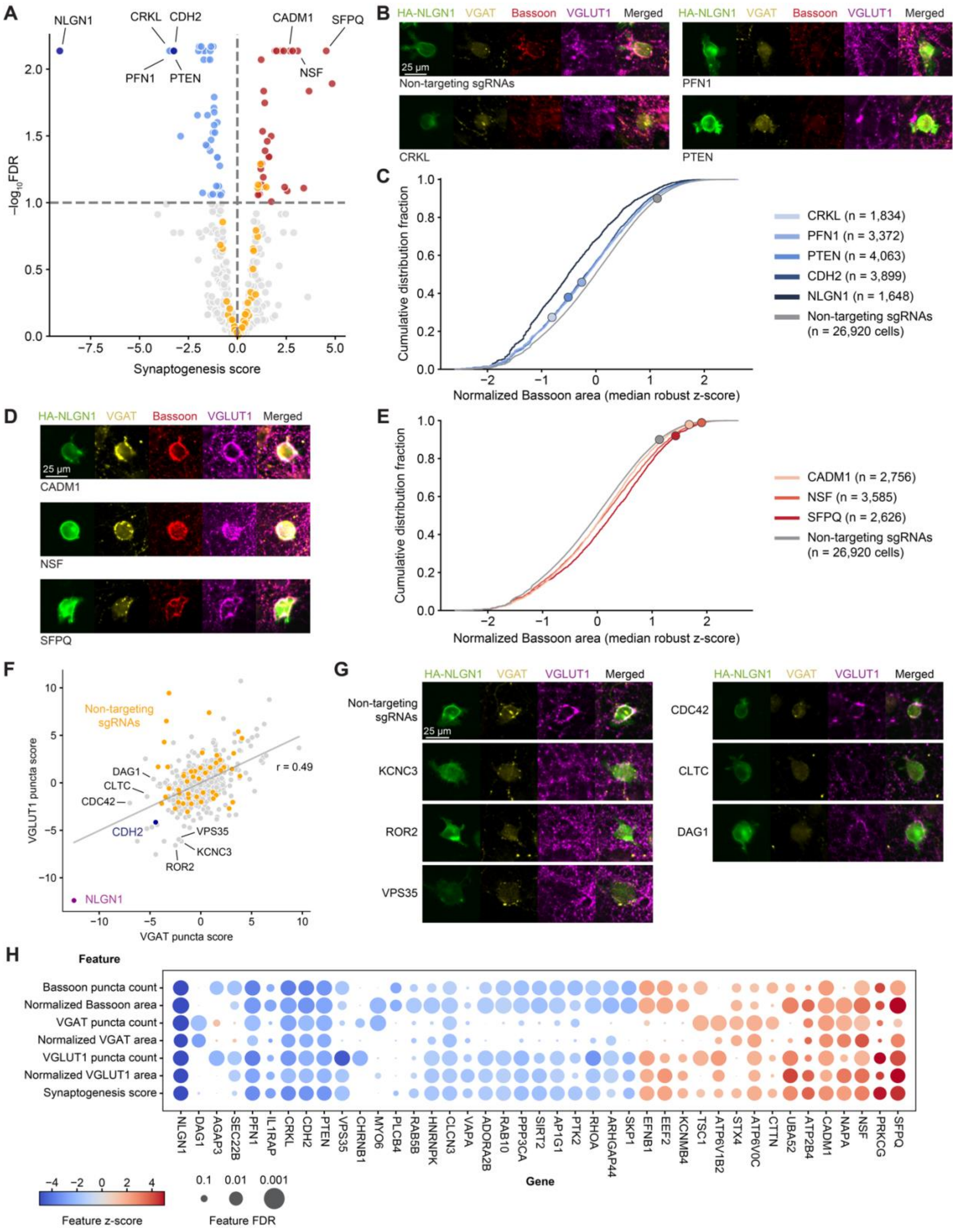


Figure 2.3 | Identification of positive and negative regulators of NLGN-1-induced synapse formation.

(A) Volcano plot of synaptogenesis scores (z-scored), highlighting genes whose knockouts resulted in decreased (blue; positive controls in navy) or increased (red) presynaptic specialization (false discovery rate, FDR < 0.1, dashed line). Orange dots represent 50 random samples of 6 non-targeting sgRNAs. Raw P-values were calculated from bootstrapped null distributions of cells expressing non-targeting sgRNAs, and FDR was corrected using the Benjamini-Hochberg procedure.
(B) Illustrative images of cells with gene knockouts highlighted in (A), showing decreased induction of presynaptic specialization. Scale bar, 25 µm.
(C) Empirical cumulative distribution functions for the normalized Bassoon puncta area (median robust z-score) of single cells with *CRKL*, *PFN1*, *PTEN*, *CDH2*, and *NLGN1* knockouts (shades of blue), compared to cells with non-targeting control sgRNAs (gray). Circles indicate cells shown in (B).
(D) Illustrative images of cells with gene knockouts highlighted in (A), showing increased heterologous synapse formation. Scale bar, 25 µm.
(E) Empirical cumulative distribution functions for the normalized Bassoon puncta area (median robust z-score) of single cells with *CADM1*, *NSF*, and *SFPQ* knockouts (shades of red), compared to cells with non-targeting control sgRNAs (gray). Circles indicate cells shown in (D).
(F) Scatter plot comparing VGLUT1 and VGAT puncta scores (summation of normalized area and count) for all genes, Pearson correlation coefficient r = 0.49. Non-targeting controls are highlighted in orange, and positive controls are shown in navy (CDH2) and purple (NLGN1).
(G) Illustrative HEK293 cell images (HA-NLGN1, green) for gene knockouts labeled in (F) show a preferential decrease in excitatory synapse formation (VGLUT1, magenta) (left) or a preferential decrease in inhibitory synapse formation (VGAT, yellow) (right), compared to the non-targeting controls. Scale bar, 25 µm.
(H) Heat map of synaptogenesis score and its component features for top scoring genes from the primary screen (FDR < 0.05 and z-score > 1.5 for at least two of the features). Color indicates the corresponding feature z-score, and circle size indicates the feature FDR. Gene order was obtained from hierarchical clustering.

Next, we sought to assess the robustness of our screening approach and validate the effects of top scoring hits by performing a secondary screen at a higher statistical power. We selected a subset of 35 genes from the primary screen dataset with the most significant gene-level normalized enrichment scores (NES), which can be less sensitive to guide-to-guide variability than an average ΔAUC score across all sgRNAs targeting the same gene (Supplementary Figure 2.3H; Methods). We designed a new sgRNA library, consisting of 4 sgRNAs for each gene target and 10 non-targeting sgRNAs, and transduced it into HA-NLGN1-expressing HEK293 cells. Following the same experimental and analytic protocols used for the primary screen, we obtained phenotypic profiles of 403,688 cells, with a median of 8,773 cells per gene knockout. We compared the synaptogenesis scores and the component features and observed substantial correlation between the primary and secondary screens (r = 0.5 to r = 0.66) (Supplementary Figure 2.3I-J). Using a similar statistical approach as in the primary screen, we found that 22 of the 33 tested candidate hits significantly affected heterologous synapse formation (FDR < 0.05 in at least 2 of the 7 features) (Supplementary Figure 2.3K). This consistency in results obtained from independent screens gave us confidence in the reproducibility of the methods and the hits identified from the screening approach.

2.2.6 Candidate NLGN1 regulators implicate adhesive, cytoskeletal, and signaling mechanisms in synaptogenesis

To better understand the mechanisms involved in synaptogenesis, we examined the cellular functions of all nominated hits. As expected, we implicated several genes involved in cell adhesion, including *CADM1* and *IL1RAP*, both known synaptic organizers[59,107,108,123]. Interestingly, we found that knockout of glypican-4 (*GPC4*) resulted in an increase in NLGN1-induced presynaptic specializations. Prior studies established that glypican-4 is localized to the presynaptic membrane, acting as a ligand for LRRTM4 and thus, a potential competitor of neurexin[124–126]. Other genes whose knockouts altered synapse assembly by NLGN1 included the transmembrane proteins integrin alpha-2 (*ITGA2*) and dystroglycan (*DAG1*), which bind ligands, such as laminins, in the extracellular matrix[117,127].

Another process critical for synapse development is the regulation of cytoskeletal dynamics. Several studies have demonstrated the requirement of actin remodeling, likely mediated by the activation of signaling pathways at the cell surface, in synaptogenesis[128–130]. In response to external stimuli, the intracellular domains of ITGA2 and DAG1 interact with actin-binding proteins to reorganize cytoskeletal structures. We also found a number of other gene targets that may modulate heterologous synapse formation by regulating actin polymerization. These factors include cofilin-1 (CFL1), profilin-1 and 2 (PFN1 and PFN2), drebrin-1 (DBN1), cortactin (CTTN), and CAPZB, all of which have putative neuronal functions[106,131–138]. Additionally, our screen identified PTK2 and the Rho GTPses CDC42 and RHOA as modulators of NLGN1-induced synaptic development[115,116,139–144]. All three are involved in the formation of focal adhesions and were clustered together on the PHATE visualization (Supplementary Figure 2.3E).

We also observed that signal transduction is a key aspect of synaptogenesis, as evidenced by the large number of hits that are signaling proteins. Previous work has demonstrated the importance of Wnt signaling in synapse development and function[145]. Indeed, several candidate NLGN1 regulators are components of Wnt-mediated pathways, including CDC42, RHOA, CTTN, and DVL1[146]. Our screen also implicated a range of protein kinases (such as ephrin B1, Src, and MAPK3/ERK1) and phosphatases (including subunits of protein phosphatases 2 and 3, PTEN, and PTPRA)[147–153,104,103,154]. Taken together, our findings support a diversity of known mechanisms of synapse development in which cell-matrix adhesions and trans-synaptic interactions by cell adhesion molecules result in signaling cascades that reorganize the cytoskeleton and recruit receptors and other synaptic proteins.

2.2.7 DAG1 and PTEN impair the induction of presynaptic specializations by NLGN1

To validate the role of identified factors in synaptogenesis, we investigated the effects of a subset of top scoring hits from the secondary screen. For each gene target, we generated clonal HEK293 cell lines expressing HA-tagged NLGN1, inducible Cas9, and an sgRNA targeting a given gene. Each clone was verified by sequencing and indel analysis to have a complete Cas9-induced knockout of the targeted gene. We then performed arrayed ASF assays with these clonal knockout cell lines (Supplementary Figure 2.4A). Knockouts of *PFN1*, *ROR2*, and *VCAN* (encoding versican, an extracellular matrix proteoglycan with diverse functions in proliferation, migration, and adhesion) show decreased synapse assembly, consistent with results from the pooled screens (Supplementary Figure 2.4B-C)[155]. We also found that *DAG1* and *PTEN* knockouts displayed significant and consistent effects across independent experiments. As we observed in our primary screen, depletion of PTEN in NLGN1-overexpressing cells showed a significant decrease in formation of both excitatory and inhibitory synapses (Figure 2.4A-B, Supplementary Figure 2.4B and D). Similarly, knockout of *DAG1* resulted in a significant decrease in the assembly of inhibitory synapses induced by NLGN1 and a smaller but still significant decrease in excitatory synapse formation (Figure 2.4A-B, Supplementary Figure 2.4B and D). These findings recapitulate in a clonal arrayed context our observation from the pooled screens that DAG1 selectively affects the development of inhibitory synapses.

Among the diverse hits we uncovered, we chose to further explore the effects of *DAG1* due to its specificity to inhibitory synapse assembly. We also looked at *PTEN* to further investigate the role of the PTEN/PI3K/AKT signaling pathway, given previous reports of its contribution to postsynaptic, but not presynaptic, synapse formation[88]. We thus analyzed the effects of these two genes on synapse development in primary neuronal cultures. We sparsely co-transfected DIV 6 primary rat cortical neurons with a vector expressing GFP-tagged human NLGN1 and a BFP-expressing vector containing a negative control shRNA or an shRNA targeting *Cdh2*, *Dag1*, or *Pten* (Methods). After 7 days of shRNA-mediated knockdown of gene targets, we stained for surface glutamate receptors (GluA) to measure changes in receptor recruitment at the postsynaptic membrane (Figure 2.4C, Supplementary Figure 2.4E-F). We observed a small but significant decrease, relative to negative controls, in the number of Bassoon-positive presynaptic terminals in contact with transfected neurons when *Cdh2* or *Dag1* were knocked down in the postsynaptic neurons, but not with *Pten*-depleted neurons (Figure 2.4D). Additionally, we found that only *Cdh2* and *Pten* knockdowns reduced the intensity of surface GluA at those Bassoon puncta (Figure 2.4E). These data suggest that in primary cultures, CDH2 affects both the induction of presynaptic specializations and postsynaptic receptor recruitment, while PTEN only modulates development at the postsynaptic surface. Our finding that *Dag1* knockdown yields fewer Bassoon-positive puncta without altering their GluA intensity indicates that DAG1 regulates the organization of presynaptic sites, but not glutamate receptor clustering at excitatory synapses. This observation is supported by prior work and our findings from ASF assays that DAG1 is specifically localized to inhibitory synapses and selectively modulates inhibitory synapse formation[122].

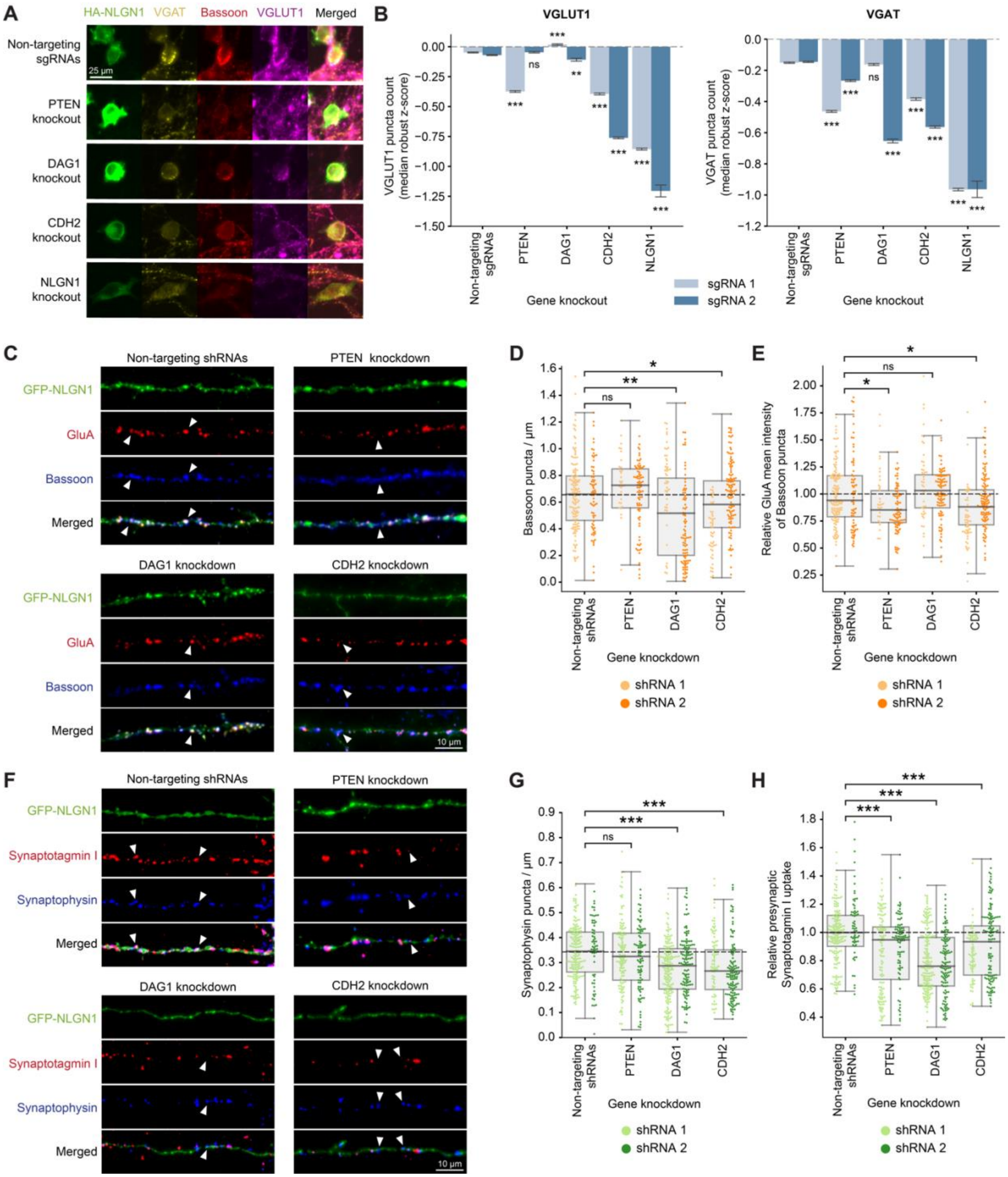


Figure 2.4 | Role of DAG1 and PTEN on presynaptic organization and vesicle recycling.

(A) Illustrative images of arrayed, single-cell clonal knockout HEK293 cells overexpressing HA-NLGN1 co-cultured with hippocampal neurons. CDH2 and NLGN1 are controls expected to reduce the number of presynaptic sites formed compared to the non-targeting control. Scale bar, 25 µm.

(B) Quantification of presynaptic specializations induced by HA-NLGN1-expressing clonal knockout HEK293 cells. Bar plots of the number of VGLUT1 (left) and VGAT (right) puncta per cell (median robust z-score) show means ± standard errors of the mean (SEM), from 4 independent co-cultures (number of cells analyzed listed in Supplementary Figure 2.4A). Statistical significance calculated by comparing each sgRNA to all non-targeting controls using one-way ANOVA test, followed by Tukey test for multiple comparisons (ns = p>0.01, **p<0.001, ***p<0001).

(C) Example images showing DIV 14 cortical neurons with shRNA-mediated knockdowns stained for surface GluA receptors (red) and presynaptic Bassoon (blue). Arrows indicate Bassoon clusters with varying GluA intensities. Scale bar, 10 µm.

(D-E) Box plots show (D) the number of Bassoon clusters per µm and (E) the relative GluA signal intensity in Bassoon clusters, from 2 independent cultures (number of cells and dendrites analyzed listed in Supplementary Figure 2.4D). Swarm plots show the distribution for each shRNA. Dashed line indicates the mean of non-targeting controls. Statistical significance calculated by comparing all observations for a given gene to the non-targeting controls using one-way ANOVA test and Tukey test for multiple comparisons (ns = p>0.05, *p<0.05, **p<0.001).

(F) Example images showing DIV 14 cortical neurons with shRNA-induced knockdowns after live antibody-labeled Synaptotagmin I (SytI) uptake (red) and fixed-cell staining for synaptophysin (blue). Arrows indicate presynaptic synaptophysin clusters with varying degrees of SytI uptake. Scale bar, 10 µm.

(G-H) Box plots show (G) the number of synaptophysin clusters per µm and (H) the relative mean SytI intensity in synaptophysin clusters, from 3 independent cultures (number of cells and dendrites analyzed listed in Supplementary Figure 2.4F). Swarm plots show the distribution for each shRNA. Dashed line indicates the mean of non-targeting controls. Statistical significance calculated by comparing all observations for a given gene to the non-targeting shRNAs using one-way ANOVA test, followed by Tukey test for multiple comparisons (ns = p>0.05, ***p<0001).

To further determine the impact of these candidate hits on presynaptic organization, we next tracked the uptake of synaptotagmin I (SytI)-bound antibodies upon depolarization as a measure of vesicle recycling from presynaptic terminals (Figure 2.4F, Supplementary Figure 2.4G-H). We again performed shRNA-mediated knockdowns of *Cdh2*, *Dag1*, and *Pten*, with concurrent expression of GFP-tagged NLGN1. We first observed that neurons with *Cdh2* or *Dag1* knockdown had fewer synaptophysin-positive presynaptic puncta at their surface (Figure 2.4G). This result agrees with our earlier finding that CDH2 and DAG1 depletion, but not that of PTEN, yielded fewer contacts with presynaptic Bassoon puncta. We then quantified the degree of SytI antibody uptake. We found that knockdowns of *Cdh2*, *Dag1*, and *Pten* all resulted in significantly lower intensities of SytI at synaptophysin clusters, indicating a lower rate of vesicle recycling at the presynaptic terminals in contact with these neurons (Figure 2.4H).

In summary, we observed that CDH2 and DAG1 affect both the number and functionality of synaptic junctions, but DAG1 functions as a specific modulator of inhibitory synapses. On the other hand, PTEN is primarily involved in regulating synaptic transmission, including through glutamate receptor clustering. Taken together, our findings demonstrate that the optical pooled screening approach can successfully identify synaptogenic regulators that impact synapse number and function through various mechanisms, allowing researchers to dissect the diverse cellular processes that underlie the complexity of synaptic development.

2.3 Discussion

In this study, we performed a pooled image-based screen of cell-cell interactions in a co-culture system to systematically characterize the role of 644 genes in synapse formation. High-content imaging and our automated analysis pipeline enabled simultaneous measurements of multiple synaptic phenotypes within the same cell, including the distinct scores indicative of inhibitory and excitatory synapses. With this approach, we captured multi-dimensional phenotypic profiles for over 1.8 million single cells, allowing us to robustly and directly compare synaptogenic effects between perturbations. We found that the regulation of the surface expression of synaptic organizers, such as NLGN1, is a key step in synapse assembly. This process, however, was not the sole determinant of heterologous synapse formation. Rather, our analysis identified many synaptogenic modulators involved in cell adhesion, extracellular matrix anchoring, and cytoskeletal (particularly actin) dynamics. We also implicated several signaling proteins in synapse development, from modulators of Wnt-mediated signaling to a number of kinases and phosphatases that interact with various pathways. We demonstrated, for example, that PTEN, a putative antagonist of the PI3K/AKT pathway, does not affect synapse number in primary cultures, but modulates postsynaptic receptor clustering. Additionally, we observed that while some genes modulate the assembly of excitatory and inhibitory synapses, others more selectively act on specific types of synapses, which further supports the need for multi-dimensional readouts of synapse formation. In particular, we found that DAG1, a cell adhesion and cytoskeletal regulator, affects the formation of inhibitory synapses, consistent with previous reports of its co-localization with GABAergic, but not glutamatergic, synapses[122]. Together, these data support the model of synaptogenesis in which strong trans-synaptic interactions and transduction of extracellular signals upon ligand binding are essential to the reorganization of the cytoskeleton to recruit and assemble factors at the synaptic terminals. By implicating many biological processes, our findings highlight the utility of a high-throughput method with multi-dimensional readout to systematically elucidate the complex network of interactions required to support synapse assembly.

Our work thus demonstrates the feasibility and scalability of OPS for the analysis of synaptogenic phenotypes and more broadly, cell-cell interactions. Single-cell, image-based measurements are suited to directly capture the spatial context of cellular phenotypes, including cell non-autonomous and time-dependent behaviors. When coupled with genetic perturbation screening, imaging assays are interpretable at the molecular level and constitute powerful tools to elucidate novel gene functions or mechanisms underlying a biological process, including synapse assembly[52–55]. Pooled genetic screens have increased throughput and less variability for many assays compared with arrayed screens, which is particularly impactful for analyses that suffer from high variability, such as the ASF assay used here[9]. In contrast to pooled image-based enrichment methods that require pre-defining scoring criteria and thresholds before the assay and cell selection, OPS, as a pooled profiling screening method, captures complex, high-dimensional phenotypes and the corresponding perturbation identity at the single-cell level, enabling post-hoc data exploration, scoring, and selective thresholding of cells with high-confidence perturbation identities[23,26,143,156–159].

This study used OPS to measure the effects of gene knockouts on synapse assembly induced by NLGN1 in an approach readily extensible to elucidate other mechanisms underlying synaptogenesis. For example, this screening workflow can be adapted to investigate regulators of other cell-adhesion molecules, including presynaptic proteins, measure additional markers of synaptic function, or include other cell types in the co-culture, such as neurons from different brain regions, to identify neuron type- or region-specific modulators of synapse development. Furthermore, while the CRISPR-mediated knockout approach limited us to probing synaptic genes that are endogenously expressed in HEK293 cells, future screens can apply other perturbation modes, including barcoded ORF overexpression or CRISPR activation for gain-of-function, to uncover the involvement of all synaptic genes in synapse formation. Our findings here, enabled by a systematic screening approach and a high-dimensional image-based dataset resolved at the single cell level in almost two million cells, present an important advance in illuminating the regulatory network driving synaptogenesis. Our work thus demonstrates the potential of large-scale genetic screening with optical readouts for many future investigations in neurobiology.

2.4 Materials and Methods

Culture conditions for immortalized cell lines

HEK293 and HEK293T were cultured in Dulbecco's Modified Eagle Medium (DMEM) with GlutaMAX and sodium pyruvate (Thermo Fisher Scientific) supplemented with 10% heat-inactivated fetal bovine serum (FBS) (Sigma-Aldrich), 100 U/mL penicillin, and 100 µg/mL streptomycin (Thermo Fisher Scientific). All cell types were passaged every 2–4 days, maintained below 90% confluency, and cultured at 37°C with 5% CO_2.

Primary neuron cultures

Hippocampal and cortical neurons were cultured from day 18 embryos isolated from pregnant female Sprague-Dawley rats (Charles River Laboratories). The hippocampus and cortex were dissected in Hibernate-E Medium (BrainBits, Thermo Fisher Scientific) supplemented with B27 (Thermo Fisher Scientific), 100 U/mL penicillin, and 100 µg/mL streptomycin. Hippocampi and cortices were then digested at 37˚C for 10 and 15 min, respectively, in a 1.2 mg/mL solution of papain (BrainBits, Transnetyx). After digestion, the tissue was washed and dissociated with pipettes of decreasing diameters in Neurobasal Medium (Thermo Fisher Scientific) supplemented with B27, GlutaMAX (Thermo Fisher Scientific), 5% FBS, 50 U/mL penicillin, and 50 µg/mL streptomycin. The cells were then pelleted and resuspended to the appropriate concentration before seeding onto 6-well (Cellvis) or 24-well (Greiner Bio-One) glass-bottom plates. The plates were coated with 50 μg/mL poly-D-lysine (PDL) (Sigma-Aldrich) prior to plating. The cultures were maintained for 3 days *in vitro* (DIV) in Neurobasal Medium supplemented with B27, GlutaMAX, 5% FBS, 50 U/mL penicillin, and 50 μg/mL streptomycin. On DIV 2, the neurons were treated with 2 μM cytosine β-D-arabinofuranoside (Sigma-Aldrich) for 24 hours. On DIV 3, the media was changed to Neurobasal Medium supplemented with B27, GlutaMAX , 50 U/mL penicillin, and 50 μg/mL streptomycin. Afterwards, a 50% medium change was performed every 3-4 days. Neurons were cultured at 37°C with 5% CO_2 up to DIV 14.

General methods for molecular cloning

Plasmids for protein expression were cloned using Gibson assembly[160]. Briefly, DNA fragments were digested with restriction enzymes (New England BioLabs, Thermo Fisher Scientific) or amplified using KAPA HiFi HotStart ReadyMix (Roche Molecular Systems) according to the manufacturers' protocols. Fragments were purified by electrophoresis with 1-2% agarose gels using QIAquick Gel Extraction Kit (QIAGEN), and then assembled with Gibson Assembly Master Mix (New England BioLabs).

Plasmids for expression of sgRNAs or shRNAs were cloned using Golden Gate assembly[161]. Briefly, a vector backbone was digested with a Type IIS restriction enzyme, such as Esp3I (Thermo Fisher Scientific), and purified via electrophoresis with 1% agarose gels using QIAquick Gel Extraction Kit (QIAGEN). DNA inserts were ordered as oligonucleotides (Integrated DNA Technologies), annealed, and ligated to the digested vector backbone using a T7 DNA ligase (QIAGEN).

Unless otherwise noted, all ligated plasmids were transformed into Stable Competent E. coli (New England BioLabs) and grown on Luria-Bertani (LB) agar with 100 μg/mL ampicillin. Bacteria containing plasmids verified by Sanger sequencing were grown in 2× YT medium with 100 μg/mL ampicillin. Plasmid DNA was isolated using the QIAGEN Plasmid Plus Midi Kit with endotoxin removal and twice the recommended amount of RNase A in Buffer P1.

General methods for lentivirus production and transduction to generate cell lines

HEK293T cells were seeded in 6-well plates (Corning) at a density of 100,000 cells/cm^2. After 12-16 h, cells were transfected at 80% confluency according to the manufacturer's protocol, with 12 μL Lipofectamine 3000 (Thermo Fisher Scientific), 12 μL P-3000, and lentiviral transfer plasmid, pMD2.G (Addgene #12259), and psPAX2 (Addgene #12260) at a mass ratio of 4:2:3 for a total of 2.5 μg. 6 h after transfection, media was exchanged to fresh pre-warmed media. 48 h after transfection, viral supernatant was harvested, centrifuged at 500 g and 4˚C for 5 min to remove cellular debris, filtered through a 0.45 μm Supor membrane syringe filter (Pall), and stored at -80 °C.

The lentivirus was first titered by transducing a range of virus volumes into HEK293 cells. The spinfection was performed by centrifuging lentivirus, 10 μg/mL polybrene (Sigma-Aldrich), and 1.5 × 10^6 cells per well in a 6-well plate (Corning) at 1000 g and 33˚C for 2 h. 4-24 h after infection, cells were expanded to a T75 flask in fresh media containing the appropriate concentration of a selection antibiotic. Once the no-infection control exhibited 100% cell death (between 2-7 days later, depending on the selection marker), the titer was calculated by counting the number of cells surviving in a field-of-view for each virus volume used. To generate cell lines with stable integration, the spinfection was scaled appropriately to achieve the desired MOI.

Cas9-expressing HEK293 cell line generation

The HEK293 cell line with inducible Cas9 used in the study was generated via piggyBac transposition. Wild-type HEK293 cells were seeded in 6-well plates (Corning) at a density of 100,000 cells/cm^2 in DMEM with GlutaMAX and sodium pyruvate (Thermo Fisher Scientific) supplemented with 10% FBS (Sigma-Aldrich). After 12-16 h, cells were transfected at 80% confluency with 8 µL of Lipofectamine 2000 (Thermo Fisher Scientific), as well as the piggyBac transposase plasmid and the PB-TRE-Cas9-EF1a-TetON-T2A-NEO transposon plasmid (Cheeseman Lab, Whitehead Institute) at a 1:2 mass ratio for a total of 2.5 µg[92]. 6 h after transfection, media was exchanged to fresh pre-warmed media. 24 h after transfection, cells were expanded to a T75 flask. Cells were then selected with 800 µg/mL G418 (Thermo Fisher Scientific) for 2 weeks.

Clonal cell lines were isolated using an SH800 Cell Sorter (Sony) and expanded. To verify Cas9 activity, clones were transduced with a guide targeting TFRC (spacer, 5′–GCTATACGCCACATAACCCCC–3′)[23]. The guide was cloned into CROPseq-puro-v2 (Addgene #127458) with Golden Gate assembly using the Esp3I restriction enzyme, and lentivirus was produced and spinfected as described above. Clones were screened after 7 days with and without 1 µg/mL doxycycline to induce Cas9 expression. Genomic DNA was extracted by resuspending cells in lysis buffer (10 mM Tris pH 7.5, 1 mM $CaCl_2$, 3 mM $MgCl_2$, 1 mM EDTA, 1% Triton X-100, and 0.2 mg/mL Proteinase K) and heating for 10 min at 65˚C, followed by 15 min at 95˚C. The target region was amplified using NEBNext High-Fidelity 2× PCR Master Mix (New England BioLabs) (P5 primer: CTCTTTCCCTACACGACGCTCTTCCGATCTATGACCTTAGGCTTATTTTAACTTAATC, P7 primer: CTGGAGTTCAGACGTGTGCTCTTCCGATCTGAGTCTCATGCACTGTTTTGC) and sequenced on an Illumina MiSeq. Sequencing was analyzed using CRISPResso2[162]. The best selected clones resulted in over 90% indel generation upon doxycycline induction and less than 5% cutting when uninduced.

NLGN1-expressing HEK293 cell line generation

The HA-tagged NLGN1 expression construct (pLV-SV40-puro-EF1a-HA-NLGN1-bc, Addgene #200167) was built via multi-step cloning. First, the backbone vector (pLV-SV40-puro-EF1a-HA-ORF-bc) was created by subcloning the SV40 promoter from pHR_dSV40-H2B-GFP (Addgene #67928), and the EF1a promoter and puromycin resistance cassette from LentiGuide-

BC-EF1a-Puro (Addgene #127170). To ensure proper membrane localization and tagging of cell adhesion molecules expressed via this vector, the N-terminal NLGN1 signal sequence was amplified from NM_014932.3 (GenScript) to include an HA-tag sequence directly following the signal sequence. Using the Gibson assembly protocol described above, these DNA inserts were then ligated into the lentiviral backbone modified from LentiGuide-BC-start (Addgene #127168). Next, a barcode was added to the 3' UTR of the open reading frame (ORF) via Golden Gate assembly with BpiI (Thermo Fisher Scientific), enabling the identification of the open reading frame (ORF) via *in situ* sequencing. Finally, the rest of the *NLGN1* open reading frame was amplified from NM_014932.3 (GenScript) and inserted into the barcoded expression vector using Gibson assembly.

Following lentivirus packaging and harvest as described, the virus was concentrated by centrifugation with a 100 kDa Amicon filter (Sigma-Aldrich) at 2500 g and 4˚C for 30 min. The best HEK293 clone with inducible Cas9 was then spinfected with the concentrated lentivirus, according to the above protocol, and selected with 1 µg/mL puromycin for 3 days. The resulting cell line was used in subsequent artificial formation assays, including the primary and secondary screens.

sgRNA library design and cloning

The gene set for the primary screen was defined based on the list of synaptic genes from the SynGO database[95]. These genes were cross-referenced with the Human Protein Atlas database to include only those that have been shown to be expressed at a protein level in HEK293 cells, resulting in a list of 731 synaptic genes endogenously expressed in HEK293[96]. Next, genes regulating general translational processes (including translation initiation factors and ribosomal proteins) were excluded, as their knockouts likely would not have effects specific to synapse formation. The final list of 644 genes was then split into 82 lowly expressed genes and 562 highly expressed genes, based on published expression data in HEK293, using a normalized RNA expression cutoff of 2[96].

sgRNA sequences for CRISPR-mediated knockout were selected from published libraries, optimizing for on-target efficiency with low off-target effects, minimizing the length of the *in situ* sequencing read from the 5′ end required to demultiplex all sgRNAs, and setting the minimum Levenshtein distance of 2 between the minimum 5′ sequencing read for all possible pairs of sgRNAs[163–165]. In total, 2,626 sgRNA sequences were selected for the primary screen, including 4 sgRNAs each for all genes and 50 non-targeting control sgRNAs. These were designed as three separate subpools – one targeting the highly expressed genes, one targeting the lowly expressed genes, and one containing the non-targeting sgRNAs. The sgRNA library for the secondary screen included 140 sgRNAs targeting 35 genes with top normalized enrichment scores from the primary screen and 10 non-targeting control sgRNAs.

The subpools of synthesized oligo arrays were ordered from Agilent. The sgRNA sequences in each subpool were amplified with Herculase II Fusion DNA Polymerase (Agilent) and cloned into CROPseq-Guide-Zeo (Addgene #127173) via Golden Gate assembly with Esp3I

(Thermo Fisher Scientific), as described above. Plasmid libraries were transformed in electrocompetent cells (Lucigen Endura) and grown in LB medium with 100 µg/mL ampicillin for 18 hours at 30˚C. Plasmid DNA was extracted using the QIAGEN Plasmid Plus Midi Kit. The plasmid libraries were then verified by PCR (P5 primer: ACACGACGCTCTTCCGATCTtcttgtggaaaggacgaaac, P7 primer: CTGGAGTTCAGACGTGTGCTCTTCCGATCaagcaccgactcggtgccac) and sequencing on an Illumina MiSeq.

Pooled knockout cell library generation

Prior to lentivirus production, sgRNA plasmid libraries were combined to a final pool fraction of 82.9% highly expressed targets, 12.1% lowly expressed targets, and 5% non-targeting controls for the primary screen. For the secondary screen, the targeting and non-targeting plasmid libraries were mixed to a 9:1 ratio.

Lentivirus containing this pooled sgRNA library was packaged and titered as described above. HEK293 cells with stably integrated inducible Cas9 and HA-tagged NLGN1 were spinfected with the lentivirus and selected with 300 µg/mL zeocin (Thermo Fisher Scientific) for 5 days. Cas9 expression was then induced with 1 µg/mL doxycycline for 7 days.

Fluorescence microscopy

Screening datasets were acquired as previously described, using a Nikon Ti-2 inverted epifluorescence microscope with an Iris 9 sCMOS camera (Teledyne Photometrics) and a CELESTA light engine (Lumencor) for fluorescence illumination[23,26]. Hardware was controlled using NIS-Elements AR.

Primary and secondary screen phenotype images, as well as arrayed artificial synapse formation assays, were acquired with a 20X 0.75 NA CFI Plan Apo Lambda objective (Nikon MRD00205), using DAPI (408 nm laser excitation, custom Chroma dual-band 408/473 dichroic and emission filter set, 2×2 pixel binning), Alexa Fluor 488 (473 nm laser excitation, Chroma T495LPXR dichroic filter, Chroma ET530/30 emission filter, 2×2 pixel binning), Alexa Fluor 532 (518 nm laser excitation, Chroma T555LPXR dichroic filter, Chroma ET575/30 emission filter, 1×1 pixel binning), Alexa Fluor 594 (545 nm laser excitation, Chroma T565LPXR dichroic filter, Semrock FF01-615/24 emission filter, 1×1 pixel binning), Alexa Fluor 647 (635 nm laser excitation, Chroma ZET635RDC dichroic filter, Semrock FF01-680/42 emission filter, 1×1 pixel binning), and DyLight 755 (750 nm laser excitation, Semrock FF765-Di01 dichroic filter, custom ET820/110 Chroma emission filter, 2×2 pixel binning) fluorescence channels.

In situ sequencing images were acquired using a 10X 0.45 NA CFI Plan Apo Lambda objective (Nikon) with 2×2 pixel binning and the following laser lines, filters, and exposure times: DAPI (408 nm laser excitation at 2% power, custom Chroma dual-band 408/473 dichroic and emission filter set, 20 ms exposure), MiSeq G (545 nm laser at 30% power and Semrock FF01-543/3 excitation filter, Chroma T555LPXR dichroic filter, Chroma ET575/30 emission filter, 200 ms exposure), MiSeq T (545 nm laser excitation at 30% power, Chroma T565LPXR dichroic filter, Semrock FF01-615/24 emission filter, 200 ms exposure), MiSeq A (635 nm laser excitation at 30% power, Chroma ZET635RDC dichroic filter, Semrock FF01-680/42 emission filter, 200 ms), and MiSeq C (635 nm laser excitation at 30% power, Chroma ZET635RDC dichroic filter, Semrock FF01-732/68 emission filter, 200 ms exposure).

Images of knockdown primary neurons were taken with a 40X 0.95 NA CFI Plan Apo Lambda objective (Nikon) on the same Nikon Ti-2 inverted epifluorescence microscope or an Andor Dragonfly 200 spinning disk confocal microscope with a Zyla sCMOS camera (Andor).

Pooled artificial synapse formation assay

Hippocampal neurons were seeded onto PDL-coated 6-well glass-bottom plates (Cellvis) at a density of 36,000 cells/cm^2 and cultured as described. On DIV 8, knockout HEK293 cells were washed in neuronal media and seeded atop these neurons at a density of 6,250 cells/cm^2. After 24 hours, the co-cultures were fixed with 4% paraformaldehyde in PBS for 30 min. Next, cells were permeabilized with 70% ethanol for 30 min and exchanged into PBS with 0.05% Tween-20 (PBS-T) over 6 washes. All subsequent antibodies were diluted in PBS with 3% bovine serum albumin (BSA) (VWR) and 0.27 U/µL Ribolock RNase inhibitor (Thermo Fisher Scientific). Samples were stained with chicken anti-MAP2 antibody (Thermo Fisher Scientific, Cat#PA110005, RRID: AB_1076848; 1:800 dilution), mouse anti-Bassoon antibody (Enzo Life Sciences, Cat#SAP7F407, RRID: AB_2313990; 1:500 dilution), rabbit anti-VGAT antibody (Synaptic Systems, Cat#131 003, RRID: AB_887869; 1:600 dilution), and guinea pig anti-VGLUT1 antibody (Sigma-Aldrich, Cat#AB5905, RRID: AB_2301751; 1:1,000 dilution) for 1 h at room temperature, and washed three times with PBS-T. Cells were then stained with goat anti-chicken IgY DyLight 755 (Thermo Fisher Scientific, Cat#SA5-10075, RRID: AB_2556655; 1:500 dilution), donkey anti-mouse IgG Alexa Fluor 594 (Thermo Fisher Scientific, Cat#A-21203, RRID: AB_141633; 1:1,000 dilution), goat anti-rabbit IgG Alexa Fluor 532 (Thermo Fisher Scientific, Cat#A-11009, RRID: AB_2534076; 1:1,000 dilution), and goat anti-guinea pig IgG Alexa Fluor 647 (Thermo Fisher Scientific, Cat#A-21450, RRID: AB_141882; 1:10,000 dilution) for 30 min at room temperature, and washed three times with PBS-T. Finally, cells were stained with mouse anti-HA antibody conjugated to Alexa Fluor 488 (Cell Signaling Technologies, Cat#2350, RRID: AB_491023; 1:800 dilution) for 1.5 h at room temperature and washed three times with PBS-T. Samples were then exchanged to 100 ng/mL DAPI and 0.4 U/µL Ribolock RNase inhibitor in 2× SSC (Thermo Fisher Scientific), and imaged with 3 z-slices at 1 µm intervals using the microscope configuration described above.

In situ sequencing of sgRNAs in pooled screen

Following imaging of synaptic phenotypes, samples underwent *in situ* amplification of sgRNA sequences as previously described[23,26]. Briefly, transcripts were reverse-transcribed *in situ* by incubating at 37˚C overnight in 1× RevertAid RT buffer (Thermo Fisher Scientific), 250 µM dNTPs (New England BioLabs), 0.2 mg/mL BSA (New England BioLabs), 1 µM LNA-modified RT primer (oRT_CROPseq, G+AC+TA+GC+CT+TA+TT+TTAACTTGCTAT, Integrated DNA Technologies), 0.8 U/µL Ribolock RNase inhibitor (Thermo Fisher Scientific), and 4.8 U/µL RevertAid H minus reverse transcriptase (Thermo Fisher Scientific). Cells were then washed 5 times with PBS-T and post-fixed with 3% paraformaldehyde and 0.1% glutaraldehyde for 30 min at room temperature. After washing 5 times with PBS-T, samples were incubated in a gap-filling reaction mix for 5 min at 37˚C and 1.5 h at 45˚C; the mix contained 1× Ampligase buffer (Lucigen), 0.4 U/µL RNase H (QIAGEN), 0.2 mg/mL BSA, 100 nM padlock probe (oPD_CROPseq, /5Phos/gttttagagctagaaatagcaagCTCCTGTTCGACACCTACCCACCTCATCCCACTCTTCAaaaggacgaaacaccg, Integrated DNA Technologies), 0.02 U/µL TaqIT polymerase (QIAGEN), 0.5 U/µL Ampligase (Lucigen), and 50 nM dNTPs. Cells were washed with PBS-T 3 times. The circularized padlock probes then served as templates for rolling circle amplification; samples were incubated at 30˚C overnight in 1× Phi29 buffer (Thermo Fisher Scientific), 250 µM dNTPs, 0.2 mg/mL BSA, 5% glycerol, and 1 U/µL Phi29 DNA polymerase (Thermo Fisher Scientific), and washed with PBS-T 3 times.

Finally, amplified sgRNA sequences were read out as previously described[23,26]. A sequencing primer (oSBS_CROPseq, CACCTCATCCCACTCTTCAaaaggacgaaacaccg, Integrated DNA Technologies), diluted to 1 µM in 2× SSC with 10% formamide, was hybridized for 30 min at room temperature. Next, *in situ* sequencing by synthesis was performed using reagents from the Illumina MiSeq 500-cycle Nano kit (Illumina MS-103-1003) . After washing with incorporation buffer (Nano kit PR2), cells were incubated in incorporation mix (Nano kit reagent 1) at 60˚C for 3 min on a flat-top thermocycler, then quickly washed 6 times with PR2, followed by 3-4 PR2 washes at 60˚C for 10 min. Samples were then exchanged to 100 ng/mL DAPI in 2× SSC and imaged using the microscope configuration described above. Following imaging, samples were incubated in cleavage mix (Nano kit reagent 4) for 6 min at 60˚C, and washed with PR2. The incorporation protocol was then repeated for the subsequent round of sequencing. In total, 9 and 8 cycles of *in situ* sequencing were performed for the primary and secondary screen, respectively.

Optical pooled screen image analysis

Phenotype images were first maximum intensity z-projected. Nuclei were detected by applying local intensity and area thresholding and watershed-based segmentation on the DAPI channel. HEK293 cells were then segmented by local thresholding on the Alexa Fluor 488 channel (HA-NLGN1), additional morphological operations to remove aberrant particles and holes, and watershed-based segmentation with the nuclear segmentations as seeds. Alexa Fluor 532 (VGAT) and Alexa Fluor 647 (VGLUT1) channels were more finely aligned using cross-correlation of each signal with that of Alexa Fluor 594 (Bassoon). Bassoon, VGAT, and VGLUT1 puncta were identified by first applying a Laplacian-of-Gaussian linear filter (kernel width $\sigma = 1$ pixel) on the corresponding channels, followed by performing local thresholding, morphological operations, and watershed segmentation. Segmented puncta were assigned to cells by comparing the puncta and cell masks. Feature extraction was performed by implementing functions derived from CellProfiler and scikit-image[166,167]. Puncta-level features were extracted from puncta segmentations for each channel, with images in the DAPI, Alexa Fluor 488, and DyLight 755 images zoomed using spline interpolation to match dimensions. Cell-level features were extracted from cellular segmentations for each channel, with Alexa Fluor 532, Alexa Fluor 594, and Alexa Fluor 647 images downsampled to 2×2 pixel binning.

In situ sequencing images were analyzed as previously described[23,26]. Nuclei were segmented by applying local thresholding and watershed-based segmentation on the DAPI channel. Cells were then detected using local thresholding of the cytoplasmic background in the sequencing channels (MiSeq G, T, A, and C) and segmented with a watershed algorithm. Sequencing reads were coarsely aligned during acquisition by calibrating the plate position and more finely aligned computationally using cross-correlation of the sequencing signal between cycles. Reads were then detected by applying a Laplacian-of-Gaussian linear filter (kernel width $\sigma = 1$ pixel) on the sequencing channels and finding the local maxima of the per-pixel, per-channel standard deviation over sequencing cycles. Base intensities were defined as the maximum value in a 3×3 window centered at the read. To correct for signal cross-talk and intensity differences between channels, a linear transformation was estimated and applied to the extracted sequencing spot intensities. Finally, bases were called based on the channel with the maximum corrected intensity in a given sequencing cycle. Homopolymers were removed to filter debris miscalled as sequencing reads.

Phenotype and *in situ* sequencing images were roughly aligned during acquisition using nuclear masks, but required further computational alignment, as the images were acquired at different magnifications. The Delaunay triangulations of nuclei centroids of each image were calculated and compared between the phenotyping and sequencing datasets to find matching fields-of-view and cell identities. Cells with no matched identity between phenotyping and sequencing images, no sequencing reads mapping to a designed sgRNA sequence, or <50% sequence agreement between reads within the same cell were excluded from further analysis.

Functions for image segmentation, feature extraction, and *in situ* sequencing analysis were written in Python and performed in parallel using the Snakemake workflow manager[168].

Processing and analysis of screen phenotypes

Segmented puncta were filtered using an intensity threshold in the DyLight 755 channel (MAP2) to limit downstream analysis to artificial synapses between HEK293 cells and neurons. Next, puncta-level feature scores were aggregated as medians or sums across puncta within a given cell. Missing values for cells containing no segmented puncta were imputed with zero. Features with skewed distributions were transformed to approximate normal distributions. To remove batch effects between wells and plates, all features for every cell were then normalized as median robust z-scores, using the median and median absolute deviation of the cells with non-targeting sgRNAs within the same well or row of the same plate (Supplementary Figure 2.1F). The synaptogenesis score was calculated for each cell by adding the median robust z-scores of 6 features: the total area of puncta normalized to the cell area ("normalized area") and the total number of puncta per cell ("puncta count") for Bassoon, VGAT, and VGLUT1.

Summary phenotype scores for each sgRNA were computed as the difference in the area under the curve (AUC) between the cumulative distribution function of a single feature of cells with non-targeting sgRNAs and that of cells containing a given sgRNA (ΔAUC). The aggregate score of each gene was then calculated by taking the mean ΔAUC across all sgRNAs targeting the same gene. To create non-targeting "genes" with more evenly matched numbers of cells as targeting genes, 6 non-targeting sgRNAs were randomly sampled without replacement 50 times for the primary screen; for the secondary screen, 10 random subsets of 2 non-targeting sgRNAs were chosen. Raw p-values for a subset of parameters were computed by a permutation test, comparing the gene-level scores to the corresponding null distributions bootstrapped from non-targeting controls. Null distributions of ΔAUC scores were first defined for each sgRNA by performing 10,000 bootstrap repetitions in which the same number of cells containing the targeting sgRNA was sampled from the population of the non-targeting control cells. To define the gene-level null distributions, ΔAUC scores were sampled 10,000 times from each of the corresponding null distributions of sgRNAs targeting the same gene and averaged. The Benjamini-Hochberg procedure was applied to the raw p-values to obtain the FDR q-values. An FDR threshold was used to determine significance, as described in the figure captions.

The normalized enrichment scores (NES) for each gene were calculated from a running-sum statistic of ranked sgRNA ΔAUC scores, similarly to the scoring method used in GSEA[169]. Briefly, all sgRNAs are ranked by their ΔAUC scores. The running-sum metric was updated moving down the rank, increasing if the sgRNA targets a given gene and decreasing otherwise. The enrichment score (ES) was defined as the maximum of the sum. Nominal p-values were obtained from a permutation test in which null distributions were produced from 10,000 permutations of sgRNA labels. The NES was computed by dividing the ES by the expected value of the corresponding null distribution. The Benjamini-Hochberg procedure was applied to the nominal p-values to obtain the FDR q-values, and an FDR threshold of 0.25 was used to determine significance.

For high-dimensional analysis, features were filtered by manual selection and iterative exclusion of those with a Pearson correlation greater than 0.9 with another feature. This resulted in a set of 95 features out of the 263 extracted features. Principal component analysis (PCA) was performed on these 95 features to identify ones that separated positive from negative controls and define an aggregate synaptogenesis score. The PHATE algorithm (default parameters except n_pca=None) was used on the filtered features to visualize a two-dimensional representation of the phenotypic profiles of the gene targets[97]. To cluster gene profiles, the diffusion operator local affinity graph produced from the PHATE algorithm was inputted into the Leiden algorithm (resolution parameter = 0.9)[98]. The Leiden resolution parameter was optimized by analyzing the cluster modularity and number of clusters detected from 20 repetitions at a range of parameters (Supplementary Figure 2.2A), as well as the robustness of clustering solutions obtained from 20 repetitions in which 90% of the genes were subsampled and clustered (Supplementary Figure 2.2B).

Generation of clonal knockout HEK293 cells for arrayed artificial synapse formation assay

sgRNA sequences for individual knockouts were selected from published libraries based on on-target and off-target performance[163–165]. These targeting sgRNAs and non-targeting controls were cloned into CROPseq-Guide-Zeo (Addgene #127173) using Golden Gate assembly, and stably integrated into HEK293 cells with inducible Cas9 and HA-tagged NLGN1, as described above. After lentivirus transduction, cells were selected with 300 µg/mL zeocin for 5 days. Single cells were then isolated using an SH800 Cell Sorter (Sony). Expanded clones were screened after 7 days of 1 µg/mL doxycycline treatment to induce Cas9 expression. To verify knockout, genomic DNA was extracted by resuspending cells in lysis buffer (10 mM Tris pH 7.5, 1 mM $CaCl_2$, 3 mM $MgCl_2$, 1 mM EDTA, 1% Triton X-100, and 0.2 mg/mL Proteinase K) and heating for 10 min at 65˚C, followed by 15 min at 95˚C. The target regions were amplified using NEBNext High-Fidelity 2× PCR Master Mix (New England BioLabs) with appropriate P5 and P7 primers and sequenced on an Illumina MiSeq. Sequencing was analyzed using CRISPResso2. For each sgRNA, clones with 100% frameshift rates were selected. For knockouts of *PFN1* and *SPFQ*, the best clones had 50-70% frameshift rates.

Clonal knockout cell lines were used for arrayed artificial synapse formation assays, following a protocol similar to the pooled assay. Briefly, hippocampal neurons were seeded onto PDL-coated 24-well glass-bottom plates (Greiner Bio-One) at a density of 36,000 cells/cm^2 and cultured as described. On DIV 8, knockout HEK293 cells were washed in neuronal media and seeded atop these neurons at a density of 6,250 cells/cm^2. After 24 hours, the co-cultures were fixed with 4% paraformaldehyde in PBS for 30 min and permeabilized with 70% ethanol for 30 min. All subsequent antibodies were diluted in PBS with 3% BSA. Samples were stained with chicken anti-MAP2 antibody (Thermo Fisher Scientific, 1:800 dilution), mouse anti-Bassoon antibody (Enzo Life Sciences, 1:500 dilution), rabbit anti-VGAT antibody (Synaptic Systems, 1:600 dilution), and guinea pig anti-VGLUT1 antibody (Sigma-Aldrich, 1:1,000 dilution) for 1 h at room temperature. Co-cultures were then incubated with goat anti-chicken IgY DyLight 755 (Thermo Fisher Scientific, 1:500 dilution), donkey anti-mouse IgG Alexa Fluor 594 (Thermo Fisher Scientific, 1:1,000 dilution), goat anti-rabbit IgG Alexa Fluor 532 (Thermo Fisher Scientific, 1:1,000 dilution), and goat anti-guinea pig IgG Alexa Fluor 647 (Thermo Fisher Scientific, 1:10,000 dilution) for 30 min at room temperature. Next, cells were stained with rabbit anti-HA antibody conjugated to Alexa Fluor 488 for 1.5 h at room temperature. Samples were then exchanged to 100 ng/mL DAPI in 2× SSC and imaged with 3 z-slices at 1 µm intervals using the microscope configuration described above.

Expression of human NLGN1 in primary neurons

To enable visualization of NLGN1 expression and localization without overexpression, a construct was created to concurrently knock down endogenous rat *Nlgn1* in primary neurons and express human NLGN1 tagged with GFP. The vector for rat *Nlgn1* knockdown and human NLGN1 expression (FSW-rNLGN1-shRNA-hSyn-GFP-hNLGN1, Addgene #200168) was built via multi-step cloning. First, a construct was created via Gibson cloning using the backbone of FSW-shLR1+GFP-LRRTM1 (Biederer Lab, Yale University) and the U6 promoter with an shRNA insertion site taken from pSpCas9(BB)-2A-Puro (PX459) V2.0 (Addgene #62988). Next, oligonucleotides of a published *Nlgn1*-targeting shRNA sequence (5′–AGACCTTCACTCGAACTTTCTCGAGAAAGTTCGAGTGAAGGTCTGCCCCTTTTT–3′) were ordered from Integrated DNA Technologies and cloned into the backbone using Golden Gate assembly with Esp3I (Thermo Fisher Scientific)[170]. Finally, the human *NLGN1* open reading frame was amplified from pLV-SV40-puro-EF1a-HA-NLGN1-bc, with the HA-tag following the signal sequence replaced with the GFP sequence taken from FSW-shLR1+GFP-LRRTM1. The vector was transfected into neurons, as described below.

shRNA design and cloning for expression in primary neurons

shRNA sequences for individual knockdowns were chosen using the shRNA design tool from the Broad Institute's Genetic Perturbation Platform. shRNA oligonucleotides were synthesized (Integrated DNA Technologies) with a U6 termination sequence and cloned into pLV-U6-shRNA-BsmBI-EF1a-puroR-BFP (Liu Lab, Broad Institute) via Golden Gate assembly with Esp3I (Thermo Fisher Scientific), as described. To mediate knockdowns in primary neurons, shRNA expression vectors were transfected, as described below.

Transfection of primary neurons

Cortical neurons were seeded onto PDL-coated 24-well glass-bottom plates (Greiner Bio-One) at a density of 36,000 cells/cm^2. On DIV 3, after the 24-h treatment with 2 μM cytosine β-D-arabinofuranoside, media was replaced with Neurobasal Medium supplemented with B27 and GlutaMAX. On DIV 6, neurons were transfected with 0.75 μL Lipofectamine-3000 (Thermo Fisher Scientific), 1 μL P-3000, and 500 ng total of pFSW-rNLGN1-shRNA-hSyn-GFP-hNLGN1 and an individual shRNA vector at a 1:1 plasmid mass ratio. On DIV 8 and 12, 50% of the media was changed. On DIV 14, imaging assays were performed on the samples.

Live staining of surface glutamate receptors

Cortical neurons were transfected with pFSW-rNLGN1-shRNA-hSyn-GFP-hNLGN1 and shRNAs, following the above protocol. On DIV 14, neurons were washed for 5 min at room temperature with 1% BSA in Tyrode's solution (15 mM D-(+)-Glucose, 108 mM NaCl, 5 mM KCl, 2 mM $MgCl_2 \cdot 6H_2O$, 2 mM $CaCl_2 \cdot 2H_2O$, 25 mM HEPES pH 7.4). Cells were then incubated with mouse anti-GluA antibody (Synaptic Systems, Cat#182 411, RRID: AB_2619876; 1:100 dilution in Tyrode's solution with 1% BSA) for 12 min at room temperature and washed with PBS. Next, samples were fixed with 4% paraformaldehyde for 15 min and permeabilized with 0.1% Triton-X 100 for 10 min. Following washes with PBS-T, neurons were stained with chicken anti-Bassoon antibody (Synaptic Systems, Cat#141 016, RRID: AB_2661779; 1:500 dilution) for 1 h at room temperature and washed with PBS-T. Samples were then stained with donkey anti-mouse IgG Alexa Fluor 594 (Thermo Fisher Scientific, Cat#A-21203, RRID: AB_141633; 1:500 dilution) and goat anti-chicken IgY DyLight 755 (Thermo Fisher Scientific, Cat#SA5-10075, RRID: AB_2556655; 1:500 dilution) for 30 min at room temperature. All antibodies after fixation were diluted in PBS with 3% BSA. Finally, cells were placed in 2× SSC and imaged with 5 z-slices at 0.5 µm intervals using the microscope configuration described above. Sections of dendrites were maximum intensity projected and cropped in Fiji[171]. Images were analyzed using a custom pipeline to segment puncta and quantify mean intensities.

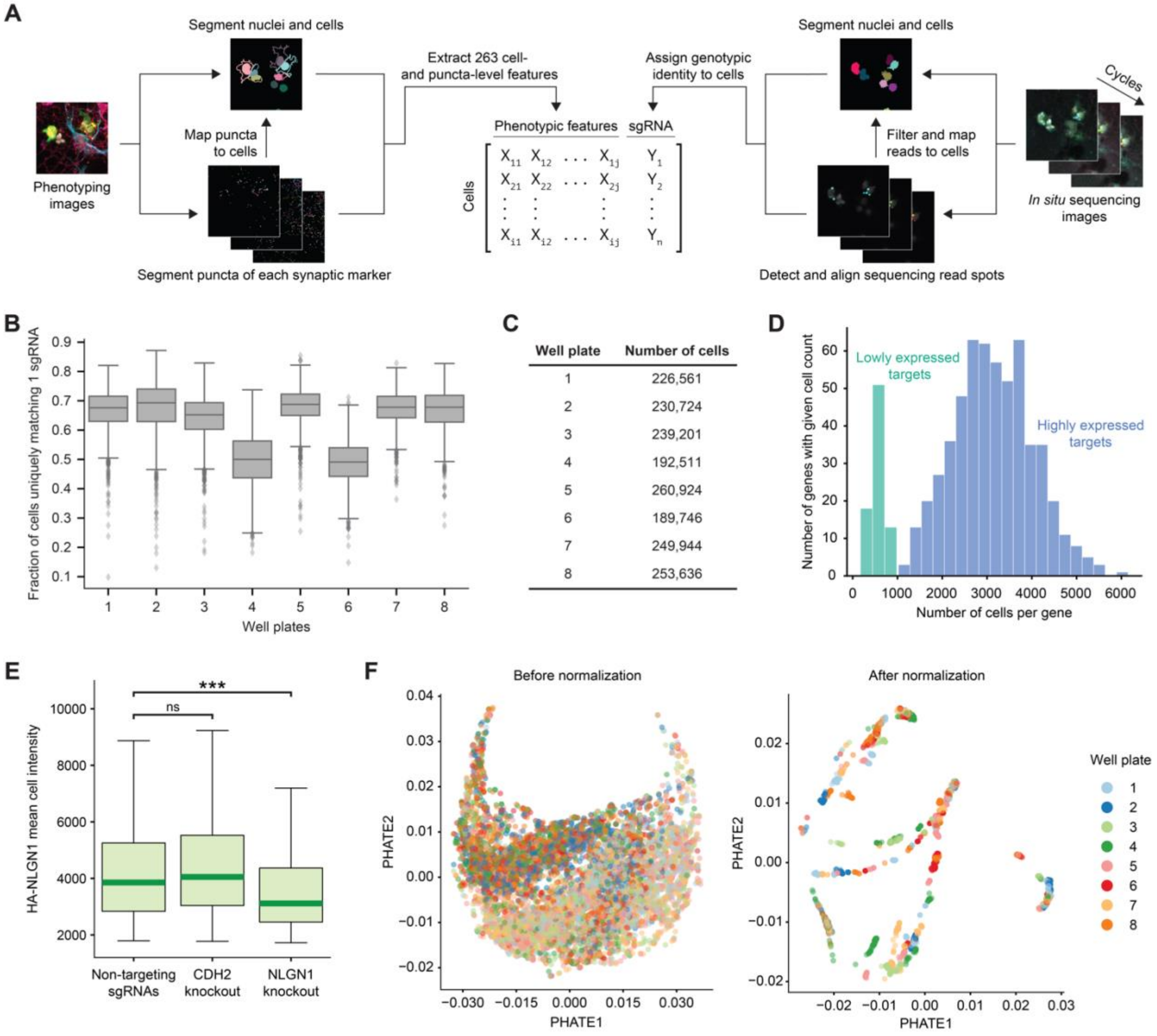
Synaptotagmin-1 antibody uptake

Cortical neurons were transfected with pFSW-rNLGN1-shRNA-hSyn-GFP-hNLGN1 and shRNAs as described. On DIV 14, well plates were placed on a heated pad at 37˚C in a tissue culture hood. Cells were carefully washed twice with warmed modified Tyrode's solution (10 mM D-(+)-Glucose, 150 mM NaCl, 4 mM KCl, 2 mM $MgCl_2$, 20 mM HEPES pH 7.4), leaving half the volumes in the wells. Meanwhile, mouse anti-Synaptotagmin-1 antibody (Synaptic Systems, Cat#105 221; RRID: AB_887834) was diluted 1:200 in loading solution, consisting of high K^+ Tyrode's solution (10 mM D-(+)-Glucose, 64 mM NaCl, 90 mM KCl, 2 mM $MgCl_2$, 2 mM $CaCl_2$, 20 mM HEPES pH 7.4), 100 µM DL-2-Amino-5-phosphonopentanoic acid (AP-5) (Sigma-Aldrich), 20 µM CNQX (Sigma-Aldrich), and additional $CaCl_2$ to a final concentration of 4 mM. This antibody solution was added to the washed wells at a final antibody concentration of 1:400 and incubated for 3 min at 37˚C. Samples were then washed 3 times with warmed modified Tyrode's solution containing 50 µM AP-5 and 10 µM CNQX and once with PBS. Cells were then fixed with 4% paraformaldehyde for 15 min and permeabilized with 0.1% Triton-X 100 for 10 min. Neurons were stained with rabbit anti-Synaptophysin antibody (Thermo Fisher Scientific, Cat#MA5-14532, RRID: AB_10983675; 1:500 dilution) for 1 h at room temperature and washed with PBS-T. Samples were then stained with donkey anti-mouse IgG Alexa Fluor 594 (Thermo Fisher Scientific, Cat#A-21203, RRID: AB_141633; 1:1000 dilution) and goat anti-rabbit IgG Alexa Fluor Plus 647 (Thermo Fisher Scientific, Cat#A-32733, RRID: AB_2633282; 1:1000 dilution). All antibodies after fixation were diluted in PBS with 3% BSA. Cells were placed in 2× SSC and imaged with 5 z-slices at 0.5 µm intervals using the microscope configuration described above. Sections of dendrites were maximum intensity projected and cropped in Fiji[171]. Images were analyzed using a custom pipeline to segment puncta and quantify mean intensities.

Materials, Data, and Software Availability

Plasmids generated in this study will be made available from Addgene. Processed data from the screen is available in the supplemental materials. All single-cell extracted phenotype measurements and raw image data will be made available on Google Cloud at gs://opspublic-east1/SynaptogenesisOpticalPooledScreen. The code used for processing optical pooled screening data is available at https://github.com/feldman4/OpticalPooledScreens and https://github.com/annale-8/OpticalPooledScreens.

2.5 Supplementary Figures



$$\begin{array}{c|cccc|c} & \multicolumn{4}{c|}{\text{Phenotypic features}} & \text{sgRNA} \\ \hline \text{Cells} & X_{11} & X_{12} & \cdots & X_{1j} & Y_1 \\ & X_{21} & X_{22} & \cdots & X_{2j} & Y_2 \\ & \vdots & \vdots & & \vdots & \vdots \\ & X_{i1} & X_{i2} & \cdots & X_{ij} & Y_n \end{array}$$

Well plate	Number of cells
1	226,561
2	230,724
3	239,201
4	192,511
5	260,924
6	189,746
7	249,944
8	253,636

Supplementary Figure 2.1 | Validation of optical screening approach for synaptogenesis.

(A) Overview of image analysis pipeline. For each segmented HEK293 cell, cell- and synaptic puncta-level features were extracted from the phenotyping images, and their sgRNA identities were obtained from the *in situ* sequencing images.

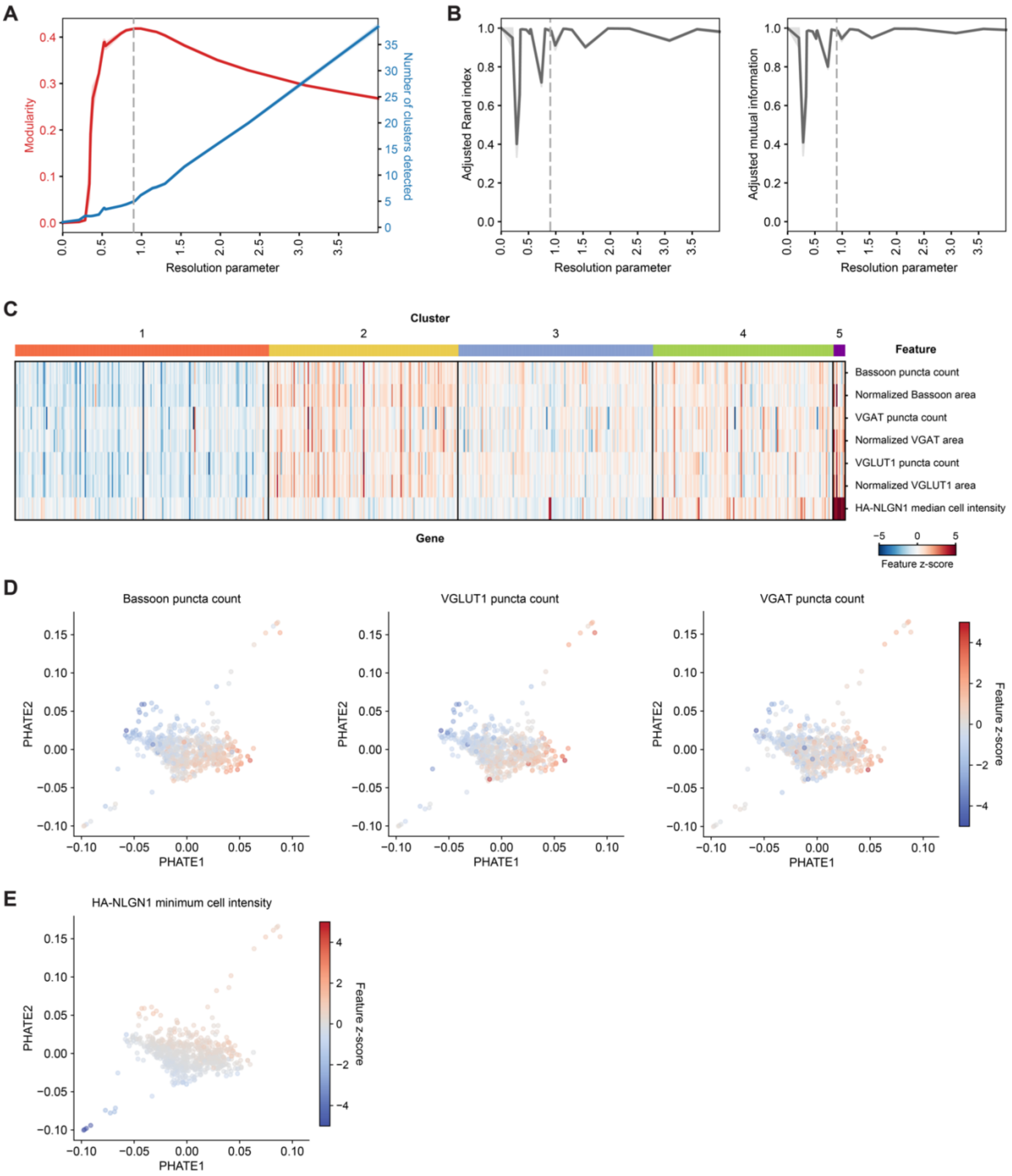
(B) Box plot of the fraction of cells within a field-of-view in each well plate mapping uniquely to 1 sgRNA by *in situ* sequencing. N = 1,281 fields-of-view per plate.

(C) Number of cells, with a mapped phenotyping image and a sgRNA sequence, obtained from each well plate.

(D) Histogram of the number of cells analyzed for both lowly (median = 545 cells) and highly expressed (median = 3,138 cells) gene targets.

(E) Box plot of HA-NLGN1 mean cell intensity in one well plate (non-targeting control, n = 3,715 cells; *CDH2* knockout, n = 510 cells; *NLGN1* knockout, n = 236 cells). Statistical significance measured by two-tailed Mann-Whitney U test comparing each condition to the non-targeting control (ns = $p>0.01$, ***$p<0.0001$).

(F) Two-dimensional PHATE visualization of single-cell phenotype profiles (random sample of 10,000 non-targeting control cells shown) before (left) and after (right) normalization. Colors indicate the well plate out of 8 plates used in the screen.



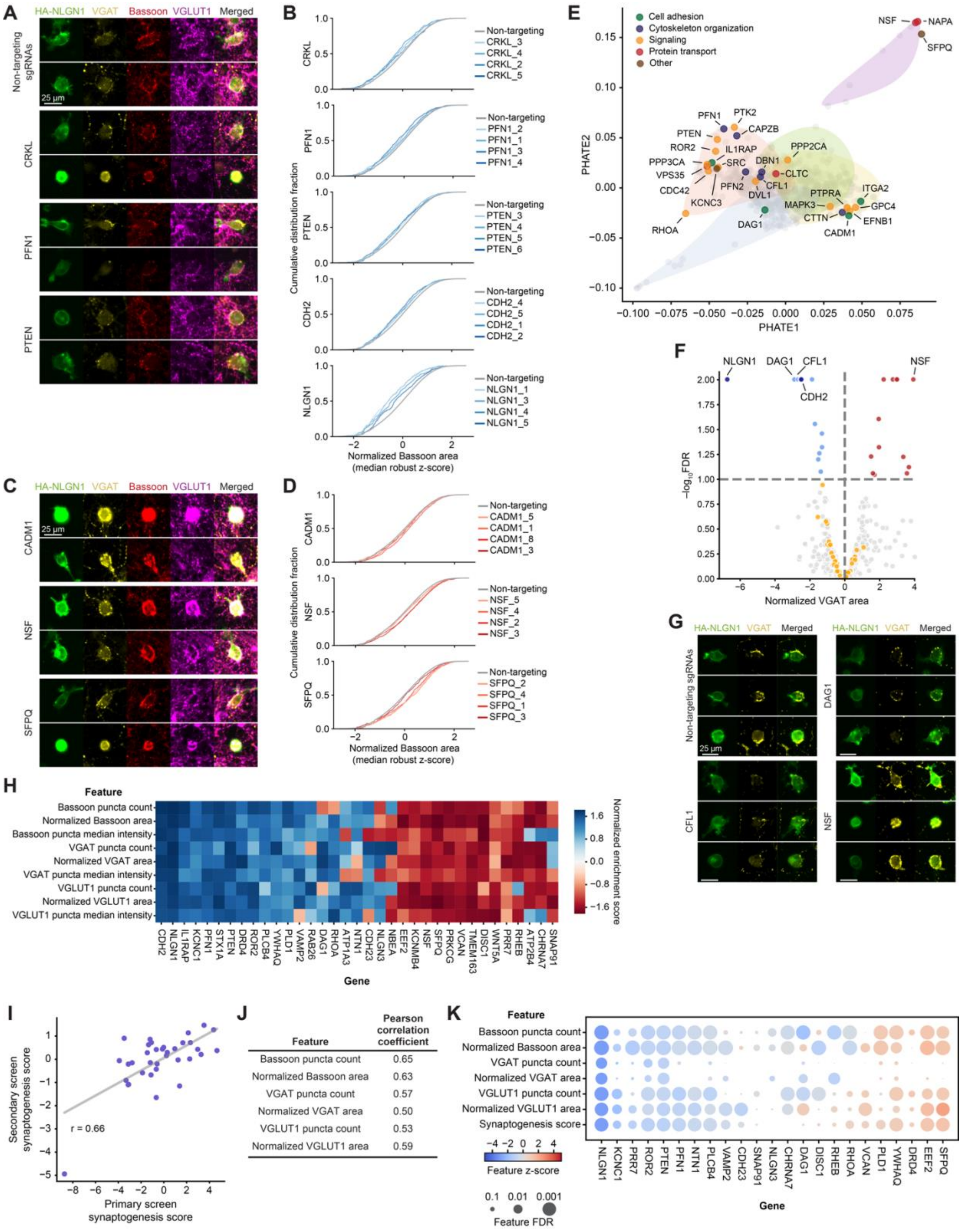
Supplementary Figure 2.2 | PHATE analysis and clustering of synaptic phenotypes.

(A) Line plots of the modularity and the number of detected clusters by Leiden clustering at a range of resolution parameters (means ± SEM from 20 clustering repetitions). Dashed line indicates the selected resolution parameter, 0.9.

(B) Line plots show performance of Leiden clustering at different resolution parameters. In each iteration, the cluster solution from a subset of genes (90% of all gene targets, sampled without replacement) was compared to the full solution using the adjusted Rand index (left) and adjusted mutual information (right). Means ± SEM are shown from 20 repetitions at each resolution parameter. Dashed lines indicate the selected resolution parameter, 0.9.

(C) Heat map of NLGN1 median cell intensity and synaptogenesis score component features for all gene knockouts, corresponding to clusters in Figure 2.2E.

(D-E) Two-dimensional PHATE representation of gene targets, colored by (D) Bassoon, VGLUT1, and VGAT puncta count scores and (E) NLGN1 minimum cell intensity.



J

Feature	Pearson correlation coefficient
Bassoon puncta count	0.65
Normalized Bassoon area	0.63
VGAT puncta count	0.57
Normalized VGAT area	0.50
VGLUT1 puncta count	0.53
Normalized VGLUT1 area	0.59

Supplementary Figure 2.3 | Analysis of primary and secondary screens for synaptogenesis.

(A) Additional example images of gene targets shown in Figure 2.3B. Scale bar, 25 μm.

(B) Empirical cumulative distribution functions for the normalized Bassoon puncta area (median robust z-score) of single cells with each individual sgRNA (shades of blue) targeting *CRKL*, *PFN1*, *PTEN*, *CDH2*, and *NLGN1* (from top to bottom), compared to cells with non-targeting control sgRNAs (gray).

(C) Additional example images of gene targets shown in Figure 2.3D. Scale bar, 25 μm.

(D) Empirical cumulative distribution functions for the normalized Bassoon puncta area (median robust z-score) of single cells with each individual sgRNA (shades of red) targeting *CADM1*, *NSF*, and *SFPQ* (from top to bottom), compared to cells with non-targeting control sgRNAs (gray).

(E) PHATE visualization of gene-level phenotypic profiles with a subset of hits from the primary screen. Labeled genes are colored by functional category. Clusters obtained from the Leiden algorithm are visualized by their convex hulls and colored as in Figure 2.2A.

(F) Volcano plot for VGAT puncta area normalized to the cell area (z-scored). Gene knockouts resulting in decreased (blue, positive controls in navy) or increased (red) normalized VGAT area relative to non-targeting controls are highlighted (FDR < 0.1, dashed line). Orange dots represent 50 random samples of 6 non-targeting sgRNAs.

(G) Illustrative images of gene knockouts shown in (F). Scale bar, 25 μm.

(H) Heat map of normalized enrichment scores of genes selected for the secondary screen. Enrichment scores were computed by ranking all sgRNAs in a given feature and calculating the running-sum statistic for each gene, from sgRNAs targeting that gene. Null distributions were obtained from 10,000 permutations of sgRNA labels. Enrichment scores were then normalized by dividing by the expected value of the corresponding null distribution.

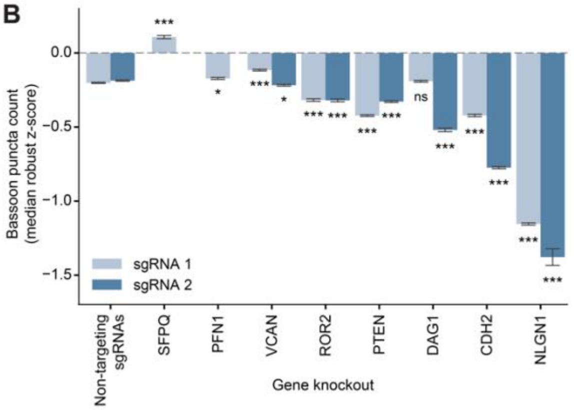
(I) Scatter plot comparing synaptogenesis scores of genes in the primary and secondary screen, Pearson correlation coefficient r = 0.66.

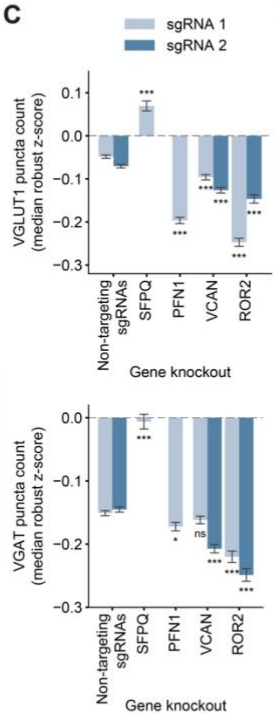
(J) Table of Pearson correlation coefficients between the Bassoon, VGAT, and VGLUT1 puncta count and normalized area scores from the primary and secondary screens.

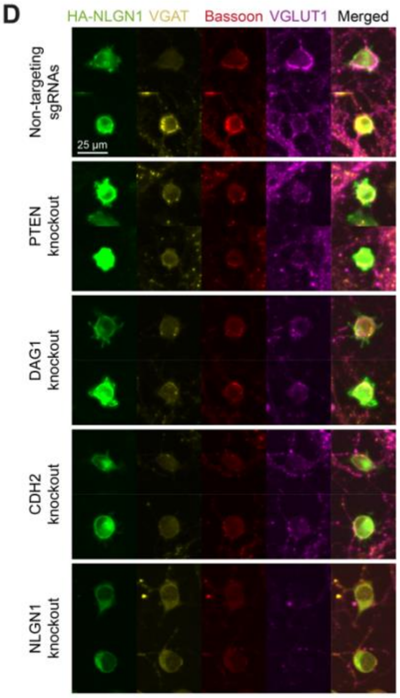
(K) Heat map of synaptogenesis score and its component features for significant genes (FDR < 0.05 for at least two features) from the secondary screen. Color indicates the corresponding feature z-score, and circle size indicates the feature FDR. Gene order was obtained from the synaptogenesis score rank.

A

	Total number of cells analyzed	
Gene	**sgRNA 1**	**sgRNA 2**
Non-targeting sgRNAs	64,328	63,185
CDH2	13,723	19,959
DAG1	30,046	6,410
NLGN1	17,023	520
PFN1	19,250	–
PTEN	29,745	30,093
ROR2	12,964	11,934
SFPQ	8,956	–
VCAN	29,570	22,922



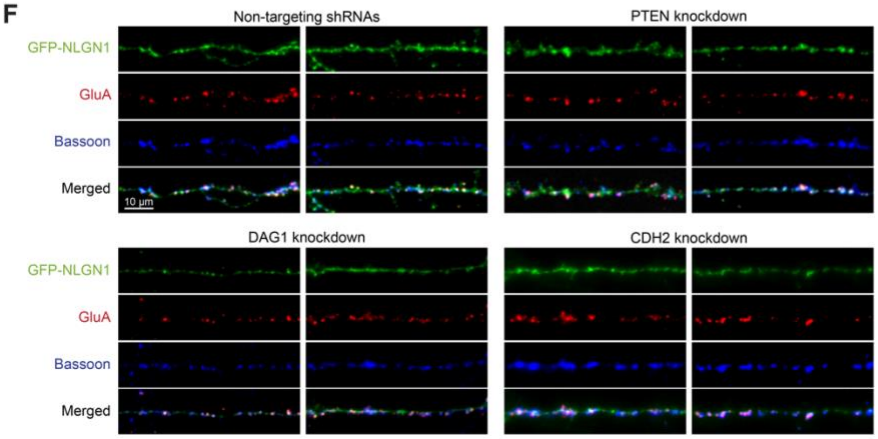




E

GluA labeling

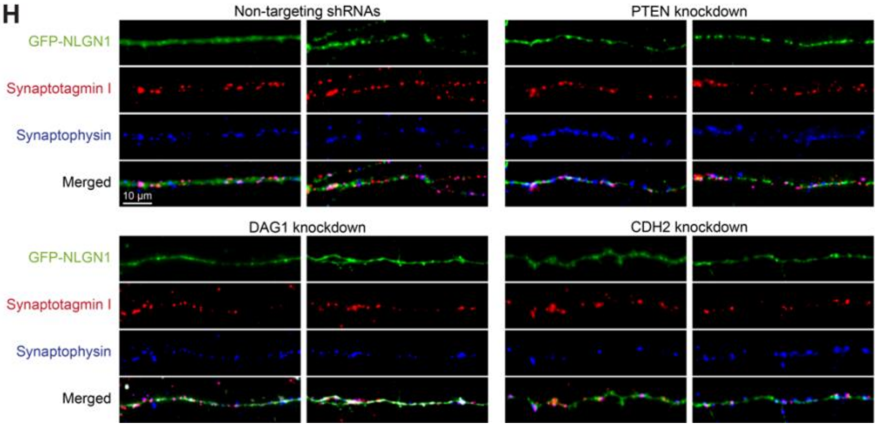
	Number of dendrites (cells)	
Gene	**shRNA 1**	**shRNA 2**
Non-targeting shRNAs	133 (39)	70 (17)
CDH2	51 (14)	112 (32)
DAG1	45 (11)	91 (23)
PTEN	30 (9)	81 (20)



G

Synaptotagmin I antibody uptake

	Number of dendrites (cells)	
Gene	**shRNA 1**	**shRNA 2**
Non-targeting shRNAs	172 (54)	64 (21)
CDH2	75 (25)	123 (41)
DAG1	170 (54)	123 (33)
PTEN	120 (42)	89 (33)



Supplementary Figure 2.4 | Additional images and details on arrayed validation experiments.

(A) The total number of clonal HEK293 cells for each gene knockout analyzed across 2-4 independent co-cultures.

(B) Bar plot of the number of Bassoon puncta per cell (median robust z-score) shows means ± SEM, from 2-4 independent cultures. Statistical significance calculated from one-way ANOVA test, followed by Tukey test for multiple comparisons (ns = p>0.01, *p<0.01, ***p<0001).

(C) Bar plots of the number of VGLUT1 (top) and VGAT (bottom) puncta per cell (median robust z-score) show means ± SEM for additional genes tested in arrayed artificial synapse formation assays, from 2-4 independent co-cultures. Statistical significance calculated from one-way ANOVA test, followed by Tukey test for multiple comparisons (ns = p>0.01, *p<0.01, ***p<0001).

(D) Additional illustrative images of clonal HEK293 knockout cells stained for HA-NLGN1 (green), VGAT (yellow), Bassoon (red), and VGLUT1 (magenta), as shown in Figure 2.4A. Scale bar, 25 µm.

(E) The total number of dendrites and neurons analyzed for surface GluA staining across 2 independent cultures.

(F) Additional example images of cortical neuron knockdowns stained for surface GluA receptors (red) and Bassoon (blue), as shown in Figure 2.4C. Scale bar, 10 µm.

(G) The total number of dendrites and neurons analyzed for synaptotagmin I (SytI) uptake measurements across 3 independent cultures.

(H) Additional example images of cortical neuron knockdowns subjected to live antibody-labeled SytI uptake (red) and fixed staining for synaptophysin (blue), as shown in Figure 2.4F. Scale bar, 10 µm.

Chapter 3. Developing an automated platform for high-throughput *in situ* sequencing

Authors: Anna Le, Tridib Biswas, Bryce Kirby, Avtar Singh, J. Owen Andrews, Nikita Podobedov, Brian Cheng, Sami Farhi, Paul C. Blainey

3.1 Introduction

Optical pooled screening is a powerful and versatile approach to investigate effects of genetic perturbations on complex, image-based cellular phenotypes. Although this technology can be applied to a range of biological systems, its scalability to tens of millions of cells and up to genome-wide libraries has only been demonstrated in simple models, such as cancer cell lines[23,28,29]. Sequencing the cell libraries for such screens can span several weeks, hindering the applicability of optical pooled screening, particularly for larger-scale screens (e.g. with combinatorial perturbations or less penetrant phenotypes requiring higher cell coverage) and screens in models that require sparser plating densities (e.g. cell lines with greater surface areas or exhibiting cell non-autonomous phenotypes, similar to the screens described in Chapter 2).

The bottleneck in sample processing throughput of optical pooled screens largely stems from the intensive time and labor requirements during iterative cycles of *in situ* sequencing by synthesis (SBS). To identify the genetic perturbation in each cell, the perturbation barcode (e.g. an sgRNA sequence) is first reverse transcribed and amplified[23]. The sequence is then read out via rounds of incorporating dye-labeled nucleotides, imaging, and cleaving the dyes. Using reagents extracted from Illumina MiSeq kits, the incorporation, cleavage, and washing steps are performed by manually pipetting into well plates for as many cycles as required to demultiplex all members of the perturbation library (typically up to 12 cycles). Depending on the number of plates needed for a given screen, this repetitive process requires weeks of hands-on time, with over a third of genotyping costs going toward labor alone.

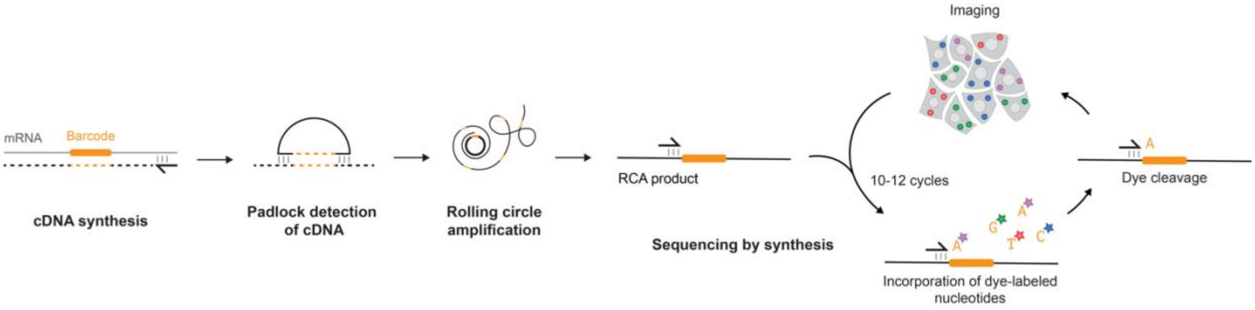


Figure 3.1 | Overview of *in situ* sequencing protocol for optical pooled screens.

The perturbation barcodes, such as sgRNA spacer sequences for CRISPR-based screens, are expressed as polyadenylated mRNA transcripts. After fixation and permeabilization, a locked nucleic acid (LNA)-modified primer is used to reverse transcribe a cDNA copy of the barcode sequence. Following glutaraldehyde and formaldehyde post-fixation, the mRNA is digested and a padlock probe is hybridized to cDNA regions flanking the sgRNA sequence. The padlock probe is then extended and ligated to copy the barcode sequence into a single-stranded circularized DNA. This circularized DNA serves as a template for rolling circle amplification with Phi29 polymerase, the amplified product of which contains tandem repeats of the barcode. This sequence is then read out by *in situ* sequencing by synthesis (SBS), consisting of successive cycles of dye incorporation, imaging, and cleavage.

Automating the cyclic chemistry steps of SBS would therefore expand the throughput and accessibility of optical pooled screens. Various approaches may be considered for fluidics automation, including using different sample formats, adapting existing instruments, or building custom systems (Table 3.1). While commercial platforms for next-generation sequencing (NGS) and spatial genomics are available and can be repurposed for *in situ* SBS, almost all of them are based on flow cells or other formats that may reduce reagent costs but are non-standard for cell culture and thus, impractical[172]. Automated liquid handling systems maintain compatibility with well plates and are more easily deployed in industry settings. However, these instruments are cost-prohibitive for most academic labs.

Here, we developed a simple, low-cost, replicable fluidics platform to automate *in situ* SBS in standard tissue culture well plates. Our custom perfusion system can perform heated reagent exchanges in 6-well plates, directly on a microscope stage, fully automating the sequencing chemistry and imaging steps, and completing 12 cycles of sequencing in 2-3 days, with minimal hands-on time (<2 hours, compared to 18 hours when manually sequencing). With resulting labor reduction and equipment utilization improvements (particularly 24-hour unattended operation), we can increase the effective throughput of OPS at least two-fold.

	Flow cell	Automated liquid handling	Custom 6-well perfusion system
Surface area	9 cm²	58 cm²	58 cm²
Compatibility with multi-well plates	No	Yes	Yes
SBS reagent usage per cycle	Low	Standard	Standard
Capital cost	Low	High	Low
Integrated imaging	Yes	No	Yes
Industry readiness	Low	High	Low

Table 3.1 | Strategies for fluidics automation of cyclic chemistry protocols.

Flow cell-based fluidics automation platforms require low capital costs and reduce reagent consumption but are not compatible with standard tissue culture well-plate formats. Automated liquid handling robots address this limitation but rely on expensive capital equipment. Custom 6-well perfusion systems maintain compatibility with common tissue culture protocols and present a lower-cost and accessible approach for academic labs.

3.2 Results

3.2.1 Optimization of in situ sequencing by synthesis protocol

To allow seamless transition between sample preparation and imaging, we envisioned the sample being placed on the microscope stage for the entire duration of the sequencing. A major hurdle for automating the SBS workflow in this way is that the standard Illumina protocol requires all chemistry steps to be performed at 60˚C. However, maintaining a plate on a microscope at 60˚C with uniform temperature distribution for extended periods of time is both technically challenging and damaging to the optical components.

We first tested whether we could eliminate the heating requirement and perform the sequencing protocol at lower temperatures and still accurately call reads. Using A549 cells containing a pooled sgRNA library, we read out the first base of the sgRNA sequences by manually incorporating (and cleaving) it at a range of temperatures between 25˚C and 60˚C. We observed that, compared to protocols at 40˚C or higher, the 25˚C protocol resulted in many cells containing multiple reads, but yielded slightly lower read signal intensities (Figure 3.2A, Supplementary Figure 3.1A). With diluted incorporation mix (MiSeq reagent 1), the sequencing quality at 25˚C decreased even further, whereas the 40˚C protocol still produced high-quality reads (Figure 3.2A, Supplementary Figure 3.1A). Thus, we concluded that while a room-temperature protocol was not robust, SBS at temperatures lower than 60˚C was feasible.

We next performed 12 cycles of SBS manually at 40˚C, 50˚C, and 60˚C, to determine the effect of the chemistry temperature on the quality of longer reads. Sequencing quality at 40˚C and 50˚C remained comparable to that at 60˚C over 12-nucleotide reads, decreasing slightly with dilutions of the incorporation and cleavage reagents (Supplementary Figure 3.1). Indeed, the base quality throughout all 12 cycles of SBS was robust even at 40˚C (Figure 3.2B). Finally, across multiple independent experiments, we showed that *in situ* SBS can be successfully performed at temperature as low as 40˚C, at least over 12 cycles of sequencing required for genome-wide screens (Figure 3.2C).

Since many commercial microscope incubators may be set to a maximum of 45˚C, our findings supported an automation platform design, in which a cage incubator is used to heat the sample plate on the microscope during both the chemistry steps and imaging, without sacrificing the sequencing quality.

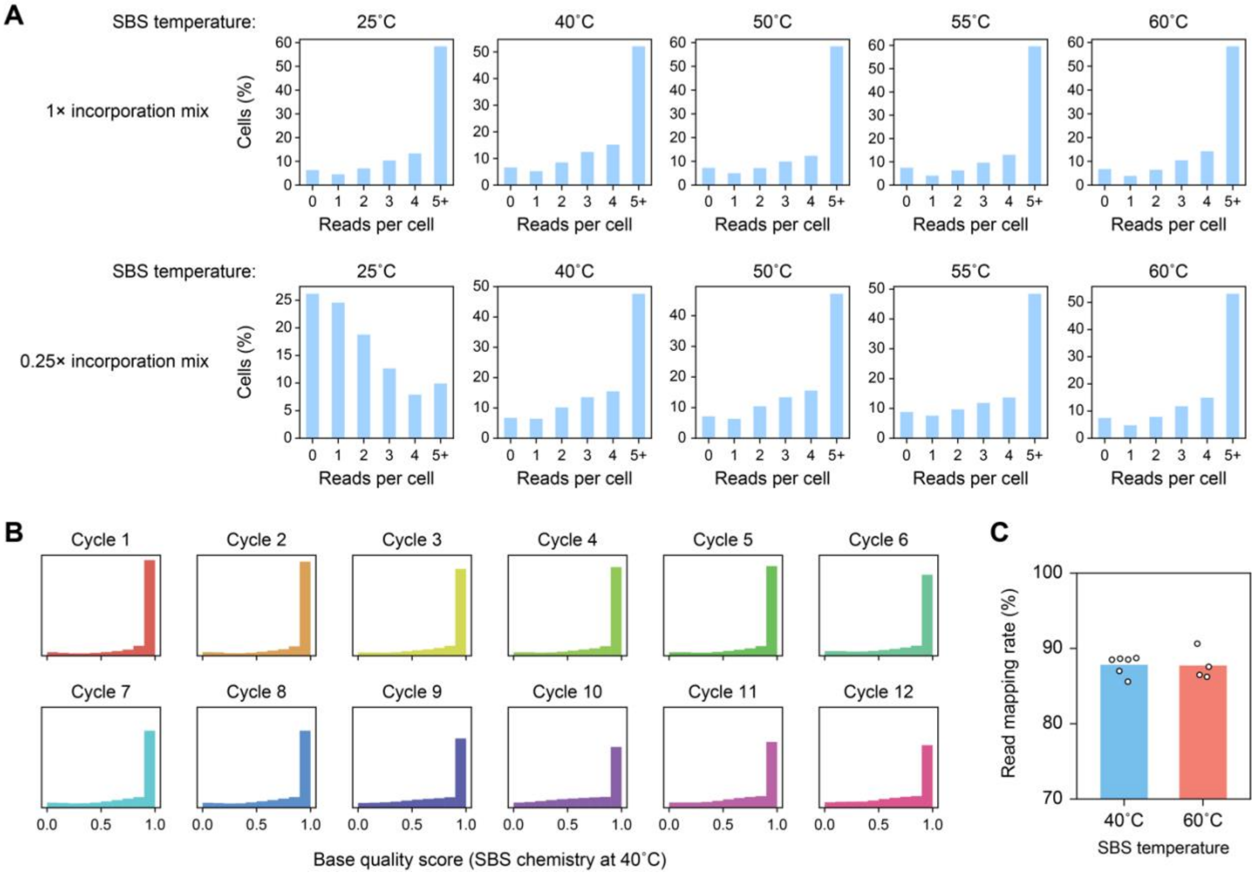


Figure 3.2 | Temperature constraints of *in situ* sequencing by synthesis (SBS).

(A) Distributions of the *in situ* barcode read counts in A549 cells expressing an sgRNA library. All chemistry steps of the SBS protocol were performed isothermally at a given temperature, ranging from room temperature (25˚C) to the standard 60˚C. The incorporation steps were performed using either undiluted or 0.25× diluted incorporation mix.

(B) Distributions of base quality scores from each SBS cycle, with the chemistry performed at 40˚C.

(C) Bar graph compares the proportion of *in situ* reads mapping to the designed library when the SBS chemistry is performed at 40˚C (n = 6 independent experiments), compared to the standard 60˚C (n = 4).

3.2.2 Custom perfusion system for automated fluidics

To automate fluidics for *in situ* SBS in a standard 6-well plate, we designed a perfusion system that includes custom-made well inserts with ports for needles, and tubing for fluid exchange realized by a peristaltic pump (Figure 3.3A-C, Table 3.2). We also developed custom software (programmed in MATLAB) to control all fluidics through a TTL microcontroller, with an I/O DAQ. Using another TTL microcontroller, we integrated our fluidics platform with existing image acquisition software (Nikon NIS-Elements) to enable continuous transitions between the SBS chemistry and imaging workflows in a multi-cycle sequencing run.

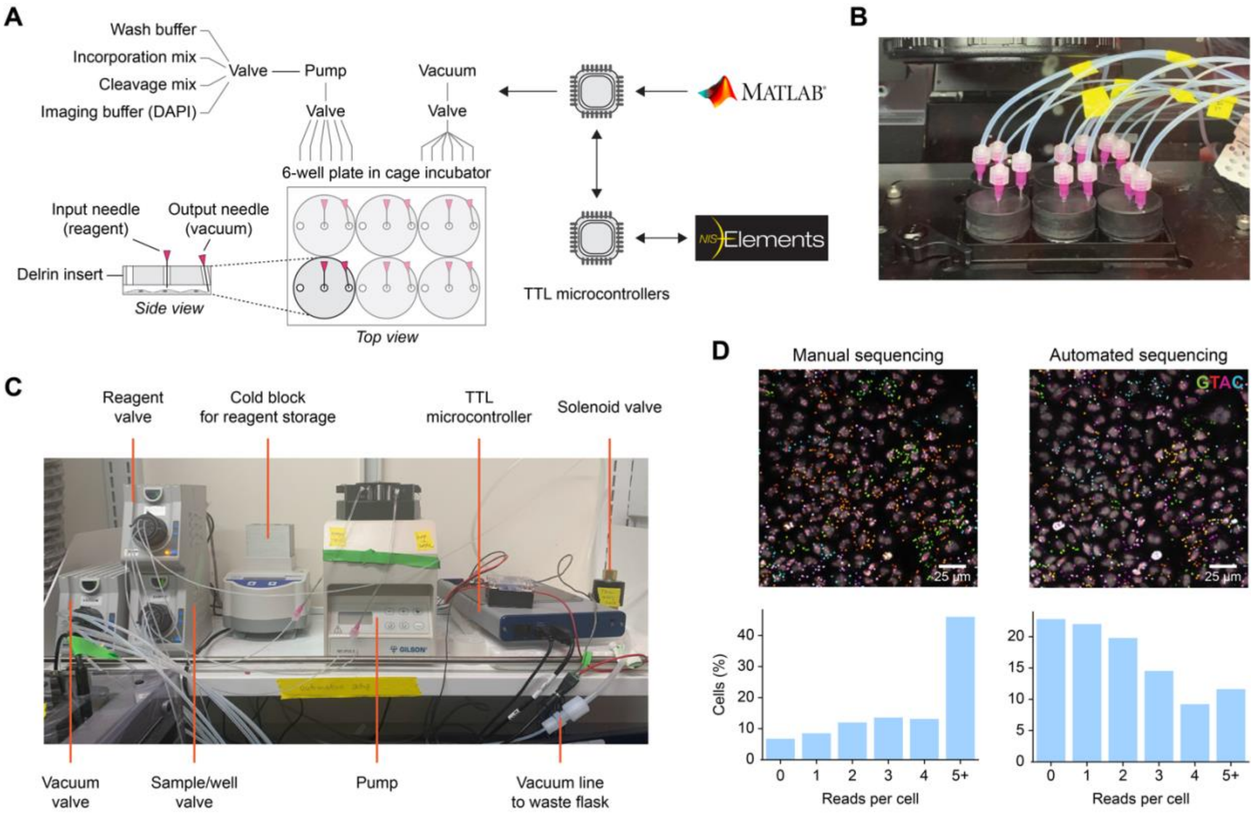
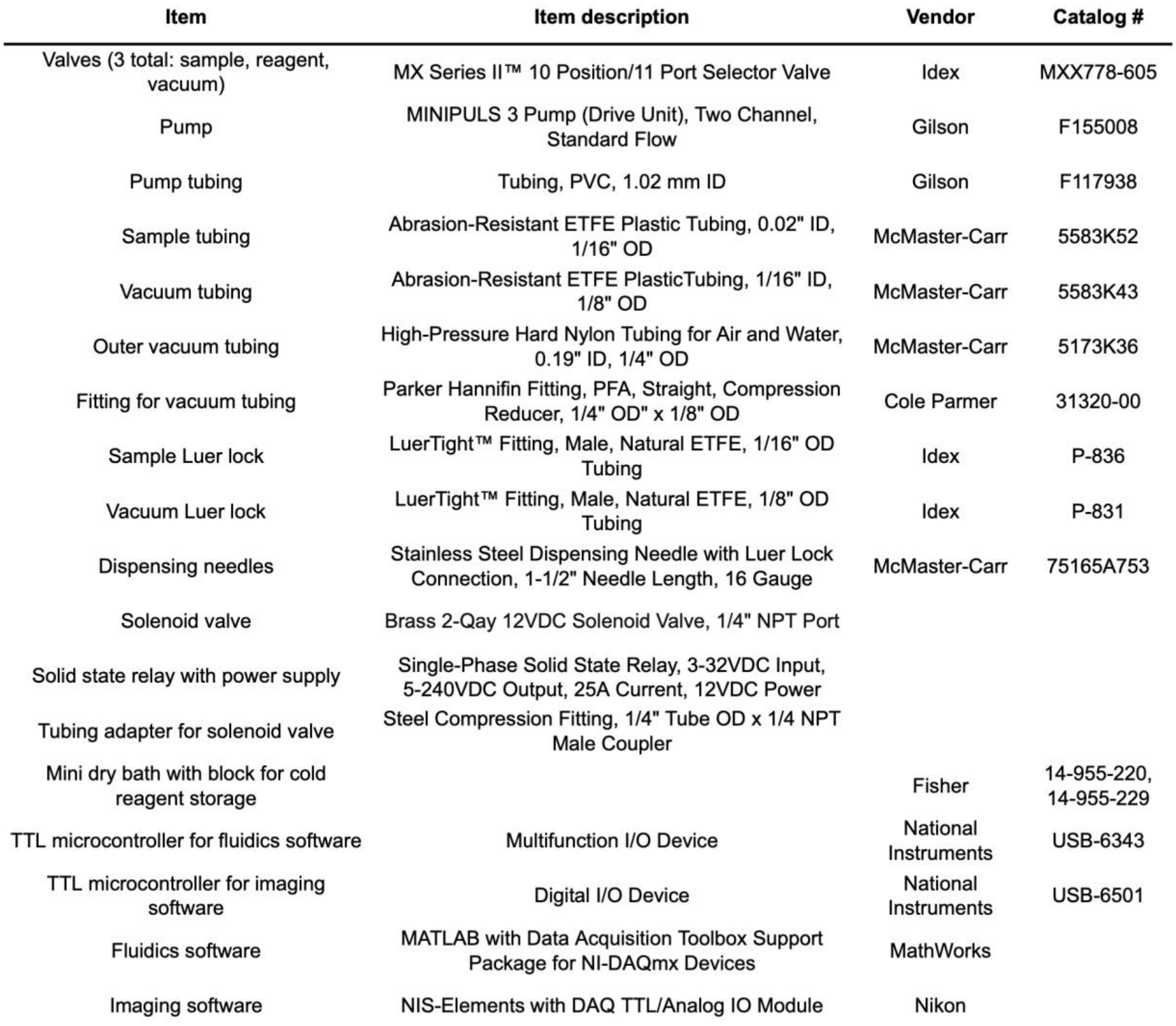


Figure 3.3 | Custom 6-well perfusion system for automated *in situ* SBS.

(A) Schematic of perfusion system for automated fluidics in 6-well plates. The plate is set on a microscope stage inside a cage incubator set at 45˚C for efficient SBS performance. Each well is fitted with a Delrin insert to hold needles for reagent addition and removal. The rest of the fluidics setup is kept nearby, outside the incubator. Reagents are stored appropriately either at room temperature (wash buffer and DAPI) or 4˚C on a cold block (incorporation and cleavage mixes). A peristaltic pump flows reagents from a reagent valve into a given well through a sample valve. Reagents are aspirated from the well through a vacuum valve connected to the house vacuum system. The fluidics components are controlled through a TTL microcontroller using custom software in MATLAB and integrated with imaging software (NIS-Elements) via another I/O device.
(B) Image of a 6-well plate inside the cage incubator, with all inserts, needles, and tubing attached.
(C) Image of fluidics setup showing key components.
(D) Comparison of *in situ* sequencing performance between the standard manual protocol (left, cycle 1) and the prototypical automated fluidics system (right, cycle 2 on the same sample). A549 cells used contained a library of sgRNAs expressing using the CROP-seq vector. (top) Representative image of SBS for the same field-of-view, showing sequencing spots (G: green, T: red, A: magenta, and C: cyan) and nuclei (DAPI, gray). Scale bar, 25 µm. (bottom) Distributions of read counts per cell.



Item	Item description	Vendor	Catalog #
Valves (3 total: sample, reagent, vacuum)	MX Series II™ 10 Position/11 Port Selector Valve	Idex	MXX778-605
Pump	MINIPULS 3 Pump (Drive Unit), Two Channel, Standard Flow	Gilson	F155008
Pump tubing	Tubing, PVC, 1.02 mm ID	Gilson	F117938
Sample tubing	Abrasion-Resistant ETFE Plastic Tubing, 0.02" ID, 1/16" OD	McMaster-Carr	5583K52
Vacuum tubing	Abrasion-Resistant ETFE PlasticTubing, 1/16" ID, 1/8" OD	McMaster-Carr	5583K43
Outer vacuum tubing	High-Pressure Hard Nylon Tubing for Air and Water, 0.19" ID, 1/4" OD	McMaster-Carr	5173K36
Fitting for vacuum tubing	Parker Hannifin Fitting, PFA, Straight, Compression Reducer, 1/4" OD" x 1/8" OD	Cole Parmer	31320-00
Sample Luer lock	LuerTight™ Fitting, Male, Natural ETFE, 1/16" OD Tubing	Idex	P-836
Vacuum Luer lock	LuerTight™ Fitting, Male, Natural ETFE, 1/8" OD Tubing	Idex	P-831
Dispensing needles	Stainless Steel Dispensing Needle with Luer Lock Connection, 1-1/2" Needle Length, 16 Gauge	McMaster-Carr	75165A753
Solenoid valve	Brass 2-Qay 12VDC Solenoid Valve, 1/4" NPT Port		
Solid state relay with power supply	Single-Phase Solid State Relay, 3-32VDC Input, 5-240VDC Output, 25A Current, 12VDC Power		
Tubing adapter for solenoid valve	Steel Compression Fitting, 1/4" Tube OD x 1/4 NPT Male Coupler		
Mini dry bath with block for cold reagent storage		Fisher	14-955-220, 14-955-229
TTL microcontroller for fluidics software	Multifunction I/O Device	National Instruments	USB-6343
TTL microcontroller for imaging software	Digital I/O Device	National Instruments	USB-6501
Fluidics software	MATLAB with Data Acquisition Toolbox Support Package for NI-DAQmx Devices	MathWorks	
Imaging software	NIS-Elements with DAQ TTL/Analog IO Module	Nikon	

Table 3.2 | Parts list for 6-well perfusion system for automated *in situ* chemistry fluidics.

For each component, the recommended manufacturer's name and catalog number is listed. The Delrin well inserts (not listed) were custom designed and machined.

Most fluidics components are kept outside the cage incubator (Figure 3.3C). The imaging buffer (DAPI) and wash buffer (MiSeq PR2 incorporation buffer) are left at room temperature, while the incorporation and cleavage mixes are stored in a cold block at 4˚C. A peristaltic pump flows one of a reagents through the reagent valve and the sample/well valve into a given well of a plate, set on the microscope stage in an Okolab cage incubator heated to 45˚C (its highest certified temperature). Each well is tightly fitted with a removable Delrin insert in black, which exhibits relatively low autofluorescence (Figure 3.3B). Each insert has three ports: one to equilibrate the pressure inside the well (and prevent the glass bottom from cracking), one for the reagent needle connected to the sample valve, and one for the vacuum needle connected to the house vacuum via another valve, which aspirates liquid from the well.

We tested our prototypical setup to benchmark its performance against manual sequencing. On the same sample, we performed the first cycle of sequencing manually, using the standard 60˚C protocol, followed by the second cycle using the automation platform. Although the automated chemistry successfully worked on some sequencing reads, many lost signal between the two cycles (Figure 3.3D). We thus sought to optimize our automated fluidics system to perform SBS more efficiently and robustly.

3.2.3 Validation of well plate-based automated in situ sequencing

Through rounds of testing and optimizations, we introduced a series of changes to our setup. First, we added foam insulation to the outside of the cage incubator to maintain the temperature more stably at 45˚C over the course of multi-day sequencing experiments and prevent fluctuations due to the surrounding room temperature.

Next, we moved the reagent needle to the center port of the well insert to allow the liquid to cover the well surface evenly more quickly and limit the dry time of the sample. The height of the reagent needle was also lifted by an additional 2mm from the surface to prevent bubble formation when reagents were dispensed into the well.

We next implemented changes to ensure the purity of reagents inside the well. In our previous manual experiments, we found that diluted incorporation and cleavage mixes yielded lower-quality reads (Supplementary Figure 3.1). Thus, we angled the vacuum needle to touch the edge between the bottom and wall of the well. This position of the vacuum needle more accurately mimics the position of a pipette tip when aspirating volume from the well manually, allowing more complete removal of reagents from the well and limiting the dilution of the subsequent reagent. To prevent reagents from mixing inside tubing throughout the fluidics system, we also added an air bubble barrier between each reagent flow in the protocol.

Finally, we observed consistently higher background signal in automated sequencing run, compared to manual experiments. To limit this background from rising significantly in later cycles, we altered the wash protocols following the incorporation and cleavage steps to include more frequent washes, as well as longer incubations in wash buffer.

We tested this optimized system by performing a fully automated 12-cycle sequencing run in a single well of a 6-well plate. We demonstrated that the automated sequencing quality is comparable to that of manual experiments, with high mapping rates across all cycles and across the area of the well (Figure 3.4). The high sequencing quality was relatively spatially uniform throughout the well, outside of focusing errors at select sites that were independent of the automation platform.

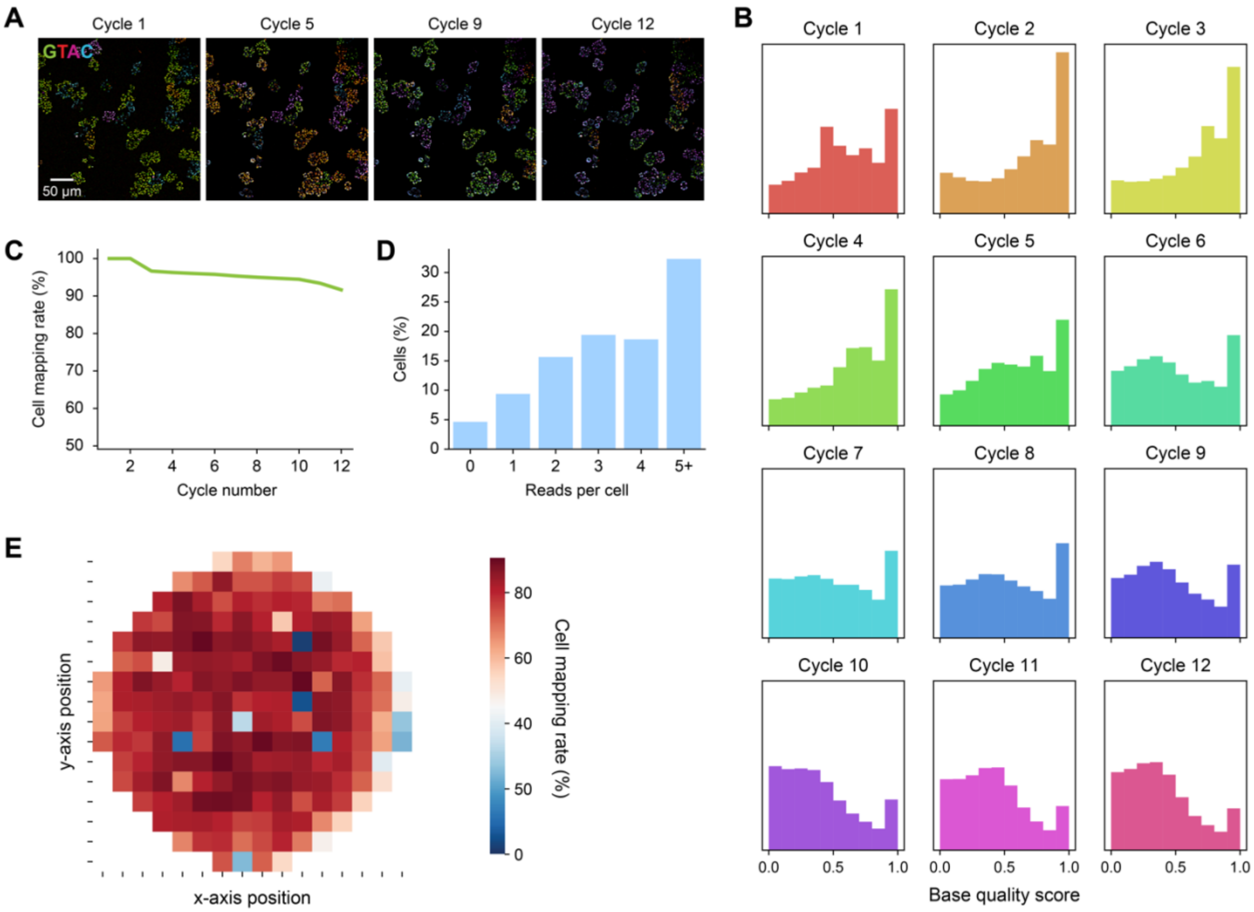


Figure 3.4 | Optimized automated *in situ* SBS protocol in a single well.

(A) Images of selected cycles of single-well automated *in situ* sequencing (Laplacian-of-Gaussian filtered) for the same field-of-view. Scale bar, 50 µm.
(B) Distributions of base quality scores from each SBS cycle.
(C) Percentage of cells with reads exactly matching a designed sgRNA over sequencing cycles. HeLa cells used contained a library of 10 sgRNAs. Only cells in which a single barcode sequence constitutes a plurality of all reads detected in that cell were included in the analysis.
(D) Distributions of read counts per cell.
(E) Cell mapping rates across fields-of-view within a well.

Next, we extended our single-well protocol to all 6 wells of a plate. In our implementation, the SBS chemistry of a given cycle is completed for a full plate before proceeding to image acquisition. Because the fluidics can operate on only one well at a time, we introduced minor changes to the protocol to better synchronize incubation steps between wells, minimizing sample processing times. We also installed a solenoid valve to enable finer regulation of the vacuum reagent removal steps, preventing liquid from being aspirated from wells upon valve switching.

Using a HeLa cell library containing 80,862 sgRNAs, we tested the performance of this multi-well automated fluidics platform across independent experiments. We showed that our custom perfusion system yields high-quality reads and high mapping rates uniformly across the area of a well and all wells of a plate (Figure 3.5, Supplementary Figure 3.2). After replicating this system on another microscope, we achieved similar results, demonstrating the reproducibility and adaptability of our approach to other labs and setups.

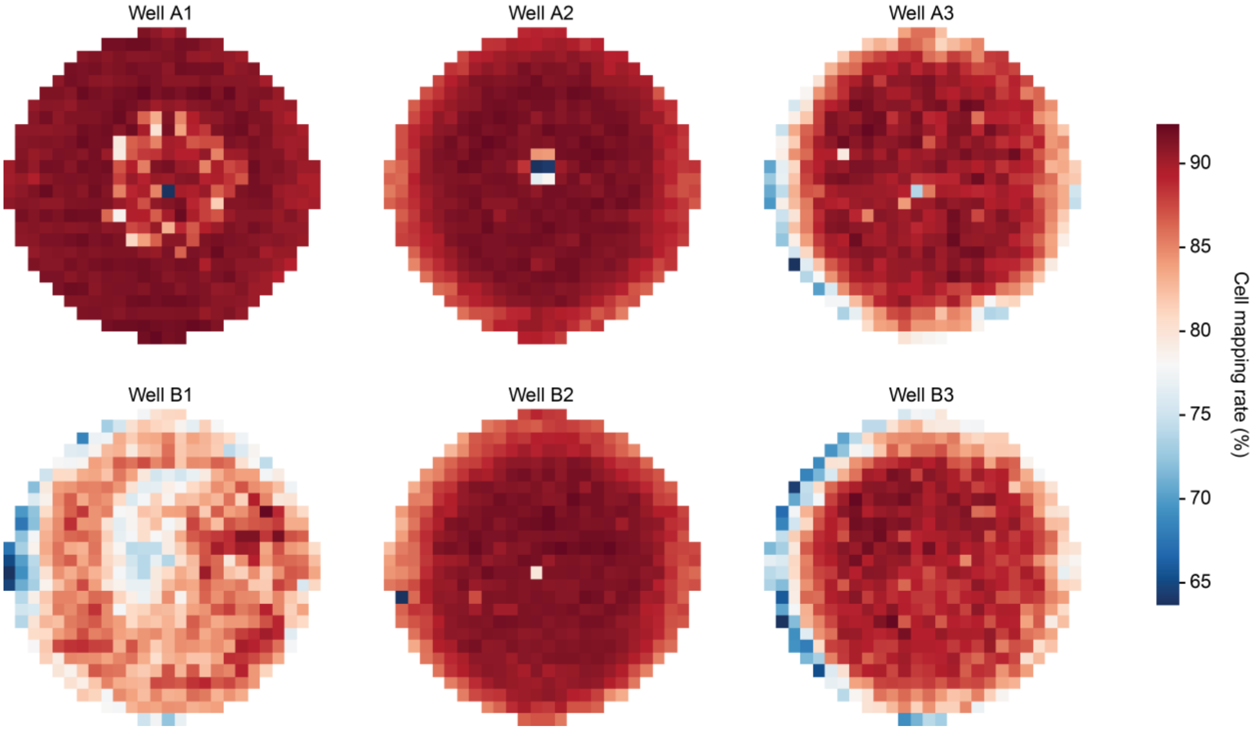


Figure 3.5 | Automated *in situ* SBS in a full 6-well plate.

Cell mapping rates across fields-of-view in a 6-well plate of HeLa cells containing a genome-wide library of 80,862 sgRNAs. In total, over 8.4 million cells were mapped in this plate (1.2-1.5 million cells mapped per well).

3.3 Discussion

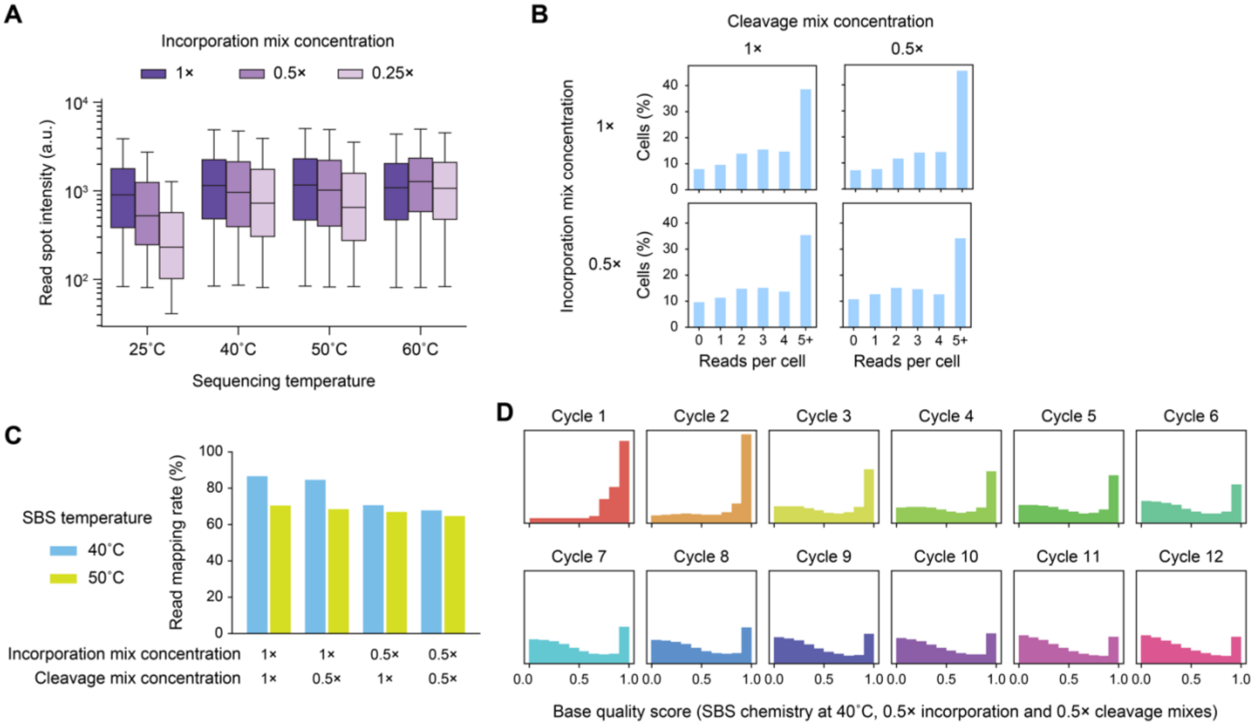
In this chapter, we describe our custom fluidics platform for automating *in situ* sequencing in standard tissue culture well plates. We showed that, at least for shorter read lengths of up to 12 nucleotides, *in situ* sequencing by synthesis (SBS) is robust even when performed at temperatures as low as 40˚C. This finding allowed us to design a perfusion system that can carry out SBS chemistry steps in a well plate set on a microscope stage, inside a heated cage incubator at constant temperature. We showed that our optimized platform can effectively, reliably, and reproducibly automate up to 12 cycles of *in situ* SBS in 6-well plates in ~3 days, with less than 2 hours of hands-on time. This increase in screening capacity will enable more complex and larger-scale optical pooled screens.

Further protocol changes may be needed as the system is tested in more screening contexts. Performing a small-scale screen with additional replicates would allow a direct comparison between results obtained with manual sequencing and those from automated sequencing. These data would then inform how to refine both the fluidics hardware and analysis pipeline (e.g. by including more robust background signal subtraction or fluorescence compensation) to improve read quality and base calling accuracy, matching or even surpassing the performance of manual SBS, especially in removing sample-to-sample variability.

Moreover, the hardware can be further optimized to maximize flow during fluid exchange, decrease reagent usage and costs, and reduce sample processing times, e.g. by evaluating various needle and tubing choices, optimizing the number and positioning of inlets and outlets within a well, and minimizing the working volume within a well by adjusting the height of the insert. Other engineering controls may also be incorporated. For example, a liquid level detection device can provide a failsafe measure in case of leaks or incomplete reagent exchange, and a custom plate insert could serve as a reproducible attachment to simplify setting up the fluidics for a multi-well plate.

More broadly, many highly multiplexed spatial genomic methods (e.g. immuno-SABER, CODEX, CycIF, 4i) require a significant investment of hands-on labor for cyclic reagent exchanges and suffer from similar sample processing bottlenecks as *in situ* SBS[17–20,173]. Although this work focused on automating SBS, our platform is highly customizable and compatible with other imaging reagents, and thus will expand the throughput and accessibility of not only optical pooled screens, but also many spatial technologies. In building and testing our fluidics setup on two different microscopes, we showed that our approach is easily replicable. Through thorough protocol documentation and additional improvements to the software, we aim to make our automation platform open-source, user-friendly, and customizable to various experimental requirements to facilitate its wide deployment. We envision other labs will be able to easily implement and build upon this perfusion system on their microscopes for diverse protocols. Our automated fluidics platform can therefore greatly increase the number and scale of projects, as well as drive the pace of innovation in spatial genomics and proteomics.

3.4 Supplementary Figures



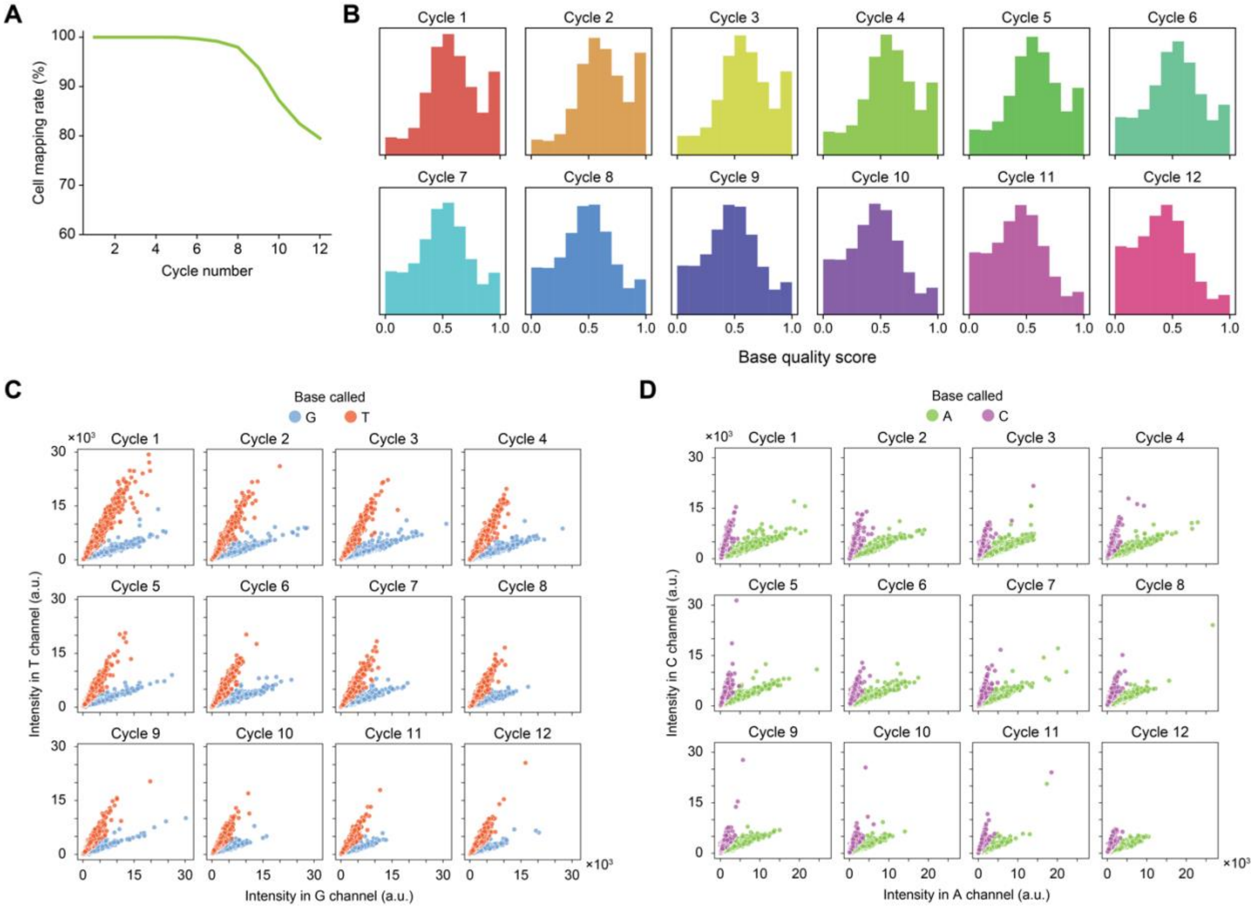
Supplementary Figure 3.1 | SBS performance with reagent dilutions.

(A) At increasing dilution factors of the incorporation mix, the signal intensity of the *in situ* sequencing reads decreases, particularly at lower chemistry temperatures.

(B) Distributions of read counts per cell. SBS was performed at 40˚C with 0.5× diluted and undiluted incorporation and cleavage mixes.

(C) Percentage of detected reads that mapped an sgRNA in the designed library. SBS was performed either at 40˚C or 50˚C.

(D) Distributions of base quality scores from each sequencing cycle, with SBS performed at 40˚C with 0.5× diluted incorporation and cleavage mixes.



Supplementary Figure 3.2 | Technical performance and quality control of automated *in situ* SBS in a 6-well plate.

(A) Percentage of cells with reads mapping to a designed sgRNA with increasing read lengths. HeLa cells used contained a genome-wide library of 80,862 sgRNAs.

(B) Distributions of base quality scores from each sequencing cycle.

(C) Scatterplots of the raw intensities of two-channel combinations with the most crosstalk (G vs. T and A vs. C) at each SBS cycle. The color indicates which base was called by the analysis pipeline.

Chapter 4. Measuring transcriptional dynamics in living cells using virus-like particles

Authors: Jacob Borrajo*, Mohamad Ali Najia*, Anna Le*, FuNien Tsai, George Q. Daley, Paul C. Blainey
*Contributed equally

4.1 Introduction

Biological research is rooted in time series analyses of living organisms and cells. The utility of light microscopy was dramatically enhanced by the introduction of genetically encodable non-toxic fluorescent proteins, which established an important interface between the dynamic analysis of cellular activity and the burgeoning field of molecular biology[174,175]. In today's genomic era, high-throughput RNA measurement using RNA sequencing (RNA-seq) techniques is indispensable for the molecular characterization of cell states and functions across the life sciences[176–178]. However, RNA-seq has been applied almost exclusively as a destructive method in which biological samples are lysed for RNA extraction or transcripts are physically retrieved from a small number of cells via invasive instrumental procedures[179–183]. Destructive RNA-seq methods cannot produce true time series data from the same sample, precluding comparison of such measurements at different time points. These limitations hobble the ability of RNA-seq to directly characterize dynamic functional activity or state changes in cells and prevent analysis of cells that are not physically accessible for destructive analysis. The popularity of computational methods that indirectly infer the time evolution of transcriptional dynamics from destructive "snapshot" data, and recognition of the limits of such "pseudotime" approaches highlight the need for non-destructive RNA-seq methods[184–190]. To record RNA levels from the same cells over time, several groups have pioneered optical methods that use endogenous or exogenous probes to monitor a few native or tagged target transcripts at a time[191–197]. More recently, molecular recording methods have allowed information about the age of transcripts or prior transcriptional states to be encoded in the DNA or RNA of living systems for readout at a single end-point[198–201]. While promising, these methods require a prior transcriptional hypothesis, provide only summary-level information, have limited temporal dynamic ranges, and do not retain the living biological sample after the measurement.

Inspired by exosome and viral biology, we sought to overcome limitations of conventional RNA-seq by engineering a synthetic RNA export pathway that enables mammalian cells to "self-report" transcriptional states in real-time. Exosomes are physiological extracellular vesicles produced by cells that contain components of cell membranes and cytosol (including RNA) and can transport molecules between cells[202]. In contrast, retroviruses have evolved the ability to specifically package and export their genomic RNA in addition to delivering and propagating their RNA genome to complete their life cycle. Retroviral export pathways are an especially attractive substrate for engineering an RNA export system, as a technology based on select retroviral proteins could be orthogonal to host biology and activated in the time, place, and strength desired by researchers to achieve tailored reporting of transcriptional information. Specifically, we hypothesized that by repurposing a retroviral RNA export pathway, we could continuously sample RNA from living cells to enable transcriptome-wide non-destructive RNA-seq measurements.

4.2 Results

4.2.1 VLPs non-specifically package cellular RNAs

Virus-like particles (VLPs) produced by mammalian cells transfected with Gag, the core structural polyprotein for retroviruses, package cellular RNAs non-selectively in the absence of cis-acting viral packaging signals[203]. While VLPs already have a broad range of technological applications including vaccinology[204], the delivery of protein payloads to cells[73,75], and measurement of protein-protein interactions[77], we hypothesized that VLPs could also be leveraged to create an engineered RNA export pathway that would allow living cells to self-report transcriptional information to the extracellular environment, as previously demonstrated for barcoded transcripts[205] (Figure 4.1A-C). We envision that live-cell transcriptional profiling will ultimately enable researchers to follow dynamic biological activities through time at the whole-transcriptome level from cells in a wide range of living biological systems, and follow up with further studies of the intact cells or organism after time series transcriptional analysis.

To evaluate retroviral-based VLPs as the basis of such a technology platform, we transduced HEK293T with doxycycline (dox)-inducible murine leukemia virus (MLV) Gag fused to a P2A-linked GFP reporter (Supplementary Figure 4.1A) to validate full-length Gag expression and translation. After induction with doxycycline, we confirmed Gag expression via flow cytometry (Supplementary Figure 4.1B) and purified cell supernatants via ultracentrifugation. We then performed electron microscopy on purified supernatant from dox-induced cells to confirm the presence of VLPs (Supplementary Figure 4.1C). Further, we measured a 5.9 ± 1.2 fold increase (mean ± SD, $n = 3$, $p = 0.0024$) in exported GAPDH mRNA via quantitative reverse-transcription PCR (RT-qPCR) in dox-induced cell supernatants relative to baseline (Supplementary Figure 4.1D), confirming that Gag expression increases RNA export from living cells. Additionally, we characterized the stability of VLPs and found that packaged RNA was stabilized for at least 48 hours in media at 37°C, whereas RNA not protected by VLPs degraded under the same conditions (Figure 4.1D). Importantly, HEK293T and iPS self-reporting cells displayed normal morphology (Supplementary Figure 4.1E-F). Finally, to further confirm that cellular RNAs were indeed packaged inside VLPs, we treated cellular media with benzonase nuclease and observed a 10-fold enrichment of RNA-seq reads that aligned to the human transcriptome from Gag-expressing cells compared to cells expressing a mutant Gag protein lacking the MA domain (GagΔMA) that is unable to form VLPs (Supplementary Figure 4.2). These results confirmed that the measured RNA was exported within VLPs induced by Gag expression in cells engineered to self-report their transcriptional state.

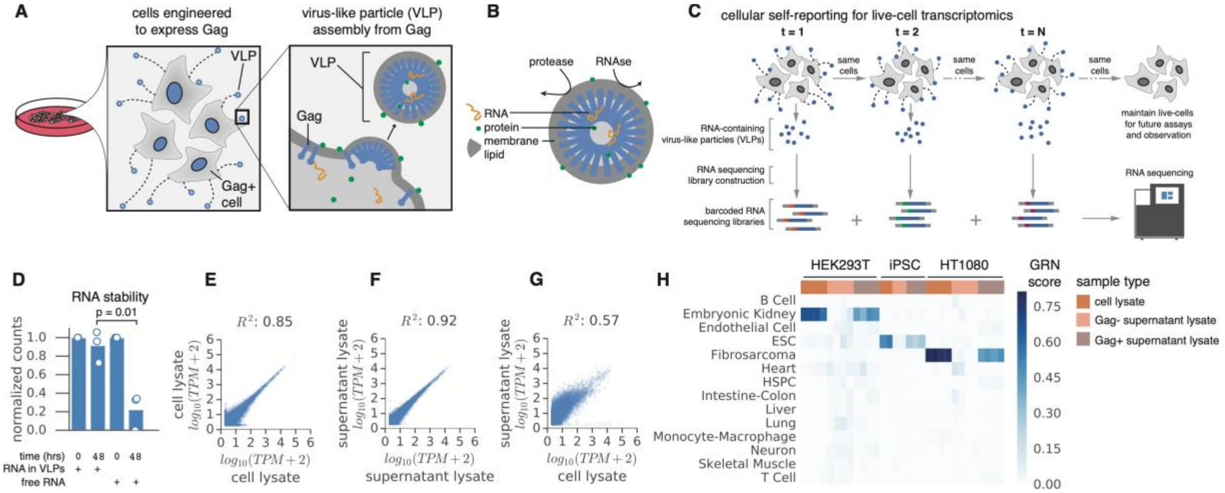


Figure 4.1 | Self-reporting via VLPs for live-cell transcriptomics.

(A) Self-reporting leverages virus-like particles (VLPs) to export RNA from living cells. Gag accumulates to assemble VLPs.
(B) VLPs can package several different types of cargos, including RNA, protein, and metabolites.
(C) Example of time-point collection using cellular self-reporting, where multiple time-points can be collected from the same biological samples.
(D) RNA stability in media at 37°C.
(E) RNA-seq replicate concordance of pGag+ HEK293T cell lysates (biological replicates).
(F) RNA-seq replicate concordance of pGag+ HEK293T supernatant lysates (biological replicates).
(G) RNA-seq sample representation showing pGag+ HEK293T lysate vs. pGag+ supernatant lysate.
(H) CellNet gene regulatory network (GRN) scores of RNA-seq from lysates and supernatants with and without Gag expression for HEK293T, iPSC, and HT1080 cells.

To characterize the RNA content of the VLPs, we performed RNA-seq[206] of VLPs harvested from HEK293T cells transfected with the self-reporting constructs. RNA-seq data showed that replicate concordance for VLP-packaged RNA was comparable to that of lysate-based RNA-seq (Figure 4.1E-F). Strikingly, we observed that the exported RNA was substantially representative of cellular lysate and that the relative abundance of transcripts measured in purified VLPs correlated with supernatant (Figure 4.1G). We then tested whether Gag-based RNA self-reporting could be generalized beyond HEK293T cells by stably integrating Gag expression constructs into the genome of induced pluripotent stem cells (iPSCs) and HT1080 cells. We assessed whether the RNAs packaged and exported in VLPs in these cell lines were reflective of the cellular transcriptome and sufficiently informative to distinguish cell states. To do so, we sequenced VLP-derived RNAs produced from each cell type and applied CellNet[207] to ascertain the active regulatory networks and predict the biological identity of the sample. The VLP-derived transcriptomes supported specific classification to the expected identities, mirroring the classification performance of conventional lysate RNA-seq controls (Figure 4.1H, Supplementary

Figure 4.3). These data highlight how the high-dimensional transcriptional readout afforded by self-reporting is amenable for gene regulatory network analysis and that VLP-derived RNAs are reflective of cellular states in the self-reporting cells.

4.2.2 Engineering Gag proteins and capsids for specific and multiplexed readouts

We hypothesized that VLPs could be engineered to tailor RNA packaging via modification of the RNA-binding domain in the Gag polyprotein, as well as display affinity-tagged envelope proteins to enable fascicle VLP purification (Figure 4.2A). Classically, the C-terminal nucleocapsid (NC) domain in the Gag polyprotein recognizes the cis-acting packaging signal in the viral RNA genome, but this basic domain also interacts nonspecifically with negatively charged RNA molecules[203,208]. While a broad range of known RNA-binding domains could, in principle, be fused with Gag to optimize RNA capture for different export applications, we focused here on poly(A)-binding domains with the goal of enhancing mRNA export. Thus, we selected a tandem RNA recognition motif domain RRM1-2 from human PABPC4 to generate a Gag fusion protein, as RRM1-2 has been shown to interact with polyadenosine RNA with high affinity[209]. We transduced HEK293T and HT1080 cells with lentivirus containing the designed fusion construct to establish integrated cell lines with constitutive expression (Supplementary Figure 4.4A). After purifying VLPs from supernatants and preparing RNA-seq libraries (Supplementary Figure 4.4B-E), we measured the increase in the number of genes detected from Gag-expressing cells relative to the wild-type controls in HT1080 cells and found that fusing RRM1-2 to Gag resulted in the detection of more genes in HT1080 cells relative to wild-type Gag (Figure 4.2B). This demonstrates that engineered Gag fusion proteins support RNA export and that RNA export profiles can be tuned by modulating the RNA binding activity of engineered Gag fusion proteins. Interestingly, we found that expression of Gag–RRM and Gag were minimally perturbative of transcriptional profiles in HT1080 and HEK293T cells (Figure 4.2C, Supplementary Figure 4.5A-B), and that Gag expression led only to detectable upregulation of SERF1A expression in iPS cells (Supplementary Figure 4.5C). We observed minimal bias in the annotated localization of transcripts in self-reported RNA-seq data relative to cell lysate controls (Figure 4.2D, Supplementary Figure 4.6). Indeed, exosome association was most predictive in the case of wild-type HEK293T and HT1080 control cells lacking an engineered export pathway, where a significant fraction of the RNA sequences may have in fact derived from naturally occurring exosomes. We also observed that VLPs exhibit a modest preference for packaging longer RNA molecules (Supplementary Figure 4.7).

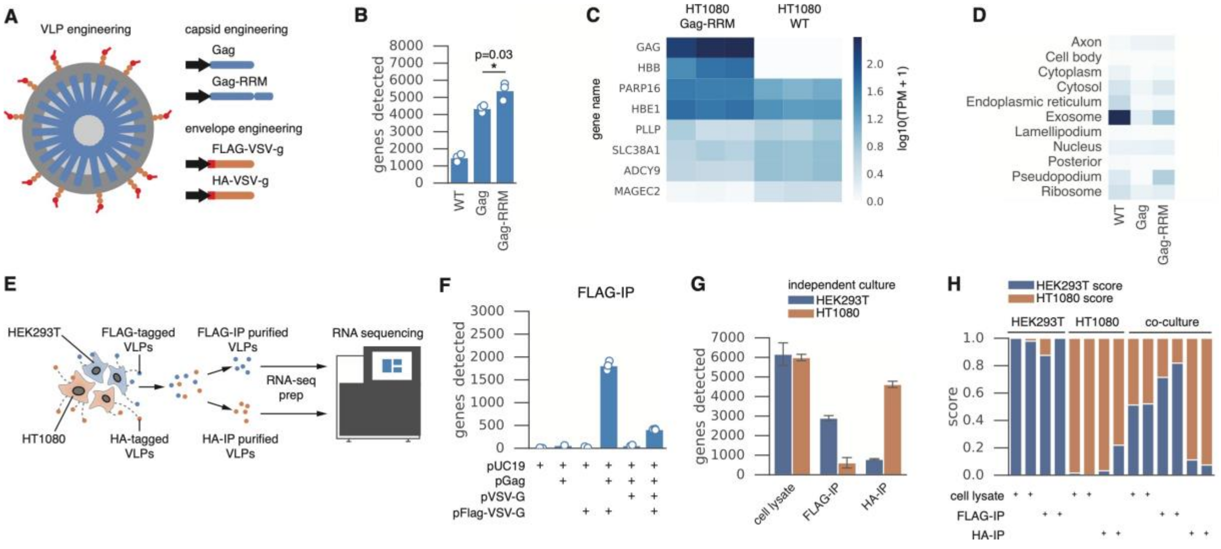


Figure 4.2 | Rational VLP engineering.

(A) Schematic of an engineered VLP.
(B) Gag–RRM leads to an increase in detected genes in HT1080 cells.
(C) Differentially expressed genes in HT1080 cells expressing Gag–RRM with columns as biological replicates (n = 3).
(D) Importance of RNA localization/association annotation in predicting abundance in supernatant samples using a gradient boosted tree model.
(E) Schematic of co-culture experiment. HT1080 and HEK293T cells are independently transduced with VSV-G envelope proteins harboring different affinity tags. Cells are mixed and VLPs are then purified for downstream RNA-seq.
(F) RNA-seq of HEK293T supernatant purified via FLAG immunoprecipitation.
(G) RNA-seq results showing high quality libraries for matching affinity-based immunoprecipitations in independent culture.
(H) Cell-line specific gene expression detected for appropriate cell types in mono-culture and co-culture.

Given that mammalian cells naturally exist in complex environments composed of multiple interacting cell types, we sought to multiplex cellular self-reporting in a workflow where we simultaneously sample VLPs produced by multiple cell populations within a co-culture (Figure 4.2E). We engineered the vesicular stomatitis virus G (VSV-G) envelope protein with an N-terminal FLAG tag and co-transfected HEK293Ts with MLV Gag to pseudotype VLPs. FLAG immunoprecipitation (IP) on cellular media revealed epitope-dependent specific isolation of VLPs by western blot (Supplementary Figure 4.8A). RNA-seq from the supernatant lysates led to detection of ~2000 genes from cells expressing Gag (Supplementary Figure 4.8B), while RNA-seq following FLAG IP led to high-quality libraries only from Gag and FLAG-VSV-G expressing cells as expected (Figure 4.2F), and resulted in the detection of ~1700 genes. Again, we observed good replicate concordance and RNA representation in the supernatant IP RNA-seq data (Supplementary Figure 4.8C-E). Furthermore, we validated a similar IP approach to isolate HA-tagged VLPs from Gag and HA-VSV-G expressing cells and observed performance comparable to our FLAG-IP results (Supplementary Figure 4.9A-B). Additionally, we confirmed the epitope specificity of each IP protocol (Supplementary Figure 4.9C-D), which suggested the possibility of purifying orthogonal VLP populations from a mixture. To test multiplexing of RNA self-reporting, we independently transduced HEK293T and HT1080 cells with Gag, along with VSV-G envelope proteins tagged with FLAG and HA, respectively (Supplementary Figure 4.10). We were able to construct high-quality RNA-seq libraries from purified VLPs of the independently cultured cell lines only when the IP antibody was cognate for each envelope tag (Figure 4.2G). The IP-based VLP purification performed well across a wide temporal range of media (cell supernatant containing VLPs) sampling durations, where we were able to detect genes significantly above background in as little as three hours of sampling (Supplementary Figure 4.11A-B). CellNet classification of transcripts detected by RNA-seq of IP-purified VLPs over various sampling windows further corroborated that the exported transcripts reflected the lysate transcriptome (Supplementary Figure 4.11C-D), underscoring the temporal resolution achievable with self-reporting. Finally, RNA-seq analysis of IP-purified, VLP-derived RNA from HEK293T–HT1080 co-cultures revealed cell line-specific gene expression (Figure 4.2H, Supplementary Figure 4.12), demonstrating the simultaneous specific analysis of independent cellular self-reporting information streams using the multiplexed IP approach with engineered VLPs.

4.2.3 VLP export enables temporal profiling of the transcriptome

Next, we aimed to determine whether live-cell RNA self-reporting, wherein signal is continuously integrated in the culture media via VLPs, can provide insight into transient transcriptional programs in living cells. To do this, we sought to detect NF-κB signaling in stimulated HT1080 cells. We cultured Gag–RRM self-reporting HT1080 cells and collected cell supernatant prior to TNF-α stimulation and every 12 hours throughout the stimulation to capture transcriptional dynamics from the same cells over a 36-hour period (Figure 4.3A). To compare self-reporting RNA-seq to traditional lysate RNA-seq, we prepared separate cell populations for destructive lysis at each timepoint and detected >1000 genes at each time point from VLP-containing cellular media and the lysate controls (Figure 4.3B). Principal component analysis (PCA) of expression data for genes with differential expression upon stimulation showed clear separation of stimulated cells at times post-stimulation in cell lysate samples as expected (Figure 4.3C). We observed similar separation when conducting PCA on genes differentially expressed in RNA-seq data from VLP-containing media samples at post-stimulation time points (Figure 4.3D). When conducting PCA with the same gene sets on supernatants from wild-type HT1080 cells lacking an engineered RNA export pathway, TNF-α-stimulated cells did not cleanly separate from unstimulated cells (Supplementary Figure 4.13A-C), indicating that background signals (i.e. from exosome or other spontaneous export processes) were insufficient to detect the transcriptional response of cells to TNF-α stimulation.

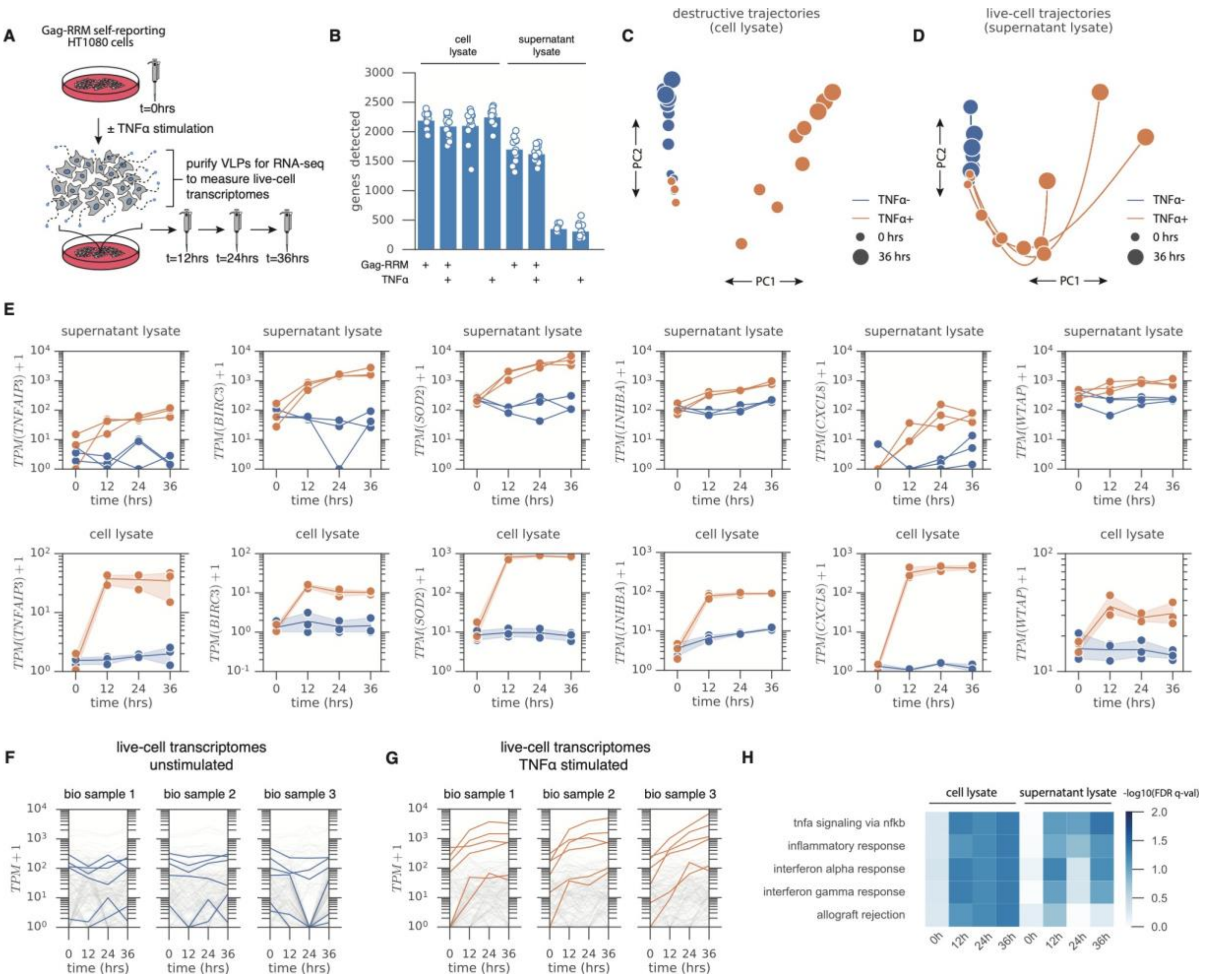


Figure 4.3 | Self-reporting reveals TNF-α response in longitudinally monitored HT1080 cells.

(A) HT1080 cells with stable single-copy integrated Gag–RRM self-reporting constructs were longitudinally monitored with and without TNF-α stimulation by harvesting supernatant with VLPs every 12 hours.

(B) Genes detected in RNA-seq libraries by sample type (± Gag–RRM expression, ± TNF-α stimulation).

(C) PCA on differentially expressed genes from lysate controls showing destructive trajectories for cells in TNF-α-stimulated (orange) and unstimulated (blue) conditions.

(D) PCA on differentially expressed genes from supernatants of self-reporting cells for live-cell transcriptional trajectories in TNF-α-stimulated (orange) and unstimulated (blue) conditions.

(E) Comparison of differentially expressed genes between TNF-α-stimulated (orange) and unstimulated (blue) conditions in supernatants of self-reporting cells. RNA-seq of VLPs from supernatant lysates (top) and RNA-seq of cell lysates (bottom).

(F) Live-cell transcriptomes for biological samples without TNF-α stimulation highlighting differentially expressed genes (grey) and TNFAIP3, BIRC3, SOD2, INHBA, CXCL8, WTAP (blue).

(G) Live-cell transcriptomes for biological samples with TNF-α stimulation highlighting differentially expressed genes (grey) and TNFAIP3, BIRC3, SOD2, INHBA, CXCL8, WTAP (orange).

(H) Gene set enrichment analysis (50 Hallmark gene sets) on lysates and supernatants from self-reporting cells (stimulated vs. unstimulated).

Next, we compared the expression of genes identified as differentially regulated in standard cell lysate RNA-seq and live-cell RNA-seq of cell supernatants. Overall, live-cell RNA-seq data were concordant with the RNA-seq expression profiles of control cell lysate samples (Figure 4.3E). This indicated that VLP-derived RNA collected from the supernatants of self-reporting cells faithfully carried transient biological information (Figure 4.3F-G). We measured upregulation of genes known to be involved in TNF-α signaling, notably TNFAIP3, BIRC3, SOD2, INHBA, and CXCL8[210,211]. Interestingly, WTAP, a component of the WTAP-METTL3-METTL14 N-6-methyltransferase complex catalyzing m6A generation in mRNA[212], was also upregulated in self-reported RNA as well as cell lysates. We next performed gene set enrichment analysis (GSEA) to investigate whether cellular self-reporting accurately reported biological pathway activity[169]. We ran GSEA for the 50 Hallmark gene sets on RNA-seq data from the supernatant of TNF-α-stimulated and unstimulated self-reporting cells at each time point. The top enriched gene set was TNF-α signaling via NF-κB, confirming that dynamic activity of biological pathways can be non-destructively resolved in living cells by RNA-seq through the engineered self-reporting approach (Figure 4.3H). Altogether, these results demonstrate that VLP-based self-reporting enables live-cell transcriptomics, which can capture both transient transcriptional responses and meaningful biological pathway information from the same living cells over time.

4.3 Discussion

Export of RNA via engineered VLPs makes live-cell transcriptomics possible through the cellular self-reporting paradigm. Live-cell RNA-seq enables new categories of experiments in which the transcriptome of a living cell sample is monitored over time without sampling the cells themselves. However, there are some limitations of our current protocol. Self-reporting requires collection of cell-free liquid samples at each time point of interest. This sets up a tradeoff between time resolution/sensitivity and alteration of the conditions experienced by the cells that is set by the amount and frequency of supernatant sampling. Further, while fold-changes in gene expression detected in self-reported RNA were strongly correlated to fold-changes in cellular lysates, they do not correspond exactly, which complicates quantitative comparisons across live-cell and destructive RNA-seq studies (Supplementary Figure 4.13D). We note that the self-reporting data from VLPs presented here represent a smaller quantity of RNA per sample per time point than is available from lysing the self-reporting cells themselves. As a result, there is less statistical power for differential gene expression calls from a single time point relative to a lysate from a similar sample.

The genetically encoded format of the live-cell self-reporting system could ultimately enable targeted longitudinal transcriptomic measurements of cells, tissues, and organs in their physiological and spatial context. For example, the expression of self-reporting machinery could be driven by cell-type-specific promoters in transgenic animals to measure the transcriptomes of target cells in their native environment. Future refinements may lead to increased export rates, thereby enabling improvements to sensitivity and time resolution, as well as single-live-cell transcriptomics experiments. The self-reporting approach may also be adapted for non-destructive proteomic and metabolomic measurements. Altogether, the results of this study demonstrate that cellular self-reporting with engineered VLPs allows longitudinal analyte monitoring in the same biological samples, circumventing the need for physical access to samples and the lysis of living samples to obtain molecular analytes while preserving the living samples for further study after the longitudinal molecular assays are complete.

4.4 Materials and Methods

Cell culture

HEK293FT (Thermo Fisher Scientific Cat. No. R70007) and HT1080 (ATCC CCL-121) were cultured in a humidified incubator at 5% CO_2 at 37˚C using high-glucose DMEM (Invitrogen) complemented with 10% FBS and 1% penicillin/streptomycin. Cell cultures were kept at low passage (<10) and regularly tested for mycoplasma contamination. iPS cells (1157-2 line, Boston Children's Hospital) were maintained in StemFlex media (Stem Cell Technologies), supplemented with 1% penicillin/streptomycin, and cultured on plasticware coated with growth factor-reduced matrigel (Corning) diluted in DMEM-F/12 according to manufacturer instructions. iPS cells were propagated as colonies using ReleSR (Stem Cell Technologies).

Cloning

All plasmid vectors were generated by Gibson Assembly using NEB's Gibson Assembly Master Mix, according to the manufacturer's protocol. The xenotropic MLV Gag ORF (UniProt Q27ID9) was synthesized by IDT as a gBlock and cloned in frame with a P2A-GFP sequence into the pcDNA3 backbone to generate the episomal expression vector, pxN01. As a negative control for VLP formation, we also created a truncation mutant with the MA domain of Gag (amino acids 2-130) removed (termed Gag∆MA) to abolish VLP budding and maturation on lipid membranes. The pLenti backbone was used to generate the lentivirus vector pLV_CSR9 utilizing a CAG promoter to drive expression of a Gag-P2A-GFP-P2A-zeoR ORF. A dox inducible Gag vector (pLxN01) was generated by Gibson Assembly using the pCW57.1 vector. Vectors expressing Gag fusion proteins were generated by cloning the minimal RNA recognition motifs of RNA binding proteins (gBlocks synthesized by IDT) to the C-terminus of Gag with a 5X serine-glycine linker.

Epitope-tagged VSV-G constructs were generated utilizing the VSV-G ORF from pMD2.G (a gift from D. Trono) to insert a FLAG (DYKDDDDK) or HA (YPYDVPDYA) sequence following the signal peptide at amino acid position 27. The epitope-tagged ORFs were cloned into pcDNA3 or pLenti backbones with a 2A-mCherry-2A-puromycin cassette to derive episomal or lentivirus vectors, respectively. Finally, piggyBac vectors for expression in iPS cells were generated with a piggyBac transposon backbone (System Biosciences) containing a CBX3 ubiquitous chromatin opening element (UCOE) upstream of an EF1-a promoter driving expression of Gag or Gag∆MA in frame with a P2A ribosomal skipping peptide and mNeonGreen reporter.

Lentivirus production

HEK293FT cells were seeded at 1 × 10^6 cells/well in 6-well plates. The following day, cells were transfected when 90-95% confluent with pMD2.G (Addgene #12259), psPAX2 (Addgene #12260), and a lentiviral transfer plasmid (2:3:4 ratio by mass) using Lipofectamine 2000 (Thermo Fisher). Media was exchanged after 6 hours and viral supernatant was harvested 48 hours after transfection and filtered through 0.45 μm cellulose-acetate filters (VWR cat. no. 28145-481).

Generation of stable MLV Gag-expressing cells

Stably self-reporting HEK293FT and HT1080 cells were generated through sequential lentiviral transductions to first introduce a constitutive Gag expression vector followed by a constitutive epitope-tagged VSV-G expression vector. Cells were transduced by adding an appropriate amount of processed viral supernatant supplemented with polybrene (8 ug/mL) to the cellular media to achieve ~30% fluorescent reporter-positive cells for single-copy integrations. The cells were incubated in virus for 24 hours followed by antibiotic selection 48 hours post-transduction at the following concentrations: 1 μg/mL puromycin (Thermo Scientific A1113802), and 300 μg/mL zeocin (Thermo Scientific R25001). Stable Gag-expressing iPS cells were generated through piggyBac transposition to achieve long-term transgene expression and avoid lentiviral silencing. iPS cells were single cell dissociated with TrypLE reagent (Thermo Scientific) and 800,000 cells were nucleofected with 2.5 μg of Super piggyBac Transposase (SBI cat. no. PB200A-1) and 10 ug of transposon plasmid using Lonza Cell Line Nucleofector Kit V (Lonza cat. no. VCA-1003) on an Amaxa IIb Nucleofector with program B-016. iPS cells post-nucleofection were cultured with 1 μM Y-27632 ROCK inhibitor (Stem Cell Technologies) for 24 hours and FACS sorted one week later to enrich for mNeon+ cells.

Transient transfection of MLV Gag

HEK293T were seeded into 6-well plates at a density of 100,000 cells/cm^2. The next day, cells were transfected with 2000 ng of total DNA comprising a Gag expression plasmid, VSV-G expression plasmid and pUC19 plasmid (2:3:4 ratio by mass) using Lipofectamine 2000 (Thermo Fisher). The media was changed 6 hours post-transfection and was collected for downstream processing 48 hours post-transfection.

VLP purification from cellular media

VLP-containing cellular media from transient transfection experiments or stable Gag-expressing cells was first processed to remove cellular debris by centrifugation at 2,000 g for 10 minutes at 4˚C. Cleared cellular media from iPS cells was treated with 200 U/mL benzonase nuclease (Sigma cat. no. E1014) at 37˚C for 60 minutes with mixing every 15 minutes to degrade free ribosomal RNAs and background RNAs. Next, the media was further filtered through a 0.45 μm cellulose acetate filter (VWR cat. no. 28145-481) and subsequently concentrated by centrifugation with a 100 kDa Amicon cutoff filter (Millipore Sigma cat. no. UFC5100) at 2500 g for 30 minutes at 4˚C. The retentate within the filter was either frozen at -80˚C for storage, used directly as a supernatant control for downstream assays, or used as an input for immunoprecipitation-based isolation of VLPs. For FLAG-based immunoprecipitation of VLPs, 20 µL of Anti-FLAG M2 Magnetic Beads (Sigma cat. no. M8823) were used per sample and washed 3x with TBS buffer. The beads were resuspended in 500 uL TBS + 1% Tween-20 and incubated with the VLP retentate on a rotisserie at 4˚C overnight. The next day, the beads were washed 3x with TBS + 1% Tween-20 and eluted in 500 ng/µL 3x FLAG peptide (Sigma cat. no. F4799) at 1200 RPM shaking for 30 minutes at 4˚C. For HA-based immunoprecipitation of VLPs, 50 µL of His-Tag Dynabeads (Invitrogen cat. no. 10103D) were used per sample according to the manufacturer's protocol. HA-tagged VLPs were eluted in 50 µL of elution buffer (300 mM imidazole, 50 mM sodium phosphate, 300 mM NaCl and 0.01% Tween-20) at 1200 RPM shaking for 30 minutes at 4˚C. The eluted VLPs were used directly in downstream assays or stored at -80˚C.

RT-qPCR of VLPs

HEK293T cells stably transduced with lentivirus packaging inducible Gag (pLxN01) were cultured in T75 culture flasks for 72 hours ± dox (10 ug/mL) before 10 mL of media was harvested from the flasks. The supernatant was stored at 4˚C for 24 hours before centrifugation at 600 g for 10 minutes at 4˚C to pellet any floating cells that were inadvertently collected with the cell media. The top 9 mL of media were then taken, and centrifuged at 2000 h for 10 minutes at 4˚C to pellet cellular debris. The top 8.5 mL of media were then collected and filtered through a 0.45 µm cellulose-acetate syringe filter (VWR cat. no. 28145-481). 8 mL of filtrate was collected and stored at 4˚C overnight before ultracentrifugation. The filtrate was prepared for ultracentrifugation using 31.5 mL tubes (Beckman Coulter cat. no. 358126) with 1 mL of 70% sucrose cushion followed by 5 mL of 20% sucrose, and filtrate diluted in 1x PBS to bring to a total of 30 mL. The prepared tube was then ultracentrifuged (Thermo (Sorvall) WX80 Ultra-Centrifuge) with a swinging-bucket rotor (AH-629 (36 mL), Thermo cat. no. 54284) at 26,000 RPM for 2 hours at 4˚C. The interface between the 70% and 20% sucrose was collected with a syringe and then frozen at -80˚C. For RT-qPCR, the purified supernatant was thawed and 5 µL was taken per replicate and combined with 5 µL of 2xTCL (Qiagen cat. no. 1070498) for lysis. Lysed supernatant was then treated with 22 µL of RNAClean XP (Beckman Coulter cat. no. A63987) and purified according to standard protocol, rinsing with 100 µL of 80% EtOH. The magnetic beads were eluted in qScript mix (Quanta Bio cat. no. 95047-025) and cDNA synthesis was performed according to the manufacturer's protocol. qPCR was performed on 1 µL of cDNA with JumpStart Taq ReadyMix (Millipore Sigma cat. no. P2893-100RXN) using Rox as a reference dye (Millipore Sigma cat. no. R4526-5ML), primers spanning GAPDH exon junctions (Fwd: gaaggctggggctcatttgc, Rev: gGaggcattgctgatgatct) and a custom TaqMan probe (Seq: atctctgccccctctgctgatg, ordered as a

FAM TaqMan probe from IDT). qPCR was conducted on a Stratagene Mx3000P qPCR System real-time PCR.

RNA sequencing

RNA from cells, supernatant, or IP-purified VLPs was extracted using 2X TCL lysis buffer (Qiagen cat. no. 1070498). At least two technical replicates were prepared per sample using the SMART-Seq2 protocol as previously published[206] with some modifications. Briefly, the lysed samples were 2.2X RNA SPRI (Beckman Coulter cat. no. A63987) cleaned and reverse transcribed in the presence of a template switching oligo (Exiqon) with Maxima RNase H-minus RT (Thermo Fisher Scientific cat. no. EP0751) using a polyT primer containing the ISPCR sequence. Whole transcriptome amplification proceeded with KAPA HiFi HotStart ReadyMix using an ISPCR primer according to the following thermal program: 98˚C for 3 minutes, 27 cycles of 98˚C for 15 seconds, 67˚C for 20 seconds, and 72˚C for 6 minutes, and a final extension step of 72˚C for 5 minutes. The amplified cDNA was cleaned with 0.8X DNA SPRI beads (Beckman Coulter cat. no. B23318). Ten nanograms of DNA was tagmented at 58˚C for 10 minutes in a 10 µL reaction containing 2 µL of 5X tagmentation buffer (50 mM Tris-HCl, 25 mM MgCl2 pH 8.0), 2 µL of Tris Buffer (10 mM Tris-HCl, 1% Tween-20 pH 8), and 4 µL Nextera (Illumina). The reaction was stopped with 1% SDS and incubated at 72˚C for 10 minutes, then 4˚C for 3 minutes. The tagmented library was cleaned with 1X DNA SPRI beads followed by 12 cycles of PCR with NEBNext High Fidelity polymerase to incorporate sample index barcodes and Illumina flow cell handles. The final libraries were pooled, diluted and sequenced on a NextSeq-500 (Illumina) in paired-end mode using a 75-cycle High Output Kit v2.

Western blot

Cells were washed with PBS and protein lysate was extracted using an extract solution (150 mM NaCl, 1% Triton X-100, 50 mM Tris-HCl pH 7.5) supplemented with a protease inhibitor (Sigma cat. no. 4693159001). The cellular lysate was incubated on a rotisserie for 30 minutes at 4˚C, followed by centrifugation at 10,000 g for 5 minutes at 4C. The protein lysate, IP-purified VLPs, or cleared cellular media was then used directly for Western-blotting. Samples were reduced using 2X Tris-Glycine SDS Sample Buffer (Life Technologies cat. no. LC2676) and 10X NuPAGE Reducing Agent (Life Technologies cat. no. NP0009) at 98˚C for 5 minutes.

The reduced samples were then run on a Novex WedgeWell 8-16% Tris-Glycine Gel (Life Technologies cat. no. XP08165BOX) for 50 minutes at 225V. The gel was then transferred to a PVDF membrane (Life Technologies cat. no. IB24002) using an iBlot 2 device (Life Technologies). The PVDF membrane was blocked in 5% milk (Bio-Rad cat. no. 1706404XTU) for 1 hour at room temperature with gentle shaking, followed by an overnight incubation at 4C with gentle shaking in primary antibodies to Gag at 1/2000 dilution (Abcam cat. no. ab100970) and to actin at 1/5000 dilution (Abcam cat. no. ab179467). The next day the membrane was washed 3x with 5% milk and incubated for 4 hours at 4˚C in an anti-rabbit secondary antibody (Sigma cat. no. 41176). The membrane was imaged on an Azure Biosystems C600 Imaging System.

TNF-α time course

25,000 HT1080 cells were plated in TC-treated 96-well plates (VWR cat. no. 62406-117) with 100 µL of media per well. After every 12 hours, media was changed with pre-warmed media ± TNF-α at a final concentration of 30 ng/mL (Invivogen cat. no. rcyc-htnfa), introducing TNF-α at each time point. Destructive lysates were collected by aspirating media and lysing with 50 µL of 1xTCL (Qiagen cat. no. 1031576) and stored at -20˚C until the final time point.

Supernatants from self-reported cells were carefully collected (to not disturb adherent cells) and stored at 4˚C until the final time point. After the final time point, supernatants were centrifuged at 2000 g for 10 minutes at 4˚C and the top 20 µL of media was carefully collected (to not disturb adherent cells) and mixed with 20 µL of 2xTCL (Qiagen cat. no. 1070498) for lysis. Samples were collected every 12 hours throughout the time course. RNA-seq libraries were prepared from 20 µL of supernatant lysate and 10 µL of cell lysate using SMART-Seq2 with 27 WTA cycles for supernatant libraries and 21 WTA cycles for cell libraries.

RNA-Seq processing

RNA sequencing paired-end reads were pseudo-aligned using Kallisto (version 0.43.1)[213]. The hg19 cDNA fasta reference from UCSC was appended with the coding sequences of Gag, Gag∆MA, mNeon and VSV-G in order to generate a custom Kallisto index via the “kallisto index” command. Differential gene expression analysis was performed using DESeq2 (version 1.30.0) with R (version 4.0.3) on the estimated count matrix output from Kallisto. Gene set enrichment analysis was performed using GSEA (version 4.1.0) on TPM output from Kallisto. Downstream analysis was performed with custom python scripts (python versions 2.7 and 3). Sequencing reads were downsampled to match sequencing depth for all samples when producing plots showing the number of genes detected.

Co-culture analysis

RNA sequencing data was filtered for genes with TPM greater than 10, and then basis vectors were generated by finding genes that were exclusively detected in cellular lysates of HEK293T or HT1080 cells. Cell-type specific scores were calculated by taking the inner product of the binarized basis vectors, and the binarized RNA-seq expression vector for each sample. The relative score was then calculated by dividing by the sum of the inner products, to determine which inner product was stronger, thus inferring the sample of origin.

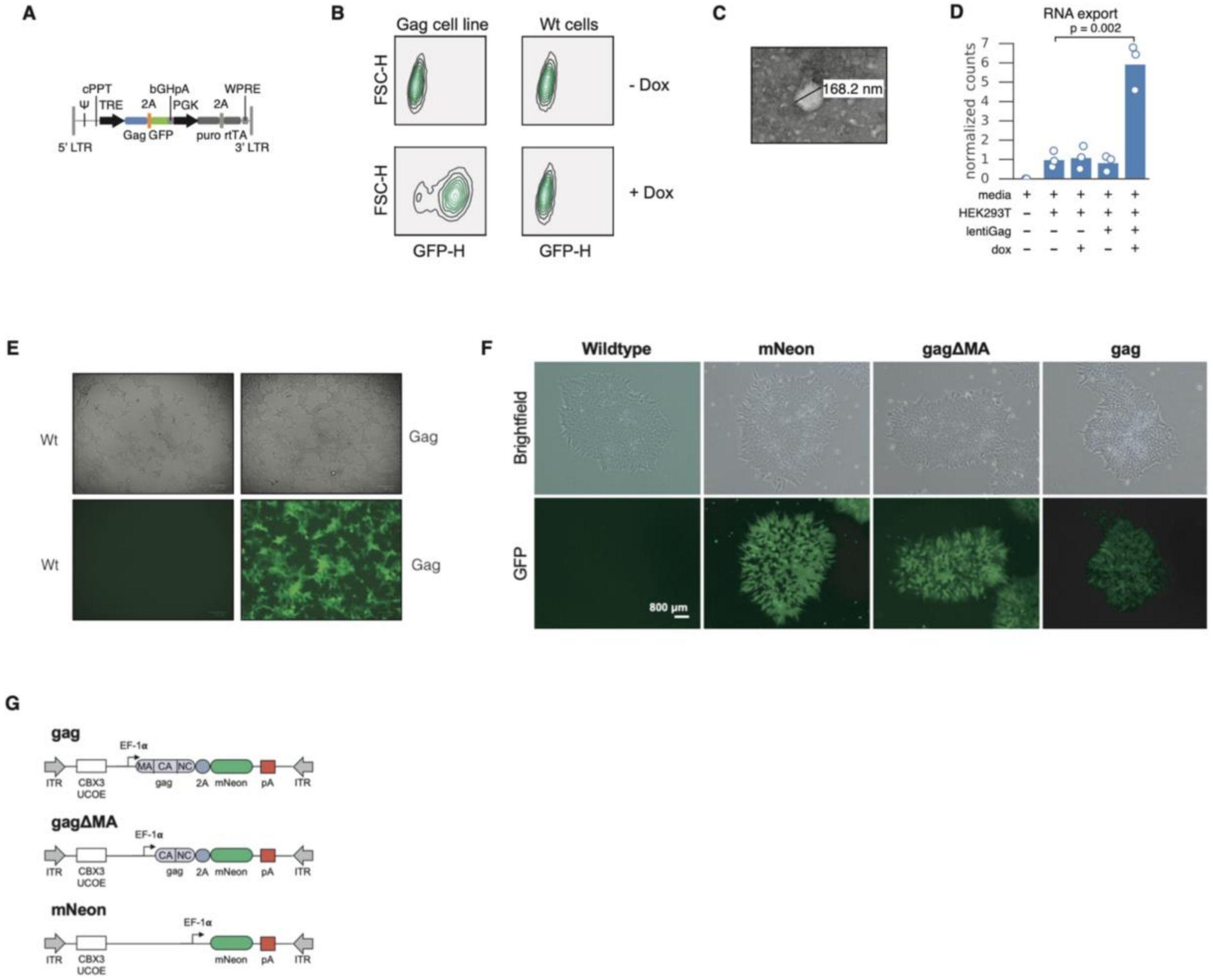
Predictive modeling and RNA localization analysis

A matrix of supernatant to lysate TPM ratios was log transformed, and then several features were selected to build a model for supernatant to lysate TPM ratio prediction. RNA location was selected as a feature, and the RNALocate[214] database was used to annotate subcellular localization. Other features were transcript length (including UTRs and CDS), GC content (%). Further, we treated homology (7-mer sequences) between each transcript and the MLV genome as a feature, to account for any structural or sequence specificity in RNA export. Briefly, we counted the matches of 7-mers from the MLV genome for each transcript, and then binned the location of the 7-mer into 100 discrete bins along the positional axis of the MLV genome. Using the measured log-transformed supernatant to lysate RNA-seq measurements from HEK293T and HT1080 cell lines, along with the features above, we split the data into a training and test set and constructed a gradient boosting regressor {'n_estimators': 500, 'max_depth': 30, 'min_samples_split': 10, 'learning_rate': 0.01, 'loss': 'ls','subsample': 0.1,'verbose':1, 'criterion':'friedman_mse','min_samples_leaf':2,'min_weight_fraction_leaf':0.0}. We then looked at feature importance scores to better understand RNA properties that would influence export bias.

Inference of gene regulatory networks with CellNet

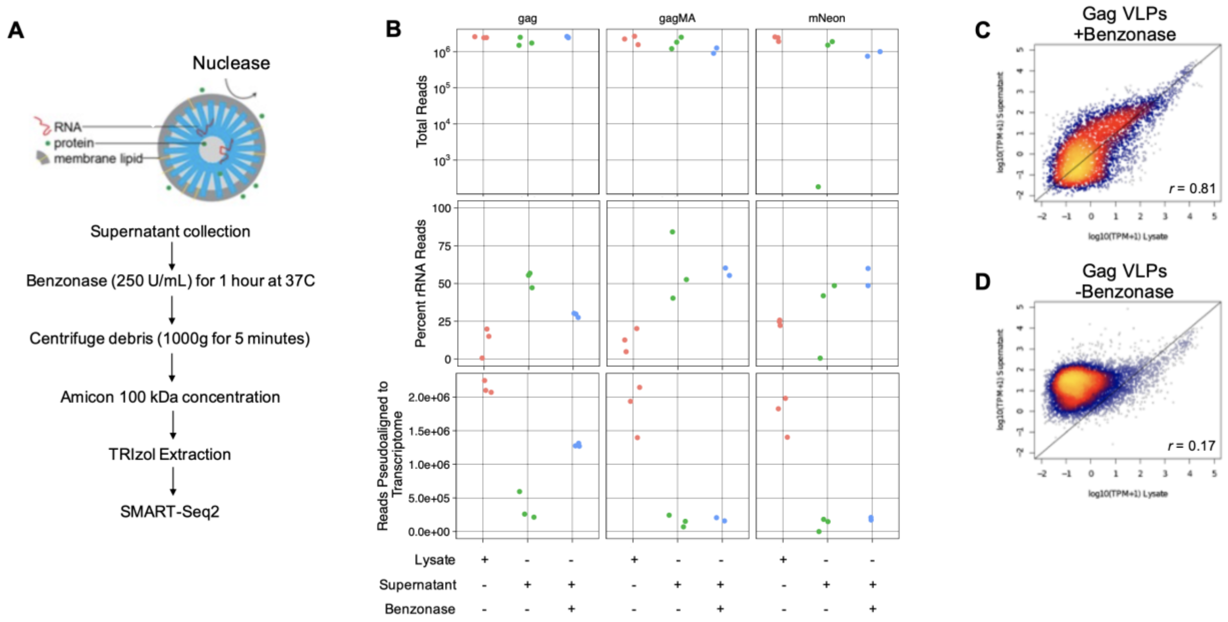
We utilized CellNet[207] to derive active gene regulatory networks (GRNs) from lysate and VLP RNA-seq samples and to assess the ability of VLP-derived RNAs to reflect the GRNs of the producing cell. We first retrained the CellNet classifier to include embryonic kidney and fibrosarcoma GRNs. We curated lysate RNA-seq samples from HEK293Ts and HT1080s and quantified transcript abundances with cn_salmon() using a pre-prepared Salmon transcript index, salmon.index.human.122116.tgz available from https://github.com/pcahan1/CellNet. We constructed new cell-type specific GRNs with cn_make_grn() using samples from the June 20, 2017 edition of the human CellNet Processor (https://github.com/pcahan1/CellNet) and our HEK293T and HT1080 samples. We assessed the random forest classifier using cn_splitMakeAssess() and generated a new CellNet processor object with cn_make_processor(). We then applied CellNet with this retrained classifier using default settings to lysate and VLP RNA-Seq samples and plotted the sample classification scores as a heatmap in R.

4.5 Supplementary Figures



Supplementary Figure 4.1 | Characterization of stably integrated Gag+ cell lines.

(A) Lentivirus construct for dox-inducible self-reporting.
(B) Flow cytometry on dox-inducible Gag+ HEK293T cell lines generated with lentiviral transduction.
(C) Negative stain electron micrograph showing a VLP.
(D) GAPDH RT-qPCR results from supernatant purified from wild-type and Gag+ HEK293T cell lines ± doxycycline. GAPDH signal was used as a proxy for exported RNA. Doxycycline induction led to VLP formation and RNA export.
(E) HEK293T cell morphology with Gag expression.
(F) Brightfield and GFP images of iPS cells stably transposed with Gag, GagΔMA, and mNeon piggyBac constructs.
(G) PiggyBac expression vector diagram.

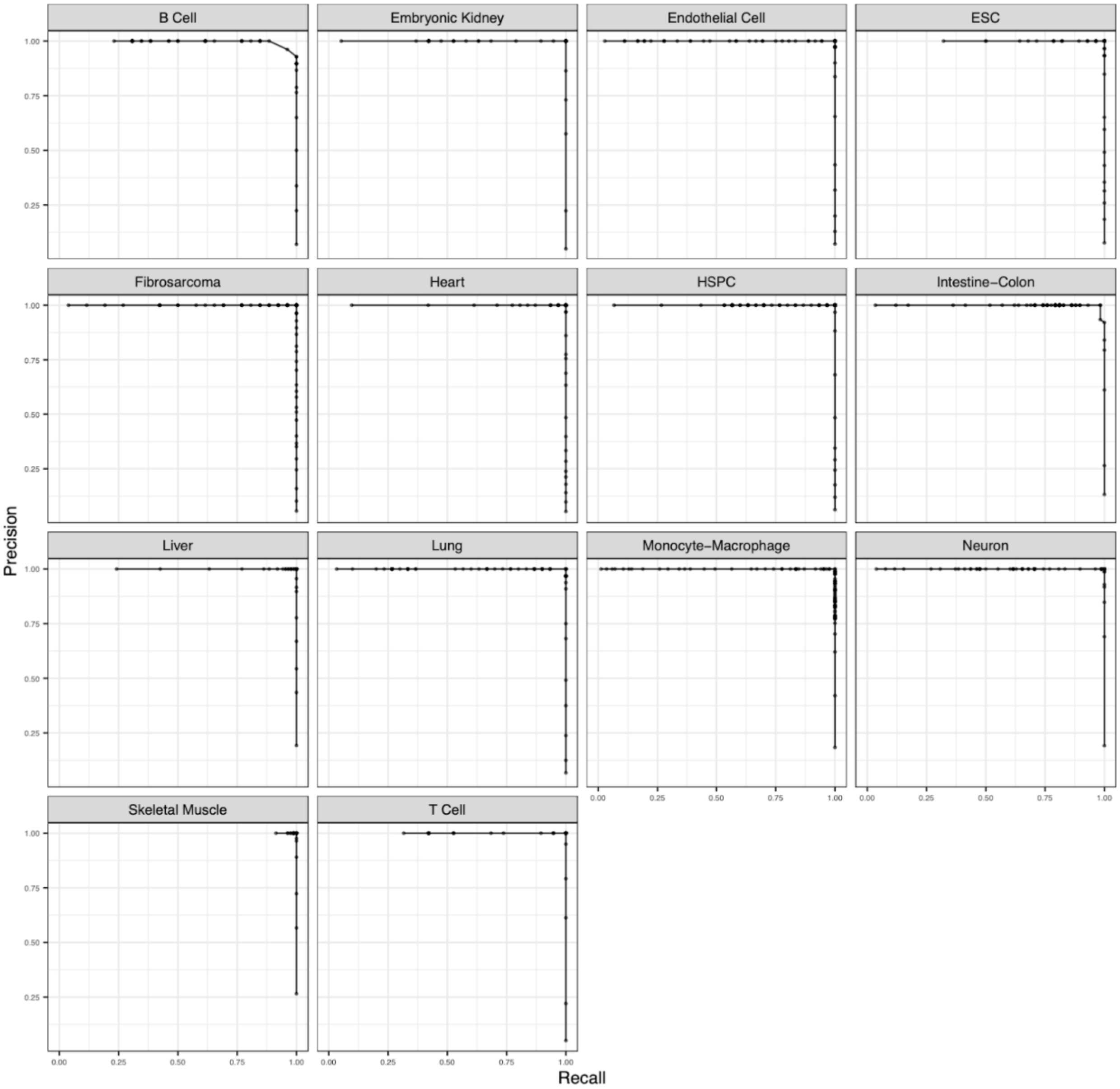


Supplementary Figure 4.2 | Characterization of RNAs packaged in VLPs via nuclease digestion of cellular media.

(A) VLP processing workflow for iPS cells.

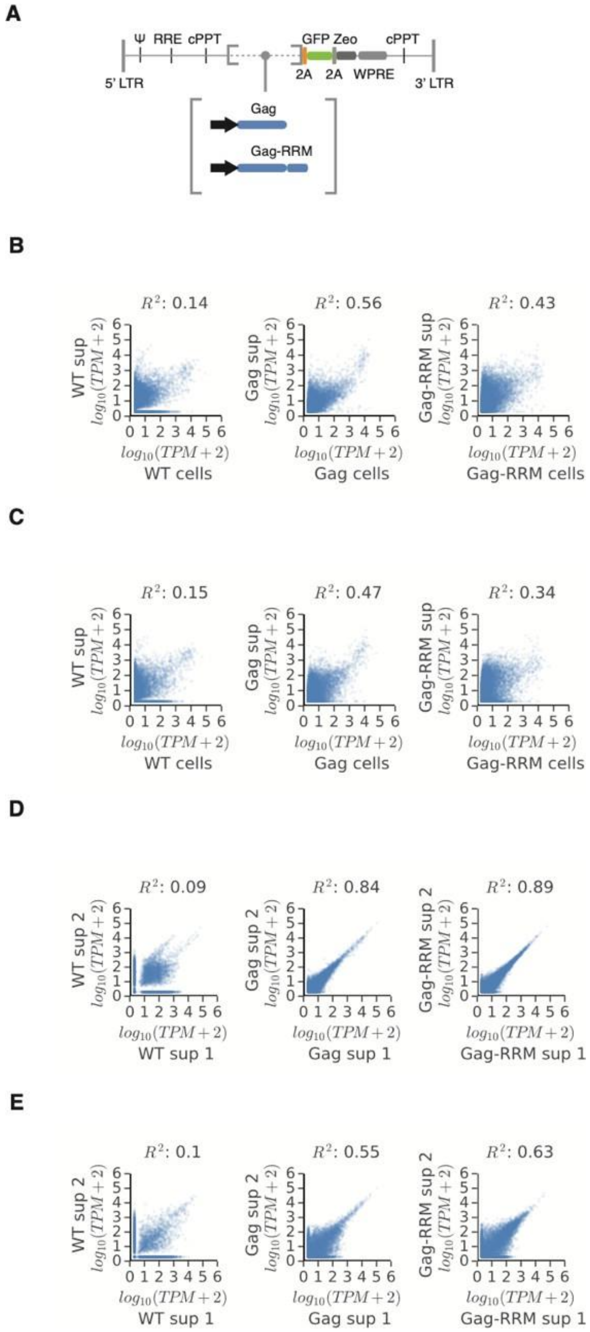
(B) Plot of RNA-Seq reads +/- benzonase treatment on media from Gag, Gag∆MA, and mNeon expressing iPS cells to provide evidence that transcriptional signal is dependent both on Gag expression and the formation of exported VLPs able to protect RNA cargo from enzymatic degradation (reads pseudoaligned to the human transcriptome).

(C, D) TPM concordance between supernatant and cell lysate +/- benzonase treatment from Gag expressing cells. As expected, RNA representation is superior by Pearson correlation, since low abundance transcripts are less likely to be detected and TPMs of higher abundance transcripts drop in the presence of excess rRNA reads.



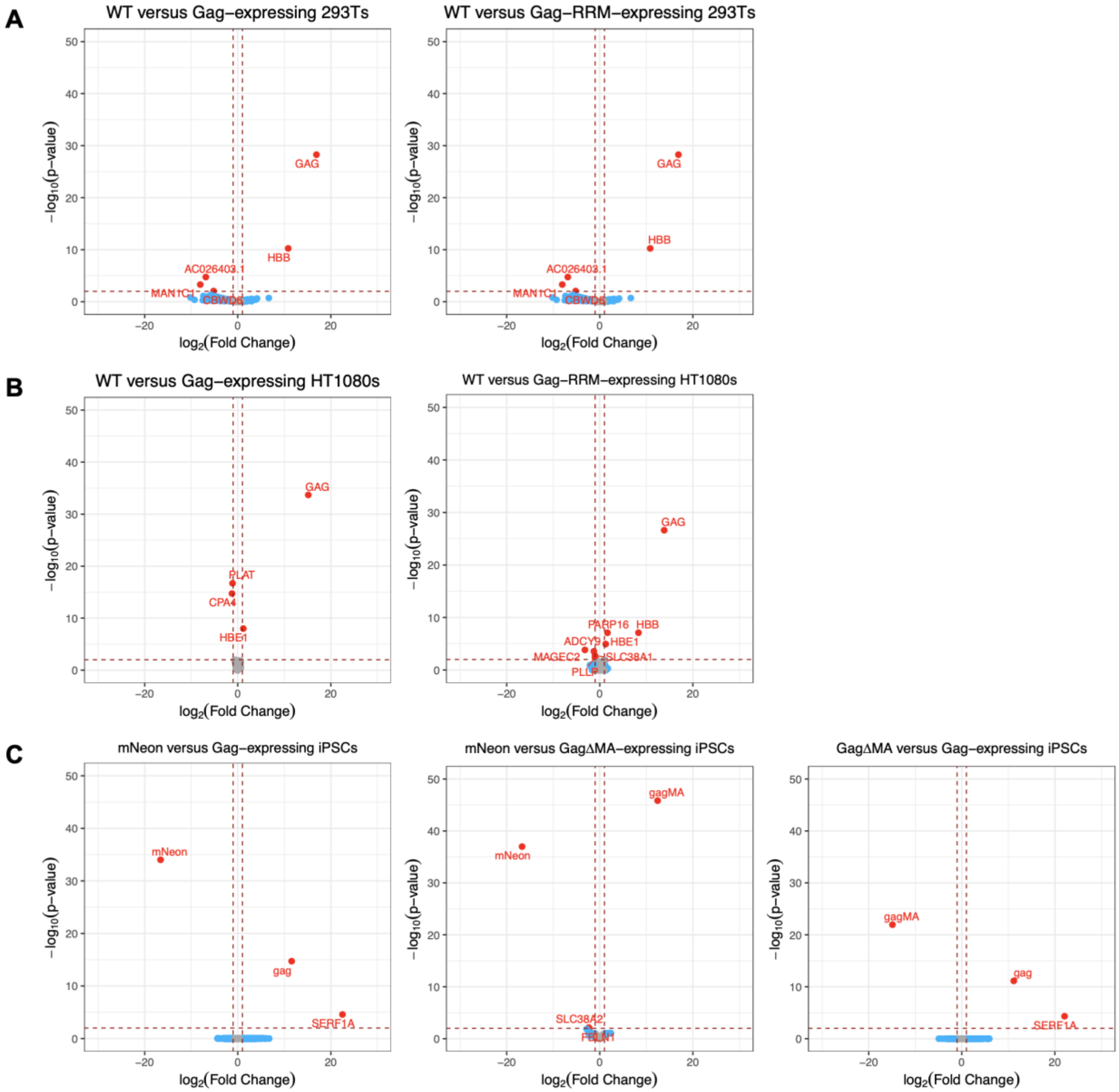
Supplementary Figure 4.3 | CellNet training.

Classification performance of a modified CellNet[207] random forest classifier trained on an expanded compendium of human cell types (including embryonic kidney and fibrosarcoma cell types, corresponding to HEK293T and HT1080 cells, respectively). Precision-recall curves show excellent performance for each cell-type classifier trained on human RNA-seq data.



Supplementary Figure 4.4 | Gag fusion characterization with stable, single-copy integrated cell lines.

(A) Lentiviral constructs for self-reporting.
(B) RNA representation from wild-type HEK293T and HEK293T expressing different Gag constructs
(C) RNA representation from wild-type HT1080 and HT1080 expressing different Gag constructs.
(D) RNA-seq biological replicates of supernatant lysates from wild-type HEK293T and HEK293T expressing different Gag constructs.
(E) RNA-seq biological replicates of supernatant lysates from wild-type HT1080 and HT1080 expressing different Gag constructs.

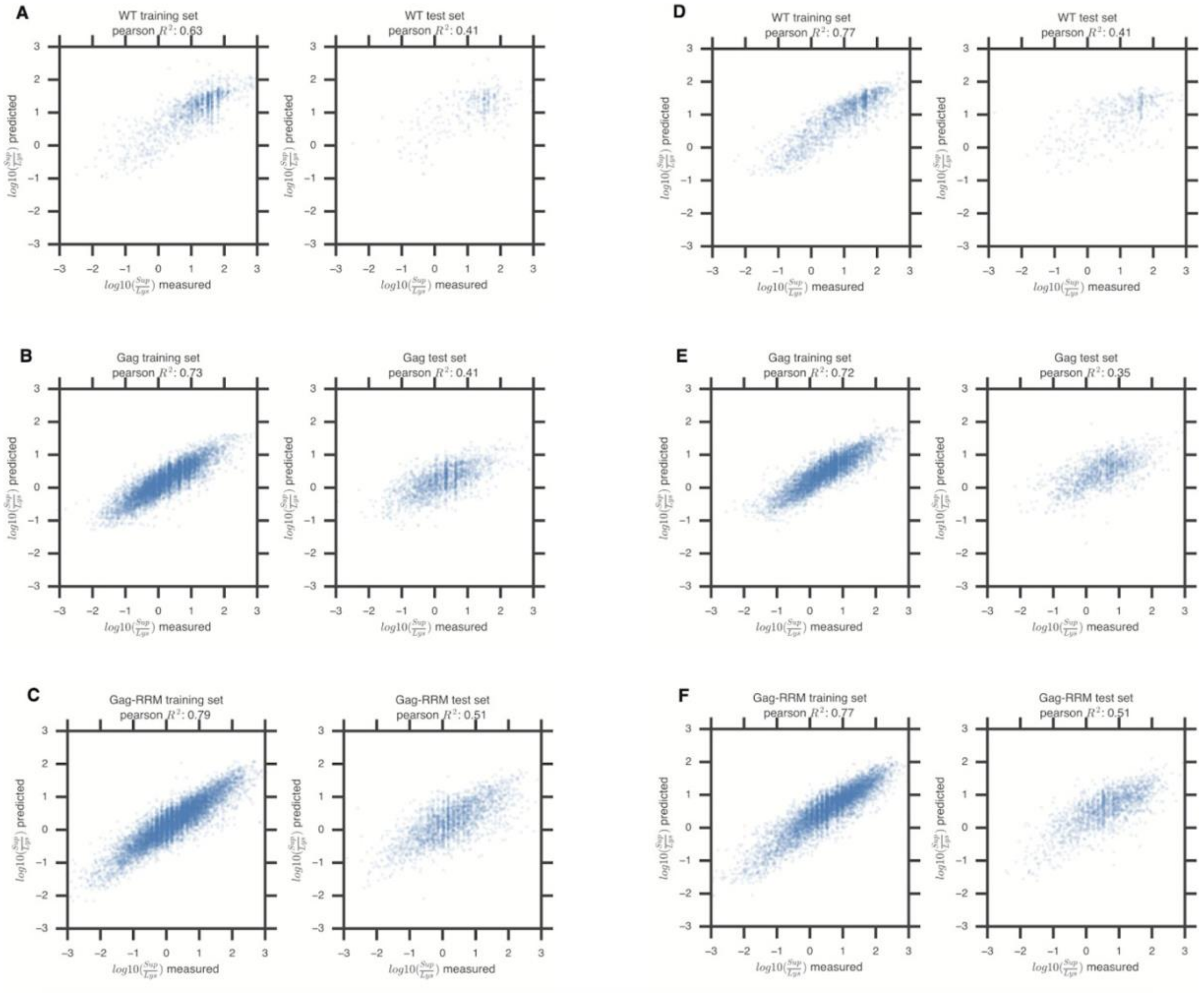


Supplementary Figure 4.5 | Differentially expressed genes in self-reporting HEK293T, HT1080, and iPS cells.

Using DESeq2, we compared various self-reporting cell types containing genomically integrated self-reporting constructs to cell type negative controls to identify differentially expressed genes (DEGs) induced from self-reporting. RNA-Seq reads from cellular lysate were mapped to a modified human transcriptome containing the exogenous transgenes.

(A) HEK293T cells and (B) HT1080 cells expressing Gag or Gag-RRM were compared to wild-type cell controls.

(C) iPS cells expressing Gag or the VLP-deficient control GagΔMA were compared to mNeon-expressing negative controls cells. DEGs were defined as log2(fold-change) > 1 or log2(fold-change) < -1 and FDR < 0.01, and are noted in red in the volcano plots. GagΔMA is depicted as gagMA.



Supplementary Figure 4.6 | Gradient-boosted tree regression performance for predicting ratios of self-reported RNA to lysate RNA.

(A) Model performance for supernatant to cell lysate ratio predictions (training and test sets) for genes in wild-type HEK293T cells.

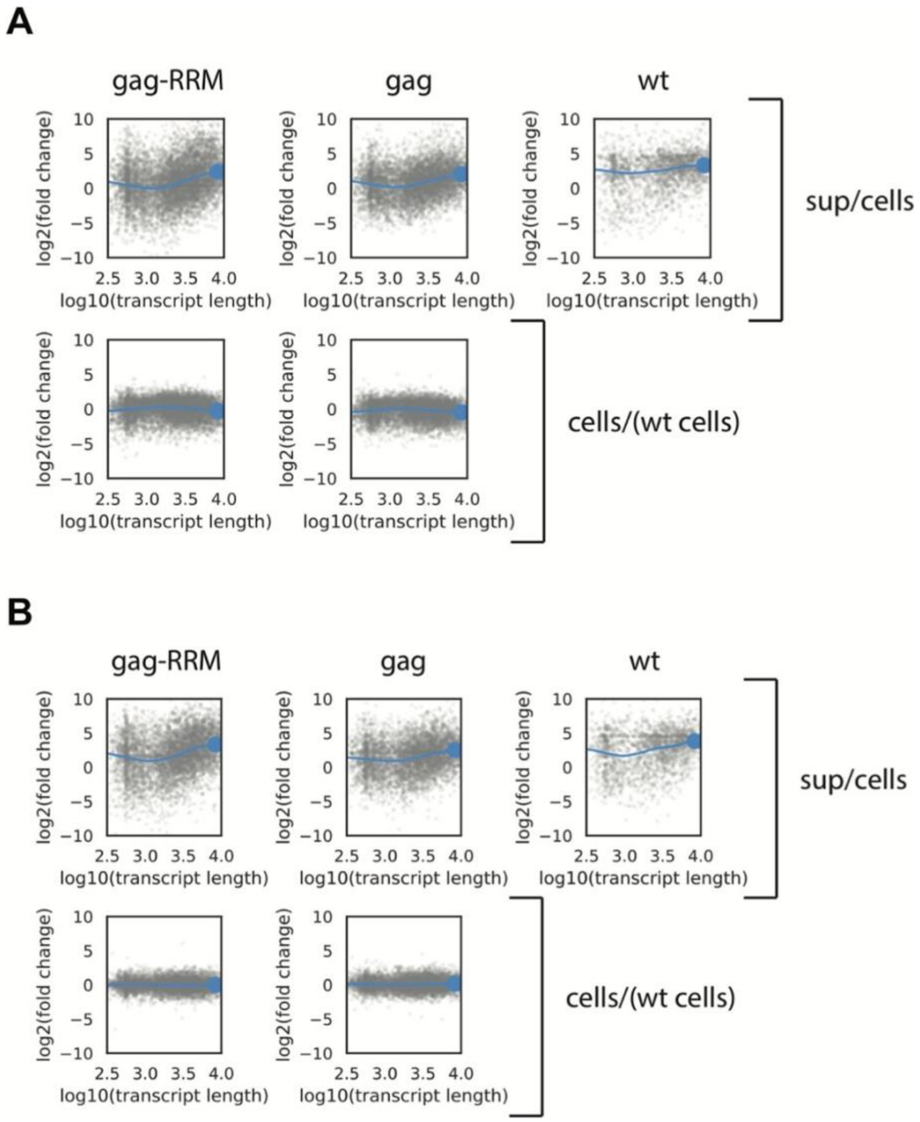
(B) Model performance for supernatant to cell lysate ratio predictions (training and test sets) for genes in Gag+ HEK293T cells.

(C) Model performance for supernatant to cell lysate ratio predictions (training and test sets) for genes in Gag–RRM+ HEK293T cells.

(D) Model performance for supernatant to cell lysate ratio predictions (training and test sets) for genes in wild-type HT1080 cells.

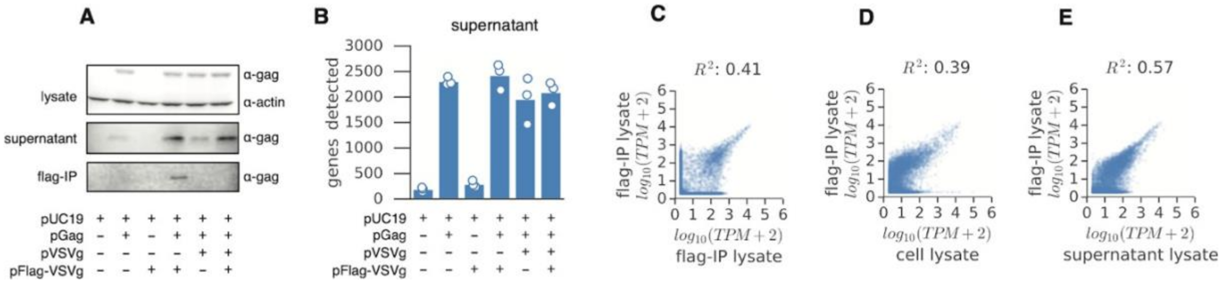
(E) Model performance for supernatant to cell lysate ratio predictions (training and test sets) for genes in Gag+ HT1080 cells.

(F) Model performance for supernatant to cell lysate ratio predictions (training and test sets) for genes in Gag–RRM+ HT1080 cells.



Supplementary Figure 4.7 | Long RNA transcripts are preferentially packaged in VLPs.

(A) Packaging size preference for HEK293T cells (lowess trendline in blue, depicting MLV genome size with a large blue dot).
(B) Packaging size preference for HT1080 cells (lowess trendline in blue, depicting MLV genome size with a large blue dot).



Supplementary Figure 4.8 | Purification and characterization of VLPs with engineered envelopes.

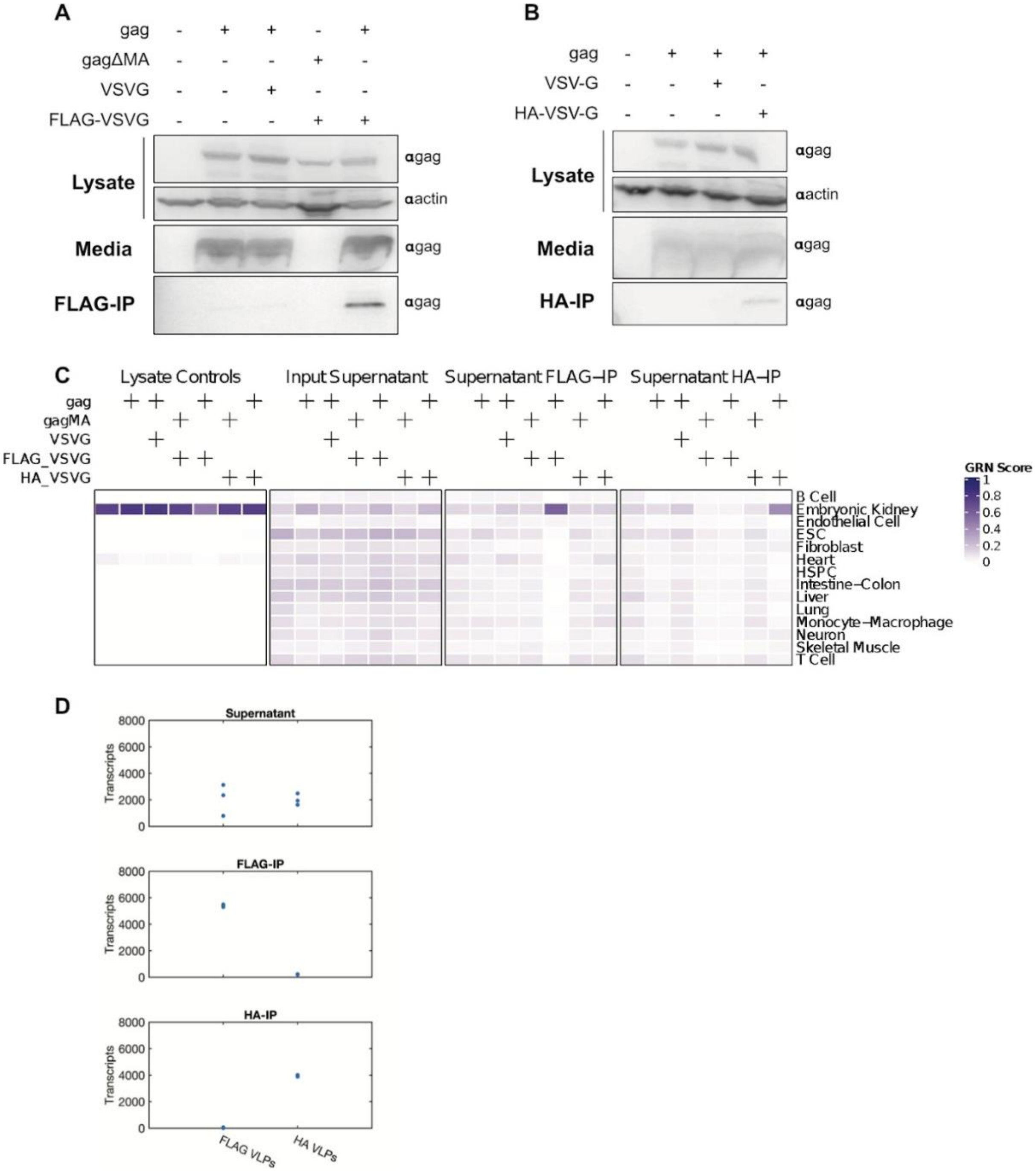
(A) Western blot on lysate, supernatant, and FLAG immunoprecipitation from HEK293T cell lines transfected with different constructs.

(B) RNA-seq of supernatant purified for immunoprecipitation input.

(C) RNA-seq replicate concordance of pGag+, pFLAG-VSV-G+ HEK293T supernatant after FLAG immunoprecipitation (FLAG-IP).

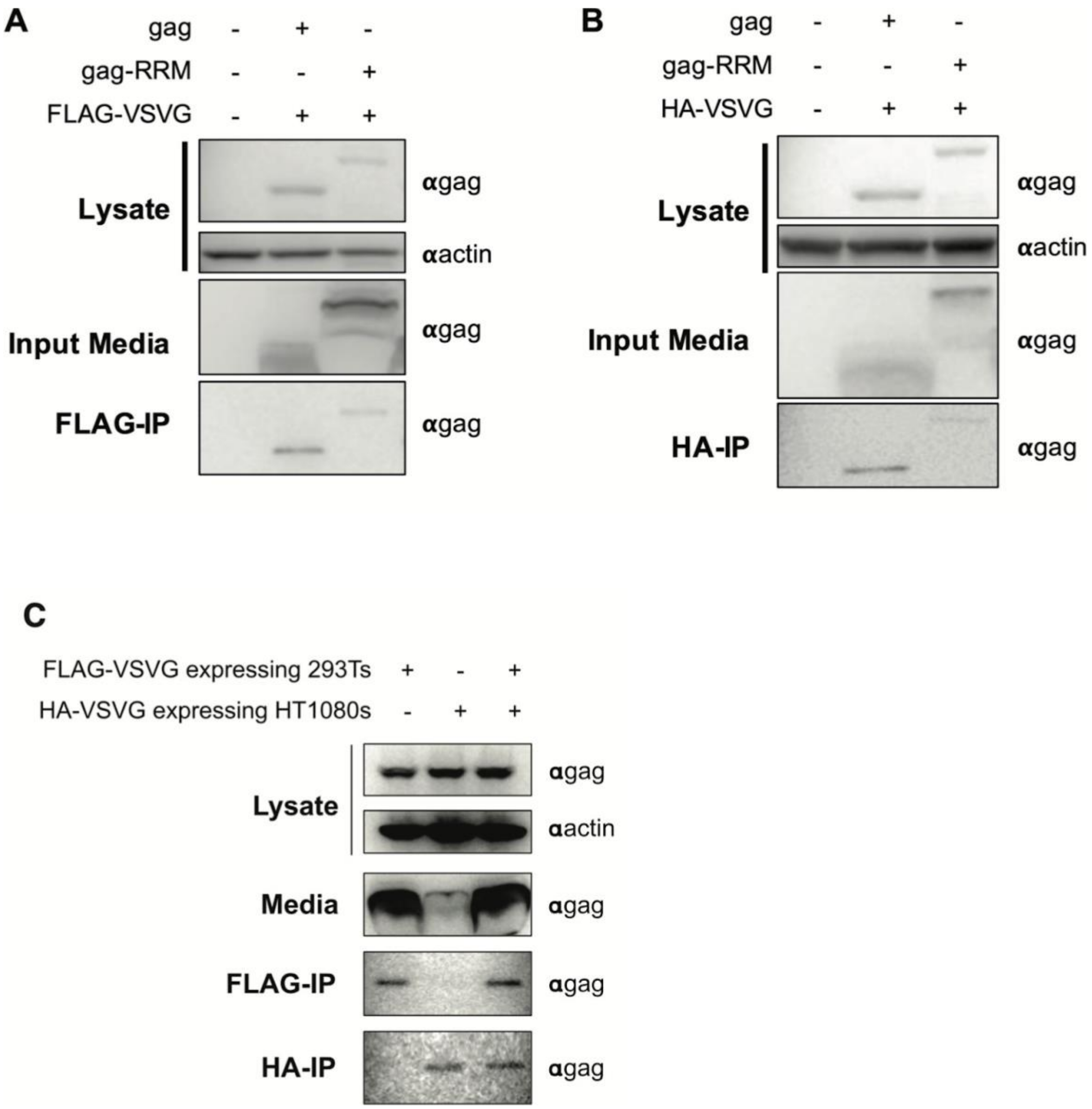
(D) Transcript abundances for FLAG-IP on self-reporting supernatant vs. cell lysate for self-reporting cells.

(E) Transcript abundances for FLAG-IP on self-reporting supernatant vs. input self-reporting supernatant.



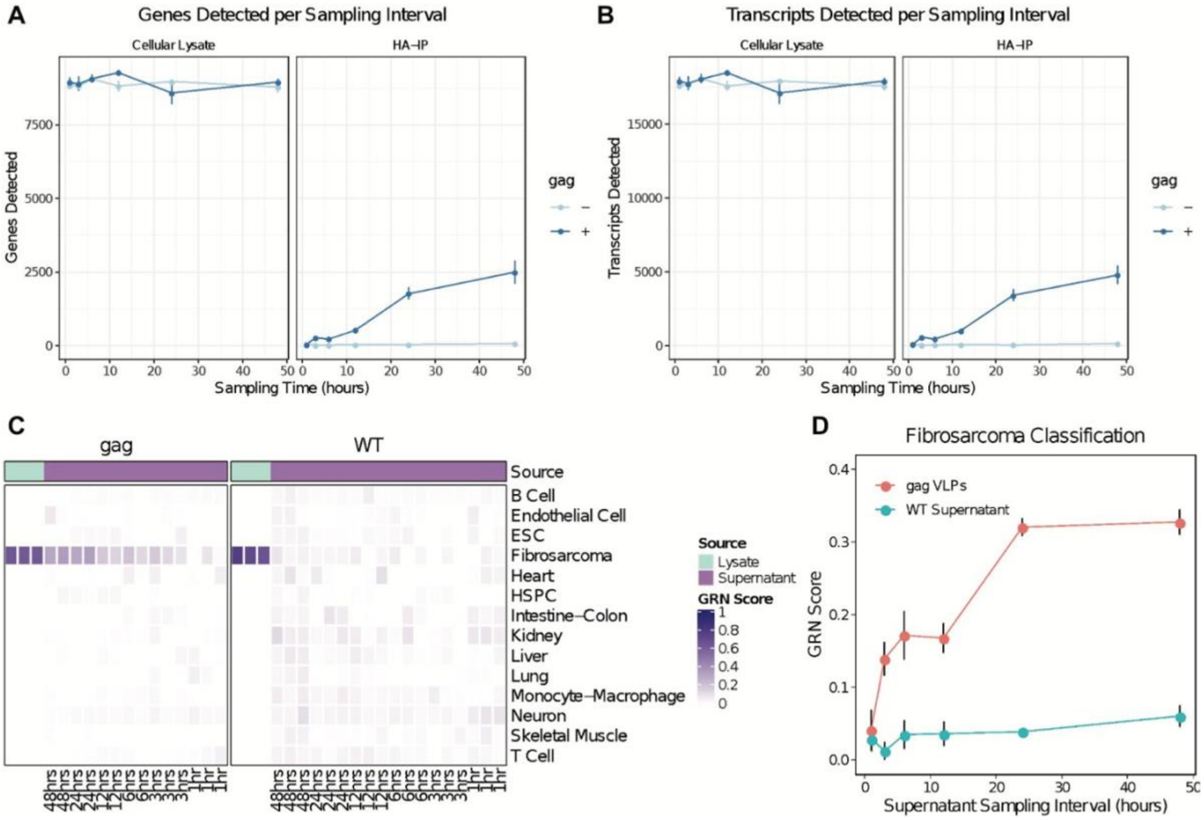
Supplementary Figure 4.9 | Specificity of multiplexed immunoprecipitation-based isolation of epitope-tagged VLPs from transfected HEK293T.

(A, B) Western blots of cellular lysate, supernatant, and IP-purified media for FLAG and HA-IP purification.
(C) CellNet classification of IP-purified VLPs from (A) and (B).
(D) FLAG- or HA-tagged VLPs produced from HEK293T cells were processed with FLAG and HA-IP to assess the specificity of the purification to the target epitope. RNA-seq libraries were prepared from the input supernatant and all IP outputs.



Supplementary Figure 4.10 | Constitutive VLP production from cells stably expressing engineered Gag polyproteins and epitope-tagged VSV-G proteins.

(A) Western blot from HEK293Ts demonstrates stable expression of Gag or Gag–RRM fusion proteins in addition to FLAG-VSV-G following lentiviral integration. Western blot for Gag on cellular media and FLAG-immunoprecipitated media demonstrates the constitutive production of FLAG-tagged VLPs.
(B) Same as (A) for HT1080s expressing HA-VSV-G.
(C) Co-culture experiment Western blot.



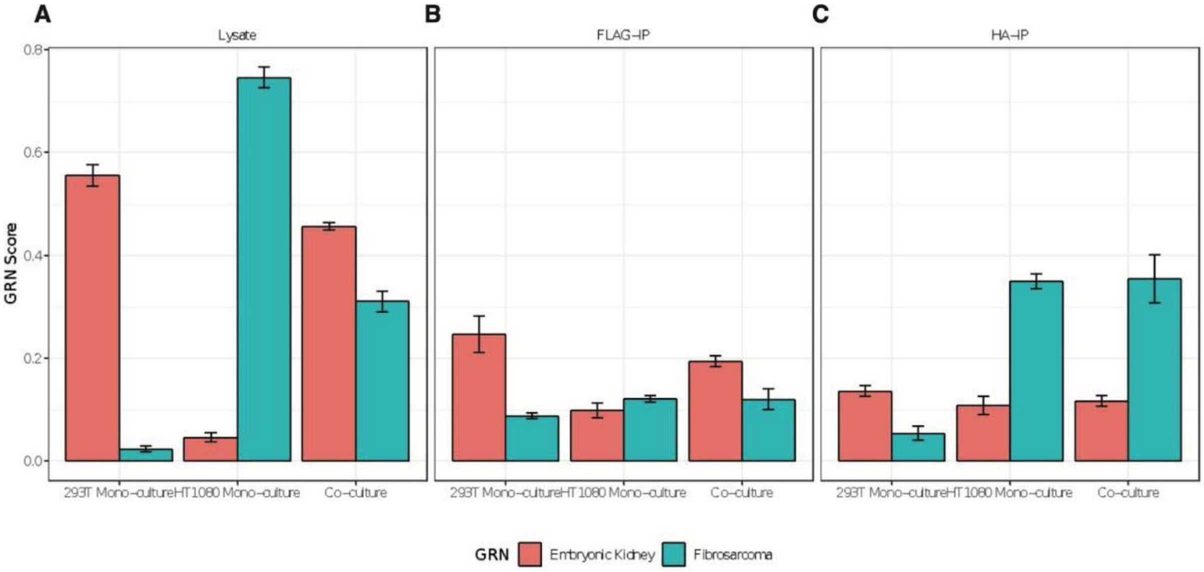
Supplementary Figure 4.11 | Characterization of the VLP-derived RNAs as a function of sampling duration.

(A) Genes detected as a function of length of sampling duration for HA-IP purified supernatants derived from Gag+ HT1080 cells.

(B) Transcript isoforms detected as a function of sampling duration (HA-IP purified VLPs and cell lysate controls) in HT1080 cells.

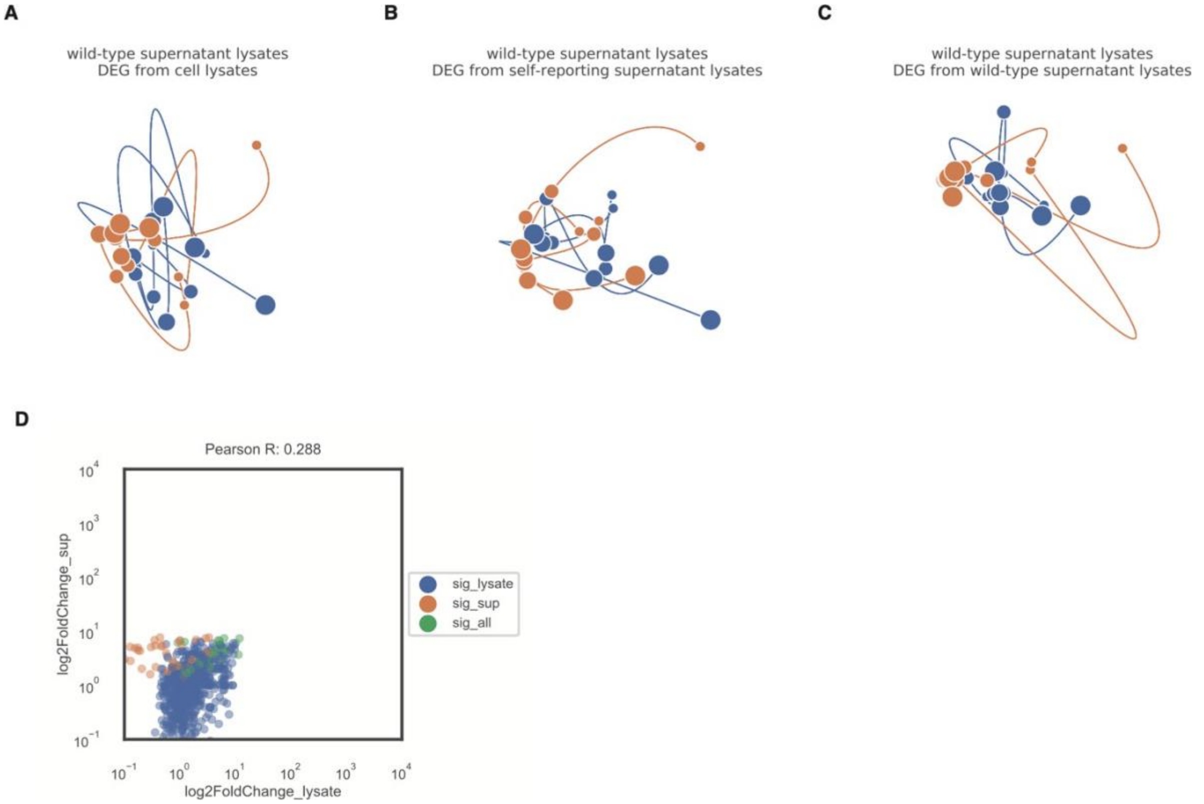
(C) CellNet classification per sampling interval length in HT1080 cells.

(D) Fibrosarcoma GRN score as a function of sampling duration in HT1080 cells.



Supplementary Figure 4.12 | CellNet classification of demultiplexed VLPs from HEK293T-HT1080 co-cultures.

(A) GRN scores for lysates from HEK293T (FLAG-tagged VLPs) and HT1080 (HA-tagged VLPs) cells cultured independently and in co-culture.
(B) GRN scores for FLAG-purified supernatants from HEK293T (FLAG-tagged VLPs) and HT1080 (HA-tagged VLPs) cells cultured independently and in co-culture.
(C) GRN scores for HA-purified supernatants from HEK293T (FLAG-tagged VLPs) and HT1080 (HA-tagged VLPs) cells cultured independently and in co-culture.



Supplementary Figure 4.13 | HT1080 TNF-α stimulation time course control.

(A) PCA on wild-type HT1080 supernatant lysates using differentially expressed genes found in cell lysate RNA-seq (no clear ± TNF-α separation observed as expected in the absence of VLP self-reporting).
(B) PCA on wild-type HT1080 supernatant lysates using differentially expressed genes found in self-reporting supernatant lysate RNA-seq (no clear ± TNF-α separation observed as expected in the absence of VLP self-reporting).
(C) PCA on wild-type HT1080 supernatant lysates using differentially expressed genes found in wild-type HT1080 supernatant lysate RNA-seq (no clear ± TNF-α separation observed as expected in the absence of VLP self-reporting).
(D) Fold-change (TNF-α vs. unstimulated) plot for supernatant lysate vs. cell lysate in self-reporting cells showing all differential expressed genes that are significant in either supernatant or lysate or both.

Chapter 5. Conclusions and Outlook

5.1 Opportunities for image-based screening of neuronal phenotypes

While the work described in this thesis demonstrates the utility of optical pooled screening to identify regulators of complex cellular behaviors, including cell non-autonomous phenotypes, further technical advances will improve the throughput, robustness, accessibility, and versatility of the platform for applications in neurobiology and beyond.

5.1.1 Limitations of optical pooled screening

Protocols for optical pooled screens, including wet-lab procedures, recommendations for imaging hardware, and custom analysis code for detecting *in situ* reads, have been published to facilitate widespread adoption by the scientific community[26]. However, several considerations still limit the accessibility of this screening approach.

First, sample processing, particularly sequencing by synthesis (SBS), poses a major throughput bottleneck. In Chapter 3, we described a custom automated fluidics platform that substantially reduces the hands-on labor required. To shorten imaging time, hardware optimizations can also be implemented, including imaging at lower magnification, using a camera with a larger field-of-view or a dual camera setup, or synchronizing microscope components with an external hardware trigger. Using 2-color instead of 4-color SBS chemistry can further decrease imaging time but would require refining the computational pipeline to accurately call *in situ* sequencing reads.

Moreover, although the cost of optical pooled screens is lower than that of single-cell sequencing-based screens, reagent costs for *in situ* barcode amplification and sequencing alone amount to upward of $1,000 per 6-well plate. Computing resources for analyses of large-scale image-based screens constitute another significant expense, and capital costs may be additionally incurred if an appropriate imaging setup must be procured or if limited microscope availability prevents screening on shared equipment. Furthermore, while published protocols serve as a basis for future optical pooled screens, different biological models, phenotyping assays, screen designs, and hardware configurations will need specific optimizations to both the wet-lab steps as well as the analysis pipeline. Since optical pooled screens require a substantial investment of resources, outsourcing certain components of the workflow to outside groups, academic or commercial, will thus enable the broader research community to leverage this screening platform for other biological applications.

5.1.2 Genetic perturbation modalities

Although this thesis focused on loss-of-function screens, other approaches for inducing genetic perturbations broaden the scope of questions answerable by functional genomics. While both Cas9-mediated knockouts and CRISPRi-based knockdowns are effective in loss-of-function screens, CRISPRi enables screening of essential genes and non-coding elements via transcriptional repression[5,6]. Conversely, CRISPR activation (CRISPRa) can be used in gain-of-function screening by using a dCas9 fused to transcriptional activator domains to increase gene expression. Gain-of-function screening may also be achieved by overexpression of open reading frames (ORFs), often associated with a unique barcode sequence. By expressing specific protein isoforms and variants, this approach allows for studies of phenotypic effects resulting from mutations of interest, including disease-associated variants, and is especially relevant for neuronal functions that are proteoform-specific. Furthermore, using gain-of-function methods, genes that are only lowly or not expressed in the screening model may still be interrogated. For example, in mixed culture assays described in Chapter 2, neuronal genes that are not endogenously expressed in non-neuronal cells can be systematically probed using such tools.

Genetic screens probing the effects of more complex genetic manipulations, beyond changes in expression of individual genes, are also now possible. Combinatorial perturbation screening to identify genetic redundancies or interactions relies on the introduction of multiple genetic perturbations, e.g. multiplexed sgRNAs, into each cell. Moreover, advances in gene editing tools that allow installation of precise mutations at endogenous loci, such as base editing and prime editing, have enabled large-scale variant screens[215–217]. In principle, these different perturbation modalities may be compatible with optical pooled screening because it is possible to *in* situ sequence sgRNA spacers, other perturbation barcodes, or actual editing outcomes.

5.1.3 Optical pooled screens with complex biological models

The work presented in this thesis and other published optical pooled screens were performed in cancer cell lines due to their tractability as cellular models for high-throughput genetic screening. Other *in vitro* models, including primary cells (isolated from animal models or human donors), cell types derived from induced pluripotent stem cells (iPSCs), or mixed cultures, may be more physiologically relevant but pose technical challenges. Maintaining coverage for sufficient screening power requires a large enough cell population that robustly exhibits the phenotypes of interest. The suitability of a cell model for screening thus depends on multiple factors, such as the penetrance of the phenotype, the efficiency of the differentiation protocol, the transduction or editing efficacy of the perturbation library, the ability to select and expand the transduced cells, and, for optical pooled screening in particular, the efficiency of *in situ* detection of perturbation barcodes. This *in situ* sequencing efficiency is partially determined by mRNA expression levels. Modifications to expression vectors, such as incorporating cell type-specific promoters or ubiquitous chromatin opening elements to reduce gene silencing, and optimizations to the *in situ* amplification protocol, including changing fixation conditions, wash protocols, or blocking steps, may improve the *in situ* detection of sequencing reads in these models. With new commercial sequencing technologies, other sequencing reagents may also be considered for *in situ* sequencing, especially in samples with higher background autofluorescence[218].

Despite their challenges, these cellular models often more accurately recapitulate certain biological processes or disease states and thus may serve as more appropriate screening systems. For example, while non-neuronal cells may be used in certain assays, as described in Chapter 2, most studies in neurobiology rely on model organisms, primary cells, differentiated cells, or co-cultures of various cell types. Organoids are another model system that are amenable for genetic screening, but image-based pooled screens in these would require additional protocol optimizations, such as sample clearing or volumetric imaging for three-dimensional structures, to maintain high *in situ* sequencing efficiency and imaging throughput[219–222]. In screens using such complex biological models, all the above considerations must therefore be carefully reviewed and incorporated into the screen design.

5.1.4 In vivo *screens with spatial readouts*

Many phenotypes depend on the physiological and spatial contexts of cells and thus cannot be recapitulated in *in vitro* systems. *In vivo* pooled CRISPR screens have been successfully performed using enrichment- or sequencing-based readouts, including in the mouse brain, to probe cellular behaviors in native settings, but these methods do not capture the spatial architecture and relationships between cells[223–225]. *In vivo* screens with image-based readouts, on the other hand, could retain this spatial information to reveal genetic effects on cellular organization. Recently, a spatial *in vivo* screen using protein barcodes integrated spatial transcriptomics with histopathology to uncover effects of 35 gene knockouts; while the detection of protein barcodes in tissues is cheaper and simpler than that of nucleic acid barcodes, it does not scale as easily to larger perturbation libraries[13,16]. Existing *ex situ* spatial transcriptomics approaches may be adapted to detect mRNA perturbation barcodes in tissues[226–228]. Other *in situ* sequencing protocols may also be applied to pooled screens, but would require further technical optimizations, e.g. to prevent sequencing background accumulation[229,230]. Tissue clearing, in addition to other methods already discussed, including new sequencing chemistries, may enable more efficient and robust capture of perturbation barcodes in tissues and, as a result, more scalable spatial *in vivo* screens.

5.1.5 Analysis of single-cell image-based screening datasets

A dataset from an image-based pooled screen contains measurements for hundreds to thousands of features and the spatial information for each single cell. Each step within the analysis pipeline, from feature extraction to perturbation-level aggregation, can be performed using many different computational approaches. Although single-cell datasets with genetic perturbations exist, analytical approaches differ between studies, and how best to handle such large high-content, single-cell datasets remains an open question[231].

In Chapter 2, we described a relatively straightforward approach to extract pre-defined features that best described a phenotype of interest. This analytical workflow, however, does not extract all information captured in images of complex cellular phenotypes. While intensity- and geometry-based metrics can conveniently quantify certain phenotypes, advances in machine learning allow for feature learning directly from acquired images. Autoencoders, for example, derive features by encoding images into latent representations and decoding them into new output images, and have been successfully applied to describe subcellular localization patterns[232,233]. These and other deep learning approaches are particularly useful for screens with few phenotypes defined *a priori* or generally complex phenotypes, obtained from multiple imaging channels.

An additional technical barrier for image-based pooled screens in neuronal systems is cell segmentation and projection mapping. Because cell body segmentation is relatively simple, phenotypes limited to the soma are amenable for pooled image-based screens. However, phenotypes that require precise assignment of neurite to cell are more challenging. Some automated segmentation algorithms have been implemented for select regions of interests in arrayed screens, but have not been applied to pooled assays[55,234]. As such, experimental approaches, including sparse plating and labeling with fluorescent markers or barcodes, may aid neurite segmentation and assignment[230,235,236]. Nevertheless, improved algorithms for high-throughput segmentation of neurons will be necessary to enable routine pooled screening in these models.

5.2 Future applications for live-cell transcriptomics

5.2.1 Single-cell transcriptional recording

Cellular responses to stimuli are often heterogenous within a population[237–241]. Understanding how the molecular state of a cell impacts its phenotypic behavior requires approaches that can measure or infer changes in these states within the same sample. Although existing cell profiling approaches can robustly capture cell states at a static time point, most are destructive and therefore do not allow repeated or orthogonal readouts from the same cell at later time points. A recently developed technology allows non-destructive, longitudinal measurements of transcriptomes from living cells[183]. Albeit an important advance, this platform relies on specialized hardware and protocols to extract picoliter-scale aliquots of cytoplasmic mRNAs and thus suffers from low throughput, further limiting its practicality and utility for capturing rarer transcriptional states.

On the other hand, since self-reporting is enabled by genetically encoded elements, the technology is more easily scalable to larger populations of cells (Chapter 4). In this thesis, we described the engineering of envelope proteins as barcodes for multiplexed readouts of different cell types or conditions from a bulk population. However, additional optimizations are necessary to provide single-cell resolution to self-reporting readouts. Increasing VLP export from individual cells, e.g. through Gag protein or cell engineering, would improve the recovery rate and robustness of transcriptional signals. Furthermore, an approach to demultiplex the source cell from which a VLP or RNA was exported is required. Advances in droplet-based sequencing technologies may enable VLP-level isolation and RNA profiling. Alternatively, an intracellular barcoding approach, e.g. through expression of a nucleic acid barcode that is packaged into VLPs with cellular mRNAs, would allow temporal single-cell transcriptional measurements from bulk populations.

5.2.2 Live-cell transcriptomics in in vitro *and* in vivo *models*

Self-reporting from bulk populations is still a powerful approach to track dynamic processes via transcriptional changes, particularly in complex biological models that are not amenable to subsampling. The VLP-based live-cell profiling approach is therefore suitable for models that cannot be easily expanded, exhibit high sample-to-sample variability, or are functionally dependent on their spatial organization. Such systems include post-mitotic cells (e.g. neurons) and 3D *in vitro* models (e.g. organoids).

Epitope tagging of VLPs, moreover, enables demultiplexing of reads from different cell types within mixed cultures, providing an additional dimension to the readout (Chapter 4). One application of this approach is to drive the expression of orthogonal epitopes with cell type- or lineage-specific promoters to elucidate differential transcriptional responses to stimuli, such as the emergence of different lineages upon differentiation.

Future applications of the self-reporting technology may also include *in vivo* models, allowing transcriptional monitoring of tissues in their physiological context. However, further optimizations in VLP export efficiency as well as extraction and purification of VLPs from bodily fluids (e.g. blood or spinal cord fluid) are necessary for longitudinal transcriptional monitoring in living animals to be feasible.

References

1. Brummelkamp, T. R., Bernards, R. & Agami, R. A System for Stable Expression of Short Interfering RNAs in Mammalian Cells. *Science* **296**, 550–553 (2002).
2. Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J. & Conklin, D. S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16**, 948–958 (2002).
3. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
4. Przybyla, L. & Gilbert, L. A. A new era in functional genomics screens. *Nat. Rev. Genet.* **23**, 89–103 (2022).
5. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
6. Gilbert, L. A. *et al.* Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
7. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
8. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
9. Walton, R. T., Singh, A. & Blainey, P. C. Pooled genetic screens with image-based profiling. *Mol. Syst. Biol.* **18**, e10768 (2022).
10. Dixit, A. *et al.* Perturb-seq: Dissecting molecular circuits with scalable single cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853-1866.e17 (2016).
11. Adamson, B. *et al.* A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* **167**, 1867-1882.e21 (2016).
12. Jaitin, D. A. *et al.* Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**, 1883-1896.e15 (2016).

13. Wroblewska, A. *et al.* Protein Barcodes Enable High-Dimensional Single-Cell CRISPR Screens. *Cell* **175**, 1141-1155.e16 (2018).
14. Rubin, A. J. *et al.* Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks. *Cell* **176**, 361-376.e17 (2019).
15. Mimitou, E. P. *et al.* Multiplexed detection of proteins, transcriptomes, clonotypes and CRISPR perturbations in single cells. *Nat. Methods* **16**, 409–412 (2019).
16. Dhainaut, M. *et al.* Spatial CRISPR genomics identifies regulators of the tumor microenvironment. *Cell* **185**, 1223-1239.e20 (2022).
17. Saka, S. K. *et al.* Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues. *Nat. Biotechnol.* **37**, 1080–1090 (2019).
18. Lin, J.-R., Fallahi-Sichani, M. & Sorger, P. K. Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nat. Commun.* **6**, 8390 (2015).
19. Gut, G., Herrmann, M. D. & Pelkmans, L. Multiplexed protein maps link subcellular organization to cellular states. *Science* **361**, eaar7042 (2018).
20. Codeluppi, S. *et al.* Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* **15**, 932–935 (2018).
21. Emanuel, G., Moffitt, J. R. & Zhuang, X. High-throughput, image-based screening of pooled genetic variant libraries. *Nat. Methods* **14**, 1159–1162 (2017).
22. Lawson, M. J. *et al.* In situ genotyping of a pooled strain library after characterizing complex phenotypes. *Mol. Syst. Biol.* **13**, 947 (2017).
23. Feldman, D. *et al.* Optical Pooled Screens in Human Cells. *Cell* **179**, 787-799.e17 (2019).
24. Wang, C., Lu, T., Emanuel, G., Babcock, H. P. & Zhuang, X. Imaging-based pooled CRISPR screening reveals regulators of lncRNA localization. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 10842–10851 (2019).
25. Shi, H. *et al.* Highly Multiplexed Spatial Mapping of Microbial Communities. *Nature* **588**, 676–681 (2020).

26. Feldman, D. *et al.* Pooled genetic perturbation screens with image-based phenotypes. *Nat. Protoc.* **17**, 476–512 (2022).
27. Feldman, D. *et al.* Pooled optical screens in human cells. *Cell* in press (2019).
28. Funk, L. *et al.* The phenotypic landscape of essential human genes. *Cell* **185**, 4634-4653.e22 (2022).
29. Carlson, R. J., Leiken, M. D., Guna, A., Hacohen, N. & Blainey, P. C. A genome-wide optical pooled screen reveals regulators of cellular antiviral responses. *Proc. Natl. Acad. Sci.* **120**, e2210623120 (2023).
30. Connors, B. W. & Long, M. A. Electrical synapses in the mammalian brain. *Annu. Rev. Neurosci.* **27**, 393–418 (2004).
31. Pereda, A. E. Electrical synapses and their functional interactions with chemical synapses. *Nat. Rev. Neurosci.* **15**, 250–263 (2014).
32. Südhof, T. C. Neurotransmitter release: the last millisecond in the life of a synaptic vesicle. *Neuron* **80**, 675–690 (2013).
33. Südhof, T. C. The presynaptic active zone. *Neuron* **75**, 11–25 (2012).
34. Phillips, G. R. *et al.* The presynaptic particle web: ultrastructure, composition, dissolution, and reconstitution. *Neuron* **32**, 63–77 (2001).
35. Collins, M. O. *et al.* Molecular characterization and comparison of the components and multiprotein complexes in the postsynaptic proteome. *J. Neurochem.* **97**, 16–23 (2006).
36. Sheng, M. & Kim, E. The postsynaptic organization of synapses. *Cold Spring Harb. Perspect. Biol.* **3**, a005678 (2011).
37. Südhof, T. C. Towards an Understanding of Synapse Formation. *Neuron* **100**, 276–293 (2018).
38. McAllister, A. K. Dynamic Aspects of Synapse Formation. *Annu. Rev. Neurosci.* **30**, 425–450 (2007).

39. Waites, C. L., Craig, A. M. & Garner, C. C. Mechanisms of vertebrate synaptogenesis. *Annu. Rev. Neurosci.* **28**, 251–274 (2005).

40. Petzoldt, A. G. & Sigrist, S. J. Synaptogenesis. *Curr. Biol.* **24**, R1076-80 (2014).

41. Yamagata, M., Duan, X. & Sanes, J. R. Cadherins Interact With Synaptic Organizers to Promote Synaptic Differentiation. *Front. Mol. Neurosci.* **11**, 142 (2018).

42. Pettem, K. L., Yokomaku, D., Takahashi, H., Ge, Y. & Craig, A. M. Interaction between autism-linked MDGAs and neuroligins suppresses inhibitory synapse development. *J. Cell Biol.* **200**, 321–336 (2013).

43. Kopczynski, C. C., Davis, G. W. & Goodman, C. S. A Neural Tetraspanin, Encoded by late bloomer, That Facilitates Synapse Formation. *Science* **271**, 1867–1870 (1996).

44. Kraut, R., Menon, K. & Zinn, K. A gain-of-function screen for genes controlling motor axon guidance and synaptogenesis in Drosophila. *Curr. Biol.* **11**, 417–430 (2001).

45. Aberle, H. *et al.* wishful thinking Encodes a BMP Type II Receptor that Regulates Synaptic Growth in Drosophila. *Neuron* **33**, 545–558 (2002).

46. Wan, H. I. *et al.* Highwire regulates synaptic growth in Drosophila. *Neuron* **26**, 313–29 (2000).

47. Yeh, E., Kawano, T., Weimer, R. M., Bessereau, J.-L. & Zhen, M. Identification of genes involved in synaptogenesis using a fluorescent active zone marker in Caenorhabditis elegans. *J. Neurosci.* **25**, 3833–41 (2005).

48. Crane, M. M. *et al.* Autonomous screening of C. elegans identifies genes implicated in synaptogenesis. *Nat. Methods* **9**, 977–80 (2012).

49. Hong, J.-H. & Park, M. Understanding synaptogenesis and functional connectome in C. elegans by imaging technology. *Front. Synaptic Neurosci.* **8**, 18 (2016).

50. Panzer, J. A. *et al.* Neuromuscular synaptogenesis in wild-type and mutant zebrafish. *Dev. Biol.* **285**, 340–357 (2005).

51. Jin, Y. Synaptogenesis: insights from worm and fly. *Curr. Opin. Neurobiol.* **12**, 71–79 (2002).

52. Linhoff, M. W. *et al.* An unbiased expression screen for synaptogenic proteins identifies the LRRTM protein family as synaptic organizers. *Neuron* **61**, 734–749 (2009).

53. Takahashi, H. *et al.* Postsynaptic TrkC and presynaptic PTPσ function as a bidirectional excitatory synaptic organizing complex. *Neuron* **69**, 287–303 (2011).

54. Sharma, K. *et al.* High-throughput genetic screen for synaptogenic factors: identification of LRP6 as critical for excitatory synapse development. *Cell Rep.* **5**, 1330–1341 (2013).

55. Nieland, T. J. F. *et al.* High Content Image Analysis Identifies Novel Regulators of Synaptogenesis in a High-Throughput RNAi Screen of Primary Neurons. *PLoS ONE* **9**, e91744 (2014).

56. Verstraelen, P. *et al.* Image-based profiling of synaptic connectivity in primary neuronal cell culture. *Front. Neurosci.* **12**, 389 (2018).

57. Biederer, T. & Scheiffele, P. Mixed-culture assays for analyzing neuronal synapse formation. *Nat. Protoc.* **2**, 670–676 (2007).

58. Scheiffele, P., Fan, J., Choih, J., Fetter, R. & Serafini, T. Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons. *Cell* **101**, 657–669 (2000).

59. Biederer, T. *et al.* SynCAM, a synaptic adhesion molecule that drives synapse assembly. *Science* **297**, 1525–1531 (2002).

60. Graf, E. R., Zhang, X., Jin, S.-X., Linhoff, M. W. & Craig, A. M. Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins. *Cell* **119**, 1013–1026 (2004).

61. Nam, C. I. & Chen, L. Postsynaptic assembly induced by neurexin-neuroligin interaction and neurotransmitter. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6137–6142 (2005).

62. Chubykin, A. A. *et al.* Dissection of synapse induction by neuroligins: effect of a neuroligin mutation associated with autism. *J. Biol. Chem.* **280**, 22365–22374 (2005).

63. Sando, R., Jiang, X. & Südhof, T. C. Latrophilin GPCRs direct synapse specificity by coincident binding of FLRTs and teneurins. *Science* **363**, eaav7969 (2019).

64. Bush, D. L. & Vogt, V. M. In Vitro Assembly of Retroviruses. *Annu. Rev. Virol.* **1**, 561–580 (2014).

65. Muriaux, D., Mirro, J., Harvin, D. & Rein, A. RNA is a structural element in retrovirus particles. *Proc. Natl. Acad. Sci.* **98**, 5246–5251 (2001).

66. Rein, A., McClure, M. R., Rice, N. R., Luftig, R. B. & Schultz, A. M. Myristylation site in Pr65gag is essential for virus particle formation by Moloney murine leukemia virus. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 7246–7250 (1986).

67. Soneoka, Y., Kingsman, S. M. & Kingsman, A. J. Mutagenesis analysis of the murine leukemia virus matrix protein: identification of regions important for membrane localization and intracellular transport. *J. Virol.* **71**, 5549–5559 (1997).

68. Hansen, M., Jelinek, L., Whiting, S. & Barklis, E. Transport and assembly of gag proteins into Moloney murine leukemia virus. *J. Virol.* **64**, 5306–5316 (1990).

69. Hansen, M. S. & Barklis, E. Structural interactions between retroviral Gag proteins examined by cysteine cross-linking. *J. Virol.* **69**, 1150–1159 (1995).

70. Andrawiss, M., Takeuchi, Y., Hewlett, L. & Collins, M. Murine Leukemia Virus Particle Assembly Quantitated by Fluorescence Microscopy: Role of Gag-Gag Interactions and Membrane Association. *J. Virol.* **77**, 11651–11660 (2003).

71. Nooraei, S. *et al.* Virus-like particles: preparation, immunogenicity and their roles as nanovaccines and drug nanocarriers. *J. Nanobiotechnology* **19**, 59 (2021).

72. Voelkel, C. *et al.* Protein transduction from retroviral Gag precursors. *Proc. Natl. Acad. Sci.* **107**, 7805–7810 (2010).

73. Kaczmarczyk, S. J., Sitaraman, K., Young, H. A., Hughes, S. H. & Chatterjee, D. K. Protein delivery using engineered virus-like particles. *Proc. Natl. Acad. Sci.* **108**, 16998–17003 (2011).

74. Choi, J. G. *et al.* Lentivirus pre-packed with Cas9 protein for safer gene editing. *Gene Ther.* **23**, 627–633 (2016).

75. Mangeot, P. E. *et al.* Genome editing in primary cells and in vivo using viral-derived Nanoblades loaded with Cas9-sgRNA ribonucleoproteins. *Nat. Commun.* **10**, 45 (2019).
76. Banskota, S. *et al.* Engineered virus-like particles for efficient in vivo delivery of therapeutic proteins. *Cell* **185**, 250-265.e16 (2022).
77. Eyckerman, S. *et al.* Trapping mammalian protein complexes in viral particles. *Nat. Commun.* **7**, 11416 (2016).
78. Segel, M. *et al.* Mammalian retrovirus-like protein PEG10 packages its own mRNA and can be pseudotyped for mRNA delivery. *Science* **373**, 882–889 (2021).
79. Pastuzyn, E. D. *et al.* The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. *Cell* **172**, 275-288.e18 (2018).
80. Ashley, J. *et al.* Retrovirus-like Gag protein Arc1 binds RNA and traffics across synaptic boutons. *Cell* **172**, 262-274.e11 (2018).
81. Chowdhury, D., Watters, K. & Biederer, T. Synaptic recognition molecules in development and disease. *Curr. Top. Dev. Biol.* **142**, 319–370 (2021).
82. Sanes, J. R. & Yamagata, M. Many paths to synaptic specificity. *Annu. Rev. Cell Dev. Biol.* **25**, 161–195 (2009).
83. Nusser, Z. Creating diverse synapses from the same molecules. *Curr. Opin. Neurobiol.* **51**, 8–15 (2018).
84. Loh, K. H. *et al.* Proteomic Analysis of Unbounded Cellular Compartments: Synaptic Clefts. *Cell* **166**, 1295-1307.e21 (2016).
85. Cijsouw, T. *et al.* Mapping the Proteome of the Synaptic Cleft through Proximity Labeling Reveals New Cleft Proteins. *Proteomes* **6**, 48 (2018).
86. Schreiner, D., Simicevic, J., Ahrné, E., Schmidt, A. & Scheiffele, P. Quantitative isoform-profiling of highly diversified recognition molecules. *eLife* **4**, e07794 (2015).
87. Gangwar, S. P. *et al.* Molecular mechanism of MDGA1: regulation of neuroligin 2:neurexin trans-synaptic bridges. *Neuron* **94**, 1132-1141.e4 (2017).

88. Jiang, X., Sando, R. & Südhof, T. C. Multiple signaling pathways are essential for synapse formation induced by synaptic adhesion molecules. *Proc. Natl. Acad. Sci.* **118**, e2000173118 (2021).
89. Song, J. Y., Ichtchenko, K., Südhof, T. C., Brose, N. & Malenka, R. C. Neuroligin 1 is a postsynaptic cell-adhesion molecule of excitatory synapses. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 1100–5 (1999).
90. Chih, B., Engelman, H. & Scheiffele, P. Control of excitatory and inhibitory synapse formation by neuroligins. *Science* **307**, 1324–1328 (2005).
91. Varoqueaux, F. *et al.* Neuroligins determine synapse maturation and function. *Neuron* **51**, 741–754 (2006).
92. McKinley, K. L. & Cheeseman, I. M. Large-Scale Analysis of CRISPR/Cas9 Cell-Cycle Knockouts Reveals the Diversity of p53-Dependent Responses to Cell-Cycle Defects. *Dev. Cell* **40**, 405-420.e2 (2017).
93. Shaw, G., Morse, S., Ararat, M. & Graham, F. L. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **16**, 869–871 (2002).
94. Lin, Y.-C. *et al.* Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nat. Commun.* **5**, 4767 (2014).
95. Koopmans, F. *et al.* SynGO: an evidence-based, expert-curated knowledgebase for the synapse. *Neuron* **103**, 217-234.e4 (2019).
96. Thul, P. J. *et al.* A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).
97. Moon, K. R. *et al.* Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019).
98. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

99. Dubiel, W., Dubiel, D., Wolf, D. A. & Naumann, M. Cullin 3-based ubiquitin ligases as master regulators of mammalian cell differentiation. *Trends Biochem. Sci.* **43**, 95–107 (2018).

100. Forgac, M. Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology. *Nat. Rev. Mol. Cell Biol.* **8**, 917–929 (2007).

101. Hallock, P. T. *et al.* Dok-7 regulates neuromuscular synapse formation by recruiting Crk and Crk-L. *Genes Dev.* **24**, 2451–2461 (2010).

102. Hallock, P. T., Chin, S., Blais, S., Neubert, T. A. & Glass, D. J. Sorbs1 and -2 Interact with CrkL and Are Required for Acetylcholine Receptor Cluster Formation. *Mol. Cell. Biol.* **36**, 262–270 (2016).

103. Knafo, S. & Esteban, J. A. PTEN: Local and Global Modulation of Neuronal Function in Health and Disease. *Trends Neurosci.* **40**, 83–91 (2017).

104. Fraser, M. M., Bayazitov, I. T., Zakharenko, S. S. & Baker, S. J. Pten deficiency in brain causes defects in synaptic structure, transmission and plasticity, and myelination abnormalities. *Neuroscience* **151**, 476–488 (2008).

105. Wu, C.-H. *et al.* Mutations in the Profilin 1 Gene Cause Familial Amyotrophic Lateral Sclerosis. *Nature* **488**, 499–503 (2012).

106. Michaelsen-Preusse, K. *et al.* Neuronal profilins in health and disease: Relevance for spine plasticity and Fragile X syndrome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 3365–3370 (2016).

107. Fogel, A. I. *et al.* SynCAMs Organize Synapses through Heterophilic Adhesion. *J. Neurosci.* **27**, 12516–12530 (2007).

108. Robbins, E. M. *et al.* SynCAM 1 Adhesion Dynamically Regulates Synapse Number and Impacts Plasticity and Learning. *Neuron* **68**, 894–906 (2010).

109. Noel, J. *et al.* Surface expression of AMPA receptors in hippocampal neurons is regulated by an NSF-dependent mechanism. *Neuron* **23**, 365–376 (1999).

110. Zhao, C., Slevin, J. T. & Whiteheart, S. W. Cellular functions of NSF: not just SNAPs and SNAREs. *FEBS Lett.* **581**, 2140–2149 (2007).

111. Lim, Y. W., James, D., Huang, J. & Lee, M. The Emerging Role of the RNA-Binding Protein SFPQ in Neuronal Function and Neurodegeneration. *Int. J. Mol. Sci.* **21**, 7151 (2020).

112. Choy, R. W.-Y. *et al.* Retromer mediates a discrete route of local membrane delivery to dendrites. *Neuron* **82**, 55–62 (2014).

113. Cerpa, W., Latorre-Esteves, E. & Barria, A. RoR2 functions as a noncanonical Wnt receptor that regulates NMDAR-mediated synaptic transmission. *Proc. Natl. Acad. Sci.* **112**, 4797–4802 (2015).

114. Paganoni, S., Bernstein, J. & Ferreira, A. ROR1-ROR2 COMPLEXES MODULATE SYNAPSE FORMATION IN HIPPOCAMPAL NEURONS. *Neuroscience* **165**, 1261 (2010).

115. Scott, E. K., Reuter, J. E. & Luo, L. Small GTPase Cdc42 is required for multiple aspects of dendritic morphogenesis. *J. Neurosci. Off. J. Soc. Neurosci.* **23**, 3118–3123 (2003).

116. Kim, I. H., Wang, H., Soderling, S. H. & Yasuda, R. Loss of Cdc42 leads to defects in synaptic plasticity and remote memory recall. *eLife* **3**, e02839 (2014).

117. Moore, C. J. & Winder, S. J. Dystroglycan versatility in cell adhesion: a tale of multiple motifs. *Cell Commun. Signal.* **8**, 3 (2010).

118. Früh, S. *et al.* Neuronal Dystroglycan Is Necessary for Formation and Maintenance of Functional CCK-Positive Basket Cell Terminals on Pyramidal Cells. *J. Neurosci. Off. J. Soc. Neurosci.* **36**, 10296–10313 (2016).

119. Liu, A. *et al.* Neuroligin 1 regulates spines and synaptic plasticity via LIMK1/cofilin-mediated actin reorganization. *J. Cell Biol.* **212**, 449–463 (2016).

120. Meng, Y. *et al.* Abnormal spine morphology and enhanced LTP in LIMK-1 knockout mice. *Neuron* **35**, 121–133 (2002).

121. Yang, B. *et al.* SH3 domain-mediated interaction of dystroglycan and Grb2. *J. Biol. Chem.* **270**, 11711–11714 (1995).

122. Lévi, S. *et al.* Dystroglycan Is Selectively Associated with Inhibitory GABAergic Synapses But Is Dispensable for Their Differentiation. *J. Neurosci.* **22**, 4274–4285 (2002).

123. Yoshida, T. *et al.* Interleukin-1 Receptor Accessory Protein Organizes Neuronal Synaptogenesis as a Cell Adhesion Molecule. *J. Neurosci.* **32**, 2588–2600 (2012).

124. de Wit, J. *et al.* Unbiased discovery of glypican as a receptor for LRRTM4 in regulating excitatory synapse development. *Neuron* **79**, 696–711 (2013).

125. Farhy-Tselnicker, I. *et al.* Astrocyte-Secreted Glypican 4 Regulates Release of Neuronal Pentraxin 1 from Axons to Induce Functional Synapse Formation. *Neuron* **96**, 428-445.e13 (2017).

126. Kamimura, K. & Maeda, N. Glypicans and Heparan Sulfate in Synaptic Development, Neural Plasticity, and Neurological Disorders. *Front. Neural Circuits* **15**, (2021).

127. Adorno-Cruz, V. & Liu, H. Regulation and functions of integrin α2 in cell adhesion and disease. *Genes Dis.* **6**, 16–24 (2018).

128. Shirao, T. & González-Billault, C. Actin filaments and microtubules in dendritic spines. *J. Neurochem.* **126**, 155–164 (2013).

129. Dillon, C. & Goda, Y. The actin cytoskeleton: integrating form and function at the synapse. *Annu. Rev. Neurosci.* **28**, 25–55 (2005).

130. Aiken, J. & Holzbaur, E. L. F. Cytoskeletal regulation guides neuronal trafficking to effectively supply the synapse. *Curr. Biol.* **31**, R633–R650 (2021).

131. Ackermann, M. & Matus, A. Activity-induced targeting of profilin and stabilization of dendritic spine morphology. *Nat. Neurosci.* **6**, 1194–1200 (2003).

132. Wolf, M. *et al.* ADF/Cofilin Controls Synaptic Actin Dynamics and Regulates Synaptic Vesicle Mobilization and Exocytosis. *Cereb. Cortex* **25**, 2863–2875 (2015).

133. Pilo Boyl, P. *et al.* Profilin2 contributes to synaptic vesicle exocytosis, neuronal excitability, and novelty-seeking behavior. *EMBO J.* **26**, 2991–3002 (2007).

134. Ivanov, A., Esclapez, M. & Ferhat, L. Role of drebrin A in dendritic spine plasticity and synaptic function. *Commun. Integr. Biol.* **2**, 268–270 (2009).

135. Takahashi, H. *et al.* Drebrin-dependent actin clustering in dendritic filopodia governs synaptic targeting of postsynaptic density-95 and dendritic spine morphogenesis. *J. Neurosci. Off. J. Soc. Neurosci.* **23**, 6586–6595 (2003).

136. Daly, R. J. Cortactin signalling and dynamic actin networks. *Biochem. J.* **382**, 13–25 (2004).

137. Hering, H. & Sheng, M. Activity-Dependent Redistribution and Essential Role of Cortactin in Dendritic Spine Morphogenesis. *J. Neurosci.* **23**, 11759–11769 (2003).

138. Mukherjee, K. *et al.* Actin capping protein CAPZB regulates cell morphology, differentiation, and neural crest migration in craniofacial morphogenesis. *Hum. Mol. Genet.* **25**, 1255–1270 (2016).

139. Fukazawa, Y. *et al.* Hippocampal LTP is accompanied by enhanced F-actin content within the dendritic spine that is essential for late LTP maintenance in vivo. *Neuron* **38**, 447–460 (2003).

140. Nakayama, A. Y., Harms, M. B. & Luo, L. Small GTPases Rac and Rho in the maintenance of dendritic spines and branches in hippocampal pyramidal neurons. *J. Neurosci. Off. J. Soc. Neurosci.* **20**, 5329–5338 (2000).

141. Govek, E.-E. *et al.* The X-linked mental retardation protein oligophrenin-1 is required for dendritic spine morphogenesis. *Nat. Neurosci.* **7**, 364–372 (2004).

142. Tolias, K. F., Duman, J. G. & Um, K. Control of synapse development and plasticity by Rho GTPase regulatory proteins. *Prog. Neurobiol.* **94**, 133–148 (2011).

143. Briz, V. *et al.* Activity-Dependent Rapid Local RhoA Synthesis Is Required for Hippocampal Synaptic Plasticity. *J. Neurosci.* **35**, 2269–2282 (2015).

144. Zhang, H., Ben Zablah, Y., Zhang, H. & Jia, Z. Rho Signaling in Synaptic Plasticity, Memory, and Brain Disorders. *Front. Cell Dev. Biol.* **9**, (2021).

145. Park, M. & Shen, K. WNTs in synapse formation and neuronal circuitry. *EMBO J.* **31**, 2697–2704 (2012).

146. Ahmad-Annuar, A. *et al.* Signaling across the synapse: a role for Wnt and Dishevelled in presynaptic assembly and neurotransmitter release. *J. Cell Biol.* **174**, 127–139 (2006).

147. Hruska, M. & Dalva, M. B. Ephrin regulation of synapse formation, function and plasticity. *Mol. Cell. Neurosci.* **50**, 35–44 (2012).

148. Arvanitis, D. N. *et al.* Ephrin B1 maintains apical adhesion of neural progenitors. *Dev. Camb. Engl.* **140**, 2082–2092 (2013).

149. Henderson, N. & Dalva, M. B. EphBs and ephrin-Bs: Trans-synaptic organizers of synapse development and function. *Mol. Cell. Neurosci.* **91**, 108–121 (2018).

150. Irie, F. & Yamaguchi, Y. EphB receptors regulate dendritic spine development via intersectin, Cdc42 and N-WASP. *Nat. Neurosci.* **5**, 1117–1118 (2002).

151. Smith, C. L., Mittaud, P., Prescott, E. D., Fuhrer, C. & Burden, S. J. Src, Fyn, and Yes are not required for neuromuscular synapse formation but are necessary for stabilization of agrin-induced clusters of acetylcholine receptors. *J. Neurosci. Off. J. Soc. Neurosci.* **21**, 3151–3160 (2001).

152. Viquez, N. M., Li, C. R., Wairkar, Y. P. & DiAntonio, A. The B′ Protein Phosphatase 2A Regulatory Subunit well-rounded Regulates Synaptic Growth and Cytoskeletal Stability at the Drosophila Neuromuscular Junction. *J. Neurosci.* **26**, 9293–9303 (2006).

153. Hoffman, A., Taleski, G. & Sontag, E. The protein serine/threonine phosphatases PP2A, PP1 and calcineurin: A triple threat in the regulation of the neuronal cytoskeleton. *Mol. Cell. Neurosci.* **84**, 119–131 (2017).

154. Petrone, A. *et al.* Receptor protein tyrosine phosphatase α is essential for hippocampal neuronal migration and long-term potentiation. *EMBO J.* **22**, 4121–4131 (2003).

155. Wu, Y. J., Pierre, D. P. L., Wu, J., Yee, A. J. & Yang, B. B. The interaction of versican with its binding partners. *Cell Res.* **15**, 483–494 (2005).

156. Hasle, N. *et al.* High-throughput, microscope-based sorting to dissect cellular heterogeneity. *Mol. Syst. Biol.* **16**, e9442 (2020).

157. Wheeler, E. C. *et al.* Pooled CRISPR screens with imaging on microraft arrays reveals stress granule-regulatory factors. *Nat. Methods* **17**, 636–642 (2020).

158. Kanfer, G. *et al.* Image-based pooled whole-genome CRISPRi screening for subcellular phenotypes. *J. Cell Biol.* **220**, e202006180 (2021).

159. Schraivogel, D. *et al.* High-speed fluorescence image-enabled cell sorting. *Science* **375**, 315–320 (2022).

160. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

161. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PloS One* **3**, e3647 (2008).

162. Clement, K. *et al.* CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).

163. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184–191 (2016).

164. Wang, T. *et al.* Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell* **168**, 890-903.e15 (2017).

165. Hart, T. *et al.* Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3 Bethesda Md* **7**, 2719–2727 (2017).

166. McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* **16**, e2005970 (2018).

167. Walt, S. van der *et al.* scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).

168. Köster, J. & Rahmann, S. Snakemake--a scalable bioinformatics workflow engine. *Bioinforma. Oxf. Engl.* **28**, 2520–2522 (2012).

169. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102**, 15545 LP – 15550 (2005).

170. Kwon, H.-B. *et al.* Neuroligin-1-dependent competition regulates cortical synaptogenesis and synapse number. *Nat. Neurosci.* **15**, 1667–1674 (2012).

171. Schindelin, J. *et al.* Fiji - an Open Source platform for biological image analysis. *Nat. Methods* **9**, 10.1038/nmeth.2019 (2012).

172. Pandit, K. *et al.* An open source toolkit for repurposing Illumina sequencing systems as versatile fluidics and imaging platforms. *Sci. Rep.* **12**, 5081 (2022).

173. Black, S. *et al.* CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nat. Protoc.* **16**, 3802–3835 (2021).

174. Shimomura, O., Johnson, F. H. & Saiga, Y. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *J. Cell. Comp. Physiol.* **59**, 223–239 (1962).

175. Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).

176. Stark, R., Grzelak, M. & Hadfield, J. RNA sequencing: the teenage years. *Nat. Rev. Genet.* **20**, 631–656 (2019).

177. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628 (2008).

178. Nagalakshmi, U. *et al.* The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**, 1344–1349 (2008).

179. Eberwine, J. *et al.* Analysis of gene expression in single live neurons. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 3010–3014 (1992).

180. Sul, J.-Y. *et al.* Transcriptome transfer produces a predictable cellular phenotype. *Proc. Natl. Acad. Sci.* **106**, 7624–7629 (2009).

181. Guillaume-Gentil, O. *et al.* Tunable Single-Cell Extraction for Molecular Analyses. *Cell* **166**, 506–516 (2016).

182. Cao, Y. *et al.* Nondestructive nanostraw intracellular sampling for longitudinal cell monitoring. *Proc. Natl. Acad. Sci.* **114**, E1866–E1874 (2017).

183. Chen, W. *et al.* Live-seq enables temporal transcriptomic recording of single cells. *Nature* **608**, 733–740 (2022).

184. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).

185. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).

186. Liu, Z. *et al.* Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. *Nat. Commun.* **8**, 22 (2017).

187. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

188. Schiebinger, G. *et al.* Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* **176**, 928-943.e22 (2019).

189. Saelens, W., Cannoodt, R., Todorov, H. & Saeys, Y. A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547–554 (2019).

190. Weinreb, C., Wolock, S., Tusi, B. K., Socolovsky, M. & Klein, A. M. Fundamental limits on dynamic inference from single-cell snapshots. *Proc. Natl. Acad. Sci.* **115**, E2467–E2476 (2018).

191. Bertrand, E. *et al.* Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* **2**, 437–445 (1998).

192. Wang, Z., Zhang, K., Wooley, K. L. & Taylor, J.-S. Imaging mRNA Expression in Live Cells via PNA·DNA Strand Displacement-Activated Probes. *J. Nucleic Acids* **2012**, 962652 (2012).

193. Tutucci, E. *et al.* An improved MS2 system for accurate reporting of the mRNA life cycle. *Nat. Methods* **15**, 81–89 (2018).

194. Nelles, D. A. *et al.* Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell* **165**, 488–496 (2016).

195. Lovatt, D. *et al.* Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. *Nat. Methods* **11**, 190–196 (2014).

196. Atmanli, A. *et al.* Multiplex live single-cell transcriptional analysis demarcates cellular functional heterogeneity. *eLife* **8**, e49599 (2019).

197. Yang, L.-Z. *et al.* Dynamic Imaging of RNA in Living Cells by CRISPR-Cas13 Systems. *Mol. Cell* **76**, 981-997.e7 (2019).

198. Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175 (2016).

199. Tang, W. & Liu, D. R. Rewritable multi-event analog recording in bacterial and mammalian cells. *Science* **360**, eaap8992 (2018).

200. Schmidt, F., Cherepkova, M. Y. & Platt, R. J. Transcriptional recording by CRISPR spacer acquisition from RNA. *Nature* **562**, 380–385 (2018).

201. Rodriques, S. G. *et al.* RNA timestamps identify the age of single molecules in RNA sequencing. *Nat. Biotechnol.* **39**, 320–325 (2021).

202. Bang, C. & Thum, T. Exosomes: new players in cell-cell communication. *Int. J. Biochem. Cell Biol.* **44**, 2060–2064 (2012).

203. Rulli, S. J. *et al.* Selective and nonselective packaging of cellular RNAs in retrovirus particles. *J. Virol.* **81**, 6623–6631 (2007).

204. Donaldson, B., Lateef, Z., Walker, G. F., Young, S. L. & Ward, V. K. Virus-like particle vaccines: immunology and formulation for clinical translation. *Expert Rev. Vaccines* **17**, 833–849 (2018).

205. Horns, F. *et al.* Engineering RNA export for measurement and manipulation of living cells. *Cell* (2023) doi:10.1016/j.cell.2023.06.013.

206. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).

207. Cahan, P. *et al.* CellNet: network biology applied to stem cell engineering. *Cell* **158**, 903–915 (2014).

208. Rein, A., Datta, S. A. K., Jones, C. P. & Musier-Forsyth, K. Diverse interactions of retroviral Gag proteins with RNAs. *Trends Biochem. Sci.* **36**, 373–380 (2011).

209. Safaee, N. *et al.* Interdomain allostery promotes assembly of the poly(A) mRNA complex with PABP and eIF4G. *Mol. Cell* **48**, 375–386 (2012).

210. Martin, J. C. *et al.* Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy. *Cell* **178**, 1493-1508.e20 (2019).

211. Sharif, O., Bolshakov, V. N., Raines, S., Newham, P. & Perkins, N. D. Transcriptional profiling of the LPS induced NF-kappaB response in macrophages. *BMC Immunol.* **8**, 1 (2007).

212. Ping, X.-L. *et al.* Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. *Cell Res.* **24**, 177–189 (2014).

213. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).

214. Zhang, T. *et al.* RNALocate: a resource for RNA subcellular localizations. *Nucleic Acids Res.* **45**, D135–D138 (2017).

215. Hanna, R. E. *et al.* Massively parallel assessment of human variants with base editor screens. *Cell* **184**, 1064-1080.e20 (2021).

216. Cuella-Martin, R. *et al.* Functional interrogation of DNA damage response variants with base editing screens. *Cell* **184**, 1081-1097.e19 (2021).

217. Erwood, S. *et al.* Saturation variant interpretation using CRISPR prime editing. *Nat. Biotechnol.* **40**, 885–895 (2022).

218. Arslan, S. *et al.* Sequencing by avidity enables high accuracy with low reagent consumption. *Nat. Biotechnol.* 1–7 (2023) doi:10.1038/s41587-023-01750-7.

219. Ringel, T. *et al.* Genome-Scale CRISPR Screening in Human Intestinal Organoids Identifies Drivers of TGF-β Resistance. *Cell Stem Cell* **26**, 431-440.e8 (2020).

220. Ungricht, R. *et al.* Genome-wide screening in human kidney organoids identifies developmental and disease-related aspects of nephrogenesis. *Cell Stem Cell* **29**, 160-175.e7 (2022).

221. Li, C. *et al.* Single-cell brain organoid screening identifies developmental defects in autism. 2022.09.15.508118 Preprint at https://doi.org/10.1101/2022.09.15.508118 (2022).

222. Lukonin, I., Zinner, M. & Liberali, P. Organoids in image-based phenotypic chemical screens. *Exp. Mol. Med.* **53**, 1495–1502 (2021).

223. Kuhn, M., Santinha, A. J. & Platt, R. J. Moving from in vitro to in vivo CRISPR screens. *Gene Genome Ed.* **2**, 100008 (2021).

224. Chow, R. D. *et al.* AAV-mediated direct in vivo CRISPR screen identifies functional suppressors in glioblastoma. *Nat. Neurosci.* **20**, 1329–1341 (2017).

225. Jin, X. *et al.* In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science* **370**, eaaz6063 (2020).

226. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).

227. Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* **39**, 313–319 (2021).

228. Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).

229. Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).

230. Chen, X. *et al.* High-throughput mapping of long-range neuronal projection using in situ sequencing. *Cell* **179**, 772-786.e19 (2019).

231. Peidli, S. *et al.* scPerturb: Harmonized Single-Cell Perturbation Data. 2022.08.20.504663 Preprint at https://doi.org/10.1101/2022.08.20.504663 (2023).

232. Kobayashi, H., Cheveralls, K. C., Leonetti, M. D. & Royer, L. A. Self-supervised deep learning encodes high-resolution features of protein subcellular localization. *Nat. Methods* **19**, 995–1003 (2022).

233. Cho, N. H. *et al.* OpenCell: Endogenous tagging for the cartography of human cellular organization. *Science* **375**, eabi6983 (2022).

234. Tian, R. *et al.* CRISPR interference-based platform for multimodal genetic screens in human iPSC-derived neurons. *Neuron* **104**, 239-255.e12 (2019).

235. Livet, J. *et al.* Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450**, 56–62 (2007).

236. Sun, Y.-C. *et al.* Integrating barcoded neuroanatomy with spatial transcriptional profiling enables identification of gene correlates of projections. *Nat. Neurosci.* **24**, 873–885 (2021).

237. Gupta, P. B. *et al.* Stochastic State Transitions Give Rise to Phenotypic Equilibrium in Populations of Cancer Cells. *Cell* **146**, 633–644 (2011).

238. Hoppe, P. S. *et al.* Early myeloid lineage choice is not initiated by random PU.1 to GATA1 protein ratios. *Nature* **535**, 299–302 (2016).

239. Shaffer, S. M. *et al.* Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**, 431–435 (2017).

240. Eling, N., Morgan, M. D. & Marioni, J. C. Challenges in measuring and understanding biological noise. *Nat. Rev. Genet.* **20**, 536–548 (2019).

241. Emert, B. L. *et al.* Variability within rare cell states enables multiple paths towards drug resistance. *Nat. Biotechnol.* **39**, 865–876 (2021).