SCUOLA DI DOTTORATO

UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Dipartimento di / Department of

Medicina e Chirurgia

Dottorato di Ricerca in / PhD program DIMET Ciclo / Cycle XXXVII

Turning darkness into light: a CRISPR-LICHT high-throughput screen of functional roles of "T-Dark" target genes of the Sox-2 transcription factor in human cerebral organoids

Cognome / Surname Baldi Nome / Name Roberta

Matricola / Registration number 892731

Tutore / Tutor: Silvia Kirsten Nicolis

Coordinatore / Coordinator: Francesco Mantegazza

ANNO ACCADEMICO / ACADEMIC YEAR 2024/2025

SUMMARY

SUMMARY

The thesis focuses on the CRISPR-LICHT technique, combining CRISPR-Cas9 genome editing with lineage tracing in 3D cerebral organoids. This advanced method enables parallel loss-of-function studies in human brain tissue at cellular resolution, overcoming limitations of traditional 2D culture systems. The central objective was to investigate T-DARK genes, functionally uncharted targets regulated by SOX2, a transcription factor critical for neural stem cell maintenance and differentiation.

Key findings highlight SOX2's regulation of NSCs and its downstream targets, including T-DARK genes implicated in neurodevelopmental disorders. Using high-throughput RNA sequencing, the study identified 122 conserved T-DARK genes in mouse and human models. These genes were functionally assessed through CRISPR-LICHT, which incorporates dual barcoding for precise lineage tracking.

A detailed methodology describes the preparation of libraries targeting T-DARK genes, the generation of cerebral organoids, and the application of tamoxifen-inducible Cas9. The dual-barcode system linked genetic edits to specific cell lineages, while next-generation sequencing revealed gene knockout impacts on neurogenesis.

The work demonstrates CRISPR-LICHT's potential to illuminate poorly understood genomic regions and its applicability in modeling human-specific brain development and disease. Future applications may uncover therapeutic targets for SOX2-related disorders, bridging gaps in neurodevelopmental research.

INTRODUCTION

1. SOX2

1.1 SOX2-related Disease and Sox-2 Target Genes

The transcription factor SOX2 is essential for neural development, as it helps maintain neural stem cells (NSCs) and guides their differentiation into different neural lineages across the central nervous system (CNS). Mutations in SOX2 can lead to various neurodevelopmental disorders, such as microphthalmia and anophthalmia (eye defects), as well as structural issues in the hippocampus, cognitive impairments, seizures, and motor difficulties. In mouse models, deleting SOX2 in neural progenitor cells has been shown to replicate these abnormalities, highlighting its crucial role in maintaining NSCs and forming brain structures, particularly the hippocampus. The impact of SOX2 during development also depends on timing; early removal of SOX2 causes more severe defects compared to later stages, as seen in different conditional knockout experiments (1).

SOX2 also regulates important signaling pathways like Sonic Hedgehog (Shh) and Wnt, which are vital for brain development, especially in the hippocampus and forebrain. Its ability to mediate long-range chromatin interactions allows it to activate key developmental genes, such as Gli3 and Cxcr4, that are critical for maintaining NSC populations and ensuring their proper differentiation. These interactions highlight SOX2's role as an "architectural" transcription factor, shaping the 3D chromatin structure necessary for gene expression in neural tissues (1)(2).

The role of SOX2 extends beyond early brain development. In adult brains, it continues to regulate NSCs, particularly in areas like the hippocampus and subventricular zone, where it supports ongoing neurogenesis. Mutations in SOX2 or its downstream targets can disrupt NSC function and neural circuitry, leading to neurodevelopmental disorders and cognitive deficits (2).

SOX2 also regulates several target genes that are essential for NSC self-renewal and differentiation. Among these targets, genes such as Gli3 and Socs3 are crucial for neural development. For instance, SOX2 directly binds to enhancers within the Gli3 gene locus, activating this transcription factor necessary for Shh signaling and proper hippocampal formation. Another significant target, Socs3, participates in a regulatory loop that maintains NSC proliferation; its downregulation in SOX2 mutants correlates with a loss of NSC self-renewal ability. Understanding these downstream pathways and targets of SOX2 not

only sheds light on its mechanisms in brain development but also opens up possibilities for therapeutic interventions in SOX2-related neurodevelopmental disorders (1).

1.2 SOX2 and Neural Stem Cells (NSCs)

SOX2 plays a critical and multi-faceted role in NSCs, particularly in maintaining their self-renewal and guiding their differentiation into neurons and glial cells. It is highly expressed in both embryonic and adult NSCs, marking undifferentiated neural progenitors. One of its primary functions is to ensure the pluripotency of these cells by regulating genes essential for keeping them in an undifferentiated, proliferative state. SOX2-expressing NSCs are found in neurogenic regions like the subventricular zone (SVZ) and the dentate gyrus (DG) of the hippocampus, where they support ongoing neurogenesis even in adulthood(3).

The role of SOX2 extends beyond maintenance; it actively regulates the transition from NSCs to mature neurons. Studies have shown that it binds to promoters of genes involved in neurogenesis, such as Notch1, and controls pathways like Sonic Hedgehog (Shh). In hypomorphic SOX2 mutants, neurogenesis is severely impaired, leading to conditions like hippocampal hypoplasia and neuronal migration defects. Moreover, SOX2's function is dose-dependent, meaning that even a partial reduction in SOX2 levels can lead to significant neural defects, reflecting the gene’s sensitivity and critical threshold requirements(3).

Additionally, SOX2 plays a role in neuronal differentiation by suppressing alternative fates, such as glial differentiation, ensuring that neural progenitors mature into neurons. This regulation is particularly vital in regions like the retina and the brain's ganglionic eminence, where SOX2 deletion leads to defects in retinal ganglion cells and GABAergic interneurons, respectively. These findings highlight that SOX2 is indispensable not only for maintaining the pool of NSCs but also for guiding their differentiation into specific neuronal subtypes(3).

1.3 SOX2 Target Genes Identified with RNA-Seq

Previously, the lab had used NSCs obtained from neonatal mice, specifically by conducting a Sox2 conditional deletion in vivo through the activation of Nestin-Cre recombinase. This technique ensures that SOX2 is specifically deleted in neural progenitor cells. Once the Sox2 deletion was induced, NSCs were isolated from the forebrain and subsequently cultured in vitro. The NSCs from Sox2-deleted and wild-type mice were propagated in a growth medium containing basic fibroblast growth factor (bFGF), which maintains NSC

proliferation. Sox2-deleted NSCs showed a progressive inability to maintain self-renewal, typically losing this capacity after 5–9 passages. This inability was a direct result of the loss of Sox2, which normally regulates the self-renewal properties of NSCs.

The study also used lentiviral transduction to reintroduce Sox2, Fos, or Socs3 into the Sox2-deleted NSCs. This transduction allowed for the examination of the compensatory roles these factors could play in restoring the self-renewal and differentiation capacities of the Sox2-deleted NSCs. This method of obtaining and manipulating NSCs provided a powerful model to study the role of SOX2 in NSC biology and neuronal differentiation(4).

Another important experiment conducted as part of the same study was the RNA-seq analysis.The primary aim was to identify differentially expressed genes that could explain the impaired ability of SOX2-deficient NSCs to maintain their stem cell characteristics and to differentiate into mature neurons.

By comparing the transcriptomes of wild-type (WT) NSCs and SOX2-deleted NSCs, we want to uncover:

1. Genes directly regulated by SOX2, especially those involved in neurogenesis and neural differentiation.
2. Pathways and biological processes affected by the absence of SOX2, particularly those related to stem cell proliferation, neuronal lineage commitment, and cell fate decisions.
3. Whether SOX2 is involved in maintaining the priming of differentiation-related genes, ensuring that NSCs are ready to transition into neurons upon receiving the appropriate signals.

Thus, the RNA-seq analysis was designed to provide a comprehensive view of the transcriptional landscape altered by SOX2 deletion, highlighting both direct and indirect downstream effects. This would enable the identification of key molecular targets and pathways affected by SOX2, contributing to a better understanding of its role in neural development and the molecular mechanisms driving neurodevelopmental disorders associated with SOX2 mutations.

The RNA-seq analysis in this study revealed a profound impact of SOX2 deletion on the transcriptome of NSCs, with approximately 1,000 genes downregulated in comparison to wild-type cells. Many of these downregulated genes are directly involved in crucial neuronal differentiation processes, highlighting the central role that SOX2 plays in regulating the gene networks essential for neural development.

The RNA-seq data indicated that genes involved in cell proliferation, neurogenesis, and cell fate determination were among those most significantly downregulated. (4)

2. T-DARK Genes

2.1 T-DARK Genes: The functionally uncharted part of the Genome

T-dark genes, also known as "dark genes" or "dark genome," are genes that lack functional annotations in traditional genomic databases and remain largely uncharacterized. These genes have minimal known biological function, and many are predicted only through computational models or non-coding RNA studies. They are of interest because they may hold key roles in complex biological processes and diseases.

Among the approximately 1,000 genes that were found to be downregulated following the deletion of SOX2 in mouse NSCs, many have human homologs, making them relevant for studying human neurodevelopment. A subset of these downregulated genes falls into the category of T-dark genes, which are genes that remain poorly characterized or understudied despite their potential importance.(4)

The *Pharos* web portal plays a crucial role in identifying and cataloging these t-dark genes. Pharos is a comprehensive resource within the Illuminating the Druggable Genome (IDG) initiative, which aims to expand knowledge about less-studied genes. Pharos provides researchers with tools to explore and prioritize genes based on their potential druggability, protein interactions, expression profiles, and disease associations. By offering data on understudied genes, Pharos helps accelerate the discovery process by highlighting targets of interest that may have otherwise been overlooked. This tool was essential for identifying the 122 t-dark genes relevant to our research interests, facilitating further investigation into their biological significance and potential clinical applications.

2.2 Rationale for Studying T-DARK Genes

The 122 T-dark genes were selected based on data from an RNA-seq analysis following the deletion of SOX2 in (NSCs derived from the postnatal day 0 forebrain of mice. The rationale behind selecting these genes lies in their significant downregulation after SOX2 was knocked out in undifferentiated, early differentiating (Day 4), and later differentiating (Day 11) NSCs. The selection focused on genes that are conserved between mouse and human, making them strong candidates for having essential roles in human neurodevelopment. This conservation implies that these genes may regulate critical functions such as NSC proliferation, GABAergic neuroblast survival, and neuronal migration, all of which are processes disrupted in Sox2 mutants.

These genes were chosen specifically because they are poorly characterized in the current scientific literature, meaning their exact roles in neurodevelopment are largely unknown.

This study aims to identify their functions by targeting these T-dark genes through high-throughput CRISPR-Cas9 screening mutagenesis in human brain organoids. The scientific question behind this selection is to determine whether these SOX2-regulated genes serve as functional effectors in neurodevelopmental processes and how their dysregulation might contribute to neurodevelopmental disorders.

3. Screening Strategies

Gene screening strategies are critical tools for understanding gene function, disease mechanisms, and identifying therapeutic targets. These approaches have evolved from single-gene studies to more comprehensive, high-throughput techniques that allow the simultaneous mutation or modulation of multiple genes. Below is an exhaustive overview of the primary gene screening strategies, with a focus on the advantages of studying multiple gene mutations simultaneously.

3.1 Classical Gene Knockout Screening

Classical gene knockout screening involves the complete deletion or disruption of a specific gene to observe its phenotypic effect. This approach has traditionally been used in model organisms such as mice, yeast, and bacteria. Gene knockouts provide clear insights into the role of a single gene, but they often miss more subtle, combinatorial interactions between genes. While effective in single-gene analysis, this method becomes inefficient when attempting to study complex traits influenced by multiple genes.

3.2 RNA Interference (RNAi) Screening

RNA interference (RNAi) is a method used to silence gene expression post-transcriptionally. High-throughput RNAi screens enable the simultaneous silencing of thousands of genes, providing insights into pathways affected by gene knockdown rather than complete knockout. RNAi is particularly useful for identifying essential genes required for survival, proliferation, or differentiation. However, off-target effects and incomplete knockdown are known limitations, reducing the precision of this technique when compared to more modern approaches like CRISPR.

3.3 CRISPR-Cas9-Based Gene Screening

CRISPR-Cas9 technology has revolutionized genome screening by enabling precise gene knockout or modulation. This method involves designing guide RNAs (gRNAs) to target specific genes, allowing the Cas9 nuclease to introduce double-strand breaks that result in loss-of-function mutations. High-throughput CRISPR screens enable the simultaneous

targeting of multiple genes, making it an ideal strategy for investigating gene function in a more comprehensive manner. CRISPR screens can be designed for genome-wide knockouts, allowing researchers to systematically disrupt every gene in a given organism or cell line.(5)(6)

3.4 Pooled CRISPR Screens

Pooled CRISPR screens allow for the simultaneous knockout or modulation of many genes in a single population of cells. In this approach, cells are transduced with a library of gRNAs, each targeting a different gene. Cells with successful gene editing are then subjected to positive or negative selection based on the phenotype of interest (e.g., drug resistance or sensitivity, cell survival, or proliferation). By sequencing the gRNAs from the selected population, researchers can identify the genes responsible for the observed phenotypes.

The advantage of pooled CRISPR screens lies in their scalability and cost-effectiveness. These screens can be applied to a wide range of cellular processes, including cell signaling, metabolism, and immune responses. By enabling the simultaneous study of thousands of genes, pooled CRISPR screens are particularly valuable for dissecting complex biological pathways and identifying key regulators of disease.

3.5 Combinatorial Gene Screens

Combinatorial gene screens are designed to assess the effects of mutating multiple genes simultaneously. This approach is especially powerful in studying genetic interactions, where the mutation of one gene may enhance or suppress the effect of another. Combinatorial screens are useful for identifying synthetic lethality, where the simultaneous disruption of two genes leads to cell death, a phenomenon that has significant implications in cancer therapy.

Gene screening strategies have evolved from single-gene approaches to high-throughput, genome-wide methods that allow for the simultaneous mutation of multiple genes and have expanded our ability to study complex biological processes and uncover genetic interactions that are essential for understanding disease mechanisms and developing novel therapies. The ability to target multiple genes simultaneously offers a powerful advantage in elucidating gene function, identifying compensatory mechanisms, and designing more effective therapeutic interventions.

4. CRISPR-Cas9

The CRISPR-Cas9 technique has revolutionized genome editing, making the process more straightforward and precise. Originally adapted from a bacterial immune system, this method allows scientists to modify DNA in living cells with high accuracy. In bacteria, CRISPR functions by storing fragments of viral DNA, which help the bacteria "remember" viruses that have attacked before. If the same virus returns, the stored DNA guides the Cas9 protein to cut and disable the viral DNA.

In a laboratory setting, this system is used to edit genomes by designing a special RNA, called a single guide RNA (sgRNA), to direct Cas9 to a specific location in the genome. The sgRNA consists of two parts: a 20-nucleotide sequence that binds to the target DNA and a scaffold RNA that connects it to the Cas9 protein. When the sgRNA binds to the correct DNA sequence, Cas9 acts like molecular scissors, cutting the DNA at that precise location.

Once the cut is made, the cell naturally attempts to repair the break. This can happen through two main processes: Non-Homologous End Joining (NHEJ) or Homology-Directed Repair (HDR). NHEJ is a quicker but error-prone method, which often introduces small mistakes, making it useful for gene knockout experiments. HDR, on the other hand, uses a DNA template to make accurate repairs, allowing for precise modifications to the genome.

For Cas9 to cut the DNA, a specific sequence called a PAM (Protospacer Adjacent Motif) must be present near the target site. The most common PAM sequence for Cas9 is "NGG" (where "N" represents any nucleotide). This ensures that Cas9 cuts only at the intended site, which enhances the accuracy of genome editing.

Despite its precision, CRISPR-Cas9 isn't flawless. There is a risk of unintended cuts at off-target sites, especially in medical applications. To address this issue, researchers have developed several improvements. These include high-fidelity versions of Cas9 that are less likely to cut in the wrong place, and shorter sgRNAs that reduce off-target effects without sacrificing on-target efficiency. Another strategy involves using two sgRNAs to guide Cas9, requiring two cuts close together, which enhances the specificity of the editing process.

Interestingly, CRISPR-Cas9 can also be used without cutting DNA at all. In techniques like CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa), scientists use a "dead" version of Cas9 (dCas9) that binds to DNA without cutting it. This allows them to either block or boost gene expression without making permanent changes to the genome.

While CRISPR-Cas9 is an incredibly powerful tool, it has sparked significant ethical debates, especially around germline editing. Changes made to an embryo’s DNA can be passed on to future generations, which raises serious concerns. There are also worries about long-term safety and unintended consequences, particularly in human gene therapy. That's why rigorous clinical trials and strong ethical guidelines are essential to ensure the technology is used responsibly.

Overall, CRISPR-Cas9 is a groundbreaking advancement in genetic engineering, offering a level of precision and versatility never seen before. But as with any powerful tool, we need to proceed with caution, carefully considering its potential risks, ethical implications, and safety challenges as the technology evolves.(7)

5. CRISPR-LICHT

The CRISPR-LICHT (CRISPR–lineage tracing at cellular resolution in heterogeneous tissue) technique is an advanced method that combines CRISPR-Cas9-mediated gene editing with lineage tracing in three-dimensional (3D) cerebral organoids. This method enables researchers to perform parallel loss-of-function (LOF) studies in human brain tissue at cellular resolution, simultaneously investigating the functional outcome of the KO of hundreds of different genes in parallel.

CRISPR-LICHT overcomes the limitations of traditional CRISPR screening methods, which are usually applied in two-dimensional (2D) cell cultures. In a 2D environment, homogeneous clonal growth is required for accurate measurements of gRNA (guide RNA) abundance and gene knockout effects. However, in cerebral organoids, where tissues develop in a more complex, heterogeneous manner, traditional methods are less effective due to uneven growth dynamics and tissue variability. CRISPR-LICHT addresses these challenges by introducing dual DNA barcoding, which links gRNA information to specific cellular lineages, thereby allowing for precise measurements even in a heterogeneous tissue context.(8)

5.1 The process

5.1.1 Inducible Cas9 Expression

(Fig 1.4) The first step in the CRISPR-LICHT methodology involves the creation of human embryonic stem cells (hESCs) that have been genetically modified to express the Cas9 endonuclease. However, rather than having Cas9 active from the start, its expression is placed under control using a tamoxifen-inducible system.

We used an eCas9 that is an enhanced version of the standard Cas9 enzyme, designed to increase editing accuracy and minimize off-target effects. Its expression is tightly controlled by the Cre-lox system, which can be activated using 4-hydroxytamoxifen (4-OHT). This approach ensures that gene editing only takes place when the system is activated by tamoxifen, providing precise regulation over the timing of gene modifications.

This inducibility is particularly important when studying developmental processes in complex tissues, such as cerebral organoids. By carefully controlling when Cas9 is activated,

researchers can selectively knock out specific genes at different stages of brain development. This enables them to track the impact of these genetic changes on processes like neural stem cell proliferation and neuronal differentiation.

In addition to temporal control, this method also provides spatial control. Since cerebral organoids develop with regional specificity that mirrors various brain regions, the technique allows researchers to study gene functions in particular tissue contexts.

5.1.2 Dual Barcoding Strategy

One of the most powerful features of the CRISPR-LICHT method is its use of a dual-barcoding system to track both the lineage of individual cells and the effects of gene knockouts. This system ensures that the lineage of each edited cell can be followed, and the impact of specific gene knockouts can be quantitatively analyzed in a complex 3D environment, like cerebral organoids.

Lineage Barcodes (LB): (Fig 1.2) These barcodes are introduced using a lentiviral vector that delivers both the gRNA (guide RNA) and a unique lineage barcode into the target cells. The gRNA (Fig 1.1) directs the Cas9 enzyme to a specific genomic location, where it induces double-strand breaks, leading to loss-of-function mutations. The lineage barcode is a sequence of nucleotides unique to each gRNA, which is integrated into the genome along with the gRNA. This allows researchers to identify which cells received specific gRNAs, essentially marking the lineage of cells derived from a common ancestor cell that received the same gRNA.

Cell Barcodes (CB): (Fig 1.3) The second barcode type is introduced during the PCR amplification step. This cell barcode allows for high-resolution tracking of individual cells. Each cell's barcode is unique, which makes it possible to count the exact number of cells derived from each lineage after gene editing. This aspect is especially important in 3D tissues, where cell proliferation or survival might be affected differently depending on the specific gene knockout. Cell barcodes increase the sensitivity of the analysis by ensuring that even small changes in cell populations are detectable.

The advantage of using two barcodes in the CRISPR-LICHT technique lies in the increased precision and sensitivity in tracking genetic perturbations within complex tissues and improve the ability to distinguish real genetic effects from inherent variability in tissue growth, providing more reliable and detailed insights in loss-of-function screens.

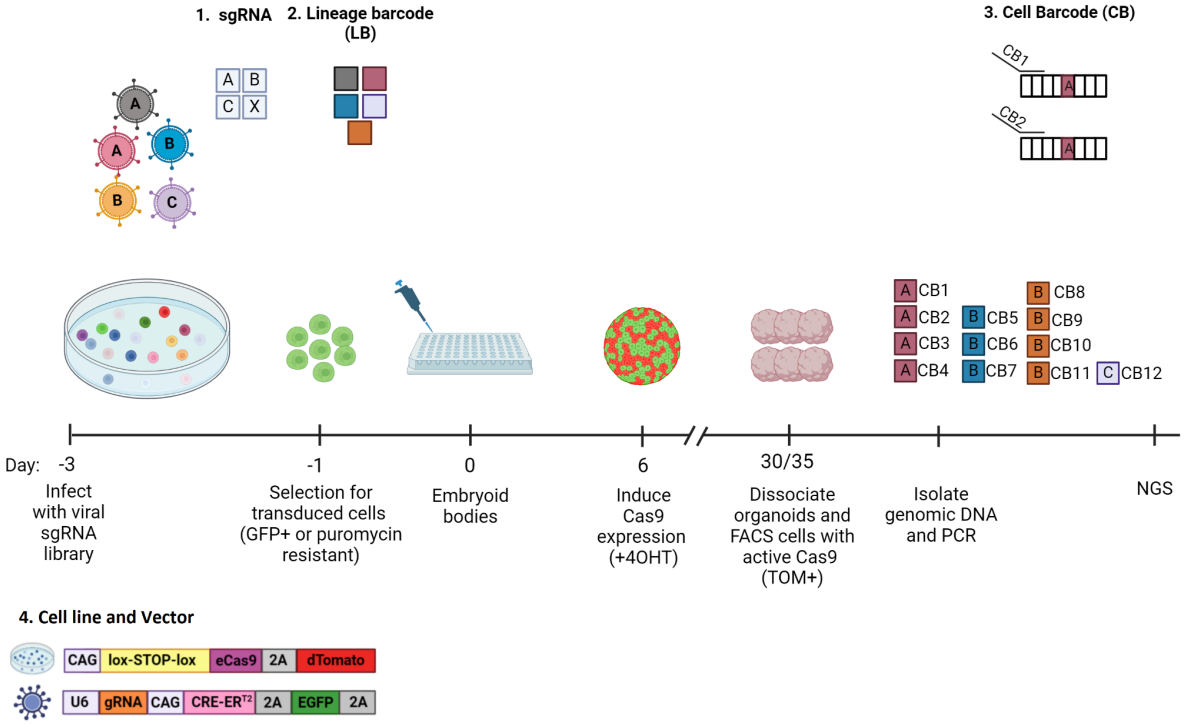

***Figure 1**: Overview of the CRISPR-LICHT workflow for inducible Cas9 expression and dual barcoding strategy in human embryonic stem cells (hESCs).*

1.1: Viral infection introduces sgRNA and lineage barcodes (LB) into target cells for gene targeting and lineage tracking.

1.2: Unique lineage barcodes (LB) are used to identify cell lineages post-editing.

1.3: Cell barcodes (CB) are added during PCR amplification, enabling precise tracking of individual cells in 3D tissues.

1.4: Genetically engineered hESCs express inducible eCas9 under control of a tamoxifen-inducible system, ensuring precise timing of gene editing.

This figure represents the process from viral infection, selection, and Cas9 activation to final analysis through genomic DNA isolation and next-generation sequencing (NGS).

5.1.3 Organoid Generation and Screening

Once the barcoding and gRNA integration are complete, the cells are cultured to form 3D cerebral organoids.

Organoids are grown from the edited stem cells, developing into complex 3D tissues that include neural progenitors, neurons, and other cell types representative of different brain

regions. This setup allows researchers to study the effects of gene knockouts in a more physiologically relevant context compared to traditional 2D culture systems.

After the organoids have formed, 4-hydroxytamoxifen (4-OHT) is added to the culture to induce Cas9 expression, triggering the gene knockouts. The use of inducible Cas9 allows for temporal control, meaning that genes can be knocked out at specific stages of development, enabling the study of gene function at different time points in neurogenesis.

5.1.4 Analysis via Next-Generation Sequencing

After the organoids are cultured for a set period, they are dissociated, and the individual cells are analyzed via next-generation sequencing (NGS).

The lineage barcodes and cell barcodes are sequenced to determine which cells carried specific gene knockouts and how those knockouts impacted their growth, differentiation, or survival within the 3D tissue. The sequencing data allows researchers to trace cell lineages and compare cell populations with different knockouts, identifying genes that are critical for brain development processes.

5.2 Conceptual Functioning

CRISPR-Cas9 lineage tracing allows the introduction of targeted genetic changes in cells, potentially leading to mutations or loss of function in specific genes. What makes this method particularly powerful is its ability to link these genetic changes to unique DNA barcodes, allowing researchers to track the effects of mutations over time.

One of the major strengths of this approach is its capacity to test loss-of-function (LOF) mutations across multiple genes simultaneously. This is especially useful for studying complex conditions like microcephaly, where understanding the genetic impact on brain size is crucial. Additionally, this technique is highly sensitive, making it ideal for detecting even subtle changes in cell populations, something that traditional methods often miss. This precision allows for a deeper understanding of how genetic mutations influence cell behavior and tissue growth, enhancing our knowledge of disease progression and potential therapeutic strategies.

5.3 Limitations of CRISPR-LICHT

One limitation is the variability inherent in organoid tissue development. Different organoids may show variations in growth patterns, making it challenging to distinguish effects caused by gene knockout from inherent tissue variability. While CRISPR-LICHT improves sensitivity through dual barcoding, the technique is not entirely free from this limitation, and distinguishing between biological noise and true gene-specific effects requires careful experimental design.

Although CRISPR-LICHT is effective in detecting moderate cell-loss phenotypes, it may not capture subtle regulatory effects that do not lead to significant changes in cell numbers. For example, certain gene knockouts may influence cellular behavior or differentiation without causing substantial cell loss or visible morphological changes. In such cases, other complementary techniques (e.g., RNA sequencing or protein analysis) may be required to fully understand the gene's function.

This technique focuses on loss-of-function studies, meaning it excels at identifying genes whose absence causes defects. However, it is less suited for studying gene regulation or subtle changes in gene expression that do not result in clear phenotypic outcomes. For instance, genes that affect cell signaling or metabolic pathways without causing immediate structural defects might be missed by this method.

6. Human Organoid Model System in Biomedical Research

The development of human organoid systems has revolutionized biomedical research, giving scientists the ability to replicate and study human organ development, homeostasis, and disease mechanisms in a lab setting. Organoids are 3D structures derived from induced pluripotent stem cells (iPSCs) or embryonic stem cells (ESCs). When grown under the right conditions, these cells self-organize into organ-like structures that mimic key aspects of real human organs(9). Researchers have created organoids for many human organs, including the brain, liver, gut, kidney, and lungs. Unlike traditional 2D cell cultures, which often fail to recreate the complexity of tissues, organoids capture the architecture, function, and cell diversity of real organs, making them a much better model for studying biology and disease(10).

Cerebral organoids, in particular, have transformed our understanding of brain development and neurological disorders. These brain-like structures can form different regions that resemble parts of the developing human brain, including cortical regions, ventricle-like cavities, and organized layers of neurons. Since they come from human stem cells, cerebral organoids allow us to study human-specific traits, like the extended development of the human cortex and unique patterns of neural growth. This is especially important for studying diseases that don't have good animal models, such as microcephaly, autism, and some forms of epilepsy(11).

The process of creating cerebral organoids begins with pluripotent stem cells, which are first guided to form embryoid bodies. These bodies are then cultivated in conditions that direct them to develop into neural tissue. Over time, they grow into complex, layered structures that resemble the human brain. One of the major benefits of organoids is that they mimic the 3D organization of real tissues, which is essential for studying brain development processes like folding, neuron migration, and the formation of cortical layers. Cerebral organoids have been particularly successful at replicating early stages of brain development, such as the growth of the ventricular zone and the formation of the cortical plate, key processes for understanding both normal brain development and the abnormalities that occur in various diseases.(12)(13)

A major application of cerebral organoids has been studying genetic disorders that affect brain development. For example, organoids have been used to model microcephaly, a condition where the brain is smaller than normal, often leading to intellectual disabilities. By growing organoids from patients with known microcephaly-causing mutations, researchers

have identified how specific genetic changes interfere with the growth, survival, or differentiation of brain cells. These findings have shed light on the molecular pathways that regulate brain growth and pointed to potential new therapeutic targets(11).

One of the most exciting uses of organoid technology is in drug discovery and personalized medicine. Organoids are better than traditional cell cultures for testing drugs because they keep the complex structure of human tissues. This makes them more reliable for checking how well a drug works or how toxic it might be. For example, in cancer research, organoids made from a patient's tumor have been used to test many different drugs, helping scientists find treatments that target the specific weaknesses of that person's cancer. This approach could change precision medicine by allowing doctors to create treatments based on the unique characteristics of each patient's cancer. Some drugs found through organoid testing are already being tried in clinical trials.

In summary, human organoids are an important model system that fills the gap between traditional cell cultures and animal models. They give us new ways to study human biology, diseases, and treatments. Organoids have many advantages: they closely copy the 3D structure and cell diversity of human tissues, they provide insights into human-specific processes that can't be studied in animals, they allow for large-scale drug testing and personalized medicine, and they help us understand complex cell interactions and how diseases progress. Overall, organoids are a powerful tool for improving our knowledge of human health and disease.(14).

RATIONALE

Understanding the role of T-DARK genes, which remain largely unexplored, is crucial for advancing our knowledge of brain development and associated disorders. These genes are a part of the so-called "dark genome," an area of the genome with unknown functions. Our interest in them stems from observations in mouse models, where RNA sequencing showed a significant reduction in the expression levels of several T-DARK genes when SOX2 was disrupted. This suggests that these genes might be regulated by SOX2 and could play important roles in brain development. Interestingly, many of these T-DARK genes are conserved in humans, indicating they might have similar functions in human brain development.

Studying these genes further could reveal new pathways and mechanisms that SOX2 controls during neural development, especially in maintaining neural stem cells and supporting their differentiation into neurons. This knowledge could be transformative, as it may help us understand neurodevelopmental disorders linked to SOX2 mutations, such as microphthalmia and cognitive impairments. If we can identify which T-DARK genes are crucial for brain development, we could potentially develop new therapies or interventions that target these pathways, offering hope for improving the lives of individuals affected by these disorders.

Another important goal was to optimize and adapt the CRISPR-LICHT technique for our specific studies. CRISPR-LICHT is an advanced technique that integrates CRISPR-Cas9 gene editing with lineage tracing in 3D cerebral organoids, allowing researchers to simultaneously screen and knock out hundreds of genes in parallel. This approach enables high-throughput loss-of-function (LOF) studies in human brain tissue at the cellular level, providing insights into the functional impact of multiple gene disruptions at once. We recognized that while CRISPR-LICHT is a powerful tool for studying gene function and lineage tracing in complex tissues, it required further adjustments to be effective in our experimental setup, particularly in cerebral organoids.

To achieve this, we focused on refining different aspects of the method, such as improving the efficiency of gene knockouts, ensuring precise timing for gene editing, and enhancing the tracking of cell lineages. This involved fine-tuning several steps of the process, from viral transduction to cell sorting, to ensure reliable results in our 3D organoid models.

By optimizing CRISPR-LICHT, we aimed to better understand the roles of specific genes in brain development and disease. Through this careful optimization process, we hoped to

create a more reliable and accurate method to answer key scientific questions about how gene function affects the growth and development of human brain tissue.

THE CRISPR-LICHT APPROACH FOR LIGHT-THROUGHPUT SCREENING OF GENE FUNCTION IN CEREBRAL ORGANOIDS

7. CRISPR-LICHT Workflow for Targeted Gene Knockouts in 3D Cerebral Organoids

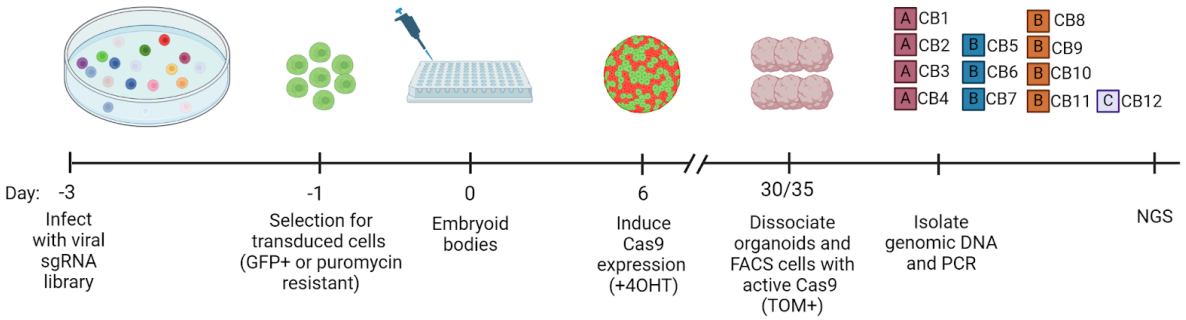

Figure 2. *CRISPR-LICHT workflow in 3D cerebral organoids. Starting on Day -3, hESCs are infected with an sgRNA library. By Day -1, cells are selected using GFP+ sorting or puromycin resistance. On Day 6, Cas9 expression is induced. Organoids are dissociated on Days 30-35, followed by FACS and genomic DNA isolation for next-generation sequencing (NGS).*

The CRISPR-LICHT methodology was designed to perform time-controlled (inducible) gene knockouts in 3D cerebral organoids, allowing for the simultaneous screening and knockout of hundreds of genes. This approach enables the study of gene function during brain development within a physiologically relevant environment, providing a comprehensive understanding of multiple genetic disruptions in parallel. Using a tamoxifen-inducible Cas9 system, we controlled the timing of gene editing, which was key for investigating how specific genes influence different stages of neurogenesis. The method's dual-barcoding system provides detailed lineage tracing, connecting genetic changes to distinct cellular behaviors. To optimize the workflow, we compared two cell selection strategies, GFP+ sorting and puromycin resistance, assessing their efficiency and reproducibility. This approach has proven especially useful for revealing the roles of genes in a 3D brain-like environment, offering deeper insights into developmental processes without the limitations of traditional 2D models.

7.1 Step-by-Step Workflow Description

Day -3: Infection with the sgRNA Library

On Day -3, we infected human embryonic stem cells (hESCs) with a lentiviral vector containing the sgRNA library and a lineage barcode (LB). This step integrated the gRNA into the genome of the cells, allowing us to perform precise gene editing and lineage tracking in the subsequent steps.

Day -1: Selection of Transduced Cells (Two Approaches Tested)

On Day -1, We selected the transduced cells using two different strategies: the first, which we will refer to as Approach 1, involved selecting GFP+ cells through sorting, while the second, which we will call Approach 2, involved selecting cells based on their resistance to puromycin.

Day 0: Formation of Embryoid Bodies

After selecting the transduced cells, we cultured them to form embryoid bodies (EBs). This transition from 2D to 3D culture allowed for the development of complex structures that better mimicked human brain development compared to traditional 2D cell cultures.

Day 6: Induction of Cas9 Expression

On Day 6, we induced Cas9 expression using 4-hydroxytamoxifen (4-OHT). The inducible Cas9 system allowed us to control the timing of gene editing, ensuring that gene knockouts occurred at specific stages of organoid development. This was particularly important for studying developmental processes over time.

Days 30-35: Organoid Dissociation and FACS

After the organoids had matured into 3D structures (by Days 30-35), we dissociated them into single cells. This step was necessary for downstream analysis. We used FACS to sort the tdTomato+ cells, which indicated successful gene editing and lineage tracing. This allowed us to collect genetically modified cells for further analysis.

Isolate Genomic DNA and PCR

After sorting, we isolate genomic DNA from the collected cells and perform PCR amplification to analyze the cell barcodes (CB) and lineage barcodes (LB). This provides a high-resolution map of cell lineage and the impact of specific gene knockouts on cell behavior.

Next-Generation Sequencing (NGS)

We perform Next-Generation Sequencing (NGS) on the isolated genomic DNA to identify which cells carried specific gene knockouts. NGS provides detailed data on how the gene knockouts affect cell proliferation, differentiation, and survival within the 3D cerebral organoids, allowing for an in-depth analysis of gene function during brain development.

This workflow demonstrates how we use the CRISPR-LICHT methodology to create gene knockouts in 3D cerebral organoids. We tested two strategies for selecting transduced cells, GFP+ sorting and puromycin selection. While GFP+ sorting offered precise cell selection, the puromycin-based approach was easier to implement without specialized equipment, making it more accessible. This workflow enabled us to study gene function in a physiologically relevant context, with final analysis done through NGS, revealing the effects of gene knockouts on brain development at the cellular level.

8. Library Design

8.1 Overview of Libraries Function

The purpose of the libraries in this CRISPR experiment is to enable the parallel analysis of gene function and cellular lineage tracing within 3D cerebral organoids. Specifically, the Oligos 10N Library is designed to introduce unique lineage barcodes into cells, allowing for the tracking of individual cell lineages after gene editing. The T-DARK Guide RNA Library targets specific T-DARK genes for knockout using CRISPR-Cas9, enabling the study of gene function in complex brain tissue. Together, these libraries allow for the precise measurement of gene knockouts and their effects on cell proliferation, differentiation, and survival in a heterogeneous 3D environment.

8.2 Oligos 10N Library (Lineage Barcode Library)

The 10N Oligos Library is important for lineage barcoding because it adds unique DNA barcodes into cells. These barcodes help scientists trace the family tree and growth of individual cells after gene editing. Each barcode is like a unique ID for cells that come from the same ancestor, allowing researchers to carefully track cell groups and their behavior in complex 3D cerebral organoids over time.

8.2.2 Rationale for design of Oligos Library

The 10N template used in the study refers to a sequence of 10 random nucleotides (N) inserted into the lentiviral constructs to serve as a lineage barcode.

The 10N random sequence in the Oligos 10N Library generates a highly diverse set of unique barcodes, allowing each infected cell and its descendants to be individually tracked throughout the experiment. This is crucial for lineage tracing in complex tissues like cerebral organoids, where understanding cellular behavior and fate relies on the ability to follow individual cell lineages. The high diversity of the 10N sequence minimizes barcode overlap,

enabling researchers to distinguish between different cell lineages even after multiple cell divisions, ensuring precise mapping of lineage fate over time. By combining the 10N barcode with gRNA sequences targeting specific genes, this system also links genetic perturbations, such as gene knockouts, to the resulting cellular phenotypes. Moreover, the scalability of the 10N barcode system makes it suitable for large populations of cells and organoids, ensuring that even in large-scale experiments, the likelihood of duplicate barcodes is extremely low. In essence, the 10N barcode provides a unique identifier for each cell, making it a powerful tool for tracking lineages and connecting genetic changes to observable outcomes in complex 3D tissues like cerebral organoids.

8.2.3 Template used for the design

The template used for the design of the 10N oligos, as shown, contains a 10-nucleotide (10N) random sequence "NNNNNNNNNN" within a fixed sequence framework. This random region allows for the creation of a diverse library of unique barcodes for lineage tracing. The template sequence begins with a specific recognition motif and ends with a region compatible with the experimental workflow, facilitating downstream processes. In this case, the oligos were ordered with desalted purification.

screening_10N_template	GAAGAGCACACGTCTGAACTCCAGTCACNNNNNNNNNNT CACACACCTAACACATCTCGTATGCCGTC

8.2.4 Amplification of Oligos 10N Library

The amplification of a 10N library using real-time PCR (qPCR), allowing for the monitoring of the reaction's progress in real time.

We performed four experimental reactions to amplify the 10N library, each containing 10N oligos, forward and reverse primers, EvaGreen qPCR Supermix, AmpliTaq Gold 360 Mastermix, and water. Additionally, we conducted two control reactions with the same setup,

but without the 10N template, to detect any non-specific amplification or contamination in the reaction mix.

The sequences of forward and reverse primers used are:

screening_10N_for	AGAGACCTCGAGCTAGCATAGATCGGAAGAGCACACGTCTGAAC
screening_10N_rev	CAATTGATAGAATTCAGATTTGACGGCATACGAGATGTG

In qPCR machine:

1. 95°C, 5 min;
2. 60°C, 10 min;
3. 95°C, 10s;
4. 60°C, 30s;
5. 72°C, 20s,
6. measure;
7. 72°C, 10s;
8. cycle back to 3. 39x; stop after measurement in logarithmic phase during step 7 at 72°C**.**

8.2.5 DNA Extraction

The rationale for DNA extraction from agarose gel after qPCR amplification of the oligos is to isolate and purify the specific amplified DNA fragment of interest. This step removes contaminants, such as non-specific amplification products, primers, and enzymes, ensuring that only the correctly sized DNA product is obtained. The purified DNA can then be used in downstream applications, such as cloning, sequencing, or further analysis, with minimal interference from unwanted by-products.

The DNA was loaded onto a 3% agarose gel and subsequently extracted using a Qiagen DNA extraction kit.

After extraction, the DNA is run on a 3% agarose gel to confirm its size and purity. The DNA was compared to a DNA ladder (Lambda DNA/HindIII Marker).

8.2.6 Digestion of Oligos 10N Library with NheI and EcoRI

The template used for generating the 10N library contains a randomized sequence of 10 nucleotides (N10), allowing for the creation of a diverse pool of sequences. After amplifying and extracting the DNA, it is digested with the restriction enzymes NheI and EcoRI.

The purpose of this step is to cut the DNA at specific recognition sites. NheI and EcoRI digestion generates sticky ends (overhangs) at the cut sites, which are essential for ligating the library into a vector (our plasmid).

The vector contain corresponding NheI and EcoRI restriction sites in their multiple cloning sites (MCS). By digesting the library DNA and the vector with the same enzymes, we created complementary sticky ends that allow for efficient ligation of the DNA into the vector.

The restriction enzyme digestion ensures that only the correctly amplified DNA is used for cloning or downstream applications. It helps eliminate any unwanted sequences or non-specific amplification products that do not contain the NheI and EcoRI recognition sites and also ensures that the insert is ligated into the vector in a specific orientation, which is crucial for maintaining the functionality of the cloned sequences.

We prepared two digestion reactions, each containing 200 ng of amplified 10N oligo mixed with NheI HF and EcoRI HF enzymes, CutSmart buffer and water.

We incubated both reactions at 37°C in a thermomixer set to 450 rpm for 1.5 hours.

After 1.5 hours, we added additional enzymes (NheI HF and EcoRI HF) to each reaction to ensure complete digestion, especially in case the initial enzyme activity had diminished over time. We then continued incubation at 37°C for an additional 1 hour.

8.2.7 Post-Digestion DNA Extraction

The rationale for DNA extraction from agarose gel after qPCR amplification of the oligos is to isolate and purify the specific amplified DNA fragment of interest.

The DNA was loaded onto a 3% agarose gel and subsequently extracted using a Qiagen DNA extraction kit.

After extraction, the DNA is run on a 3% agarose gel to confirm its size and purity. The DNA was compared to a DNA ladder (Lambda DNA/HindIII Marker)

8.3 Guide RNA Library Targeting T-DARK Genes

The gRNA T-DARK library is designed to target and knock out specific T-DARK genes in 3D cerebral organoids. Its role is to facilitate the study of these genes' functions by inducing loss-of-function mutations via CRISPR-Cas9. This allows us to observe how the disruption of T-DARK genes impacts cellular behavior, development, and phenotype in a complex tissue environment, helping to uncover their roles in brain development and function.

8.3.1 Rationale for the design of Guide RNAs

We use five gRNAs per gene to increase the chances of a successful and efficient gene knockout while minimizing variability in the results. Targeting different regions of the same gene ensures that, even if some guides are less effective or have off-target effects, others can still introduce effective mutations. This redundancy improves knockout reliability and efficiency, as some gRNAs may perform better due to differences in DNA accessibility or sequence context. Additionally, using multiple gRNAs helps mitigate off-target effects, reducing the likelihood that observed phenotypes are caused by non-specific mutations. If all the gRNAs result in similar phenotypes, it strengthens the conclusion that the changes are due to the intended gene knockout, making the results more robust and reliable for large-scale screens like CRISPR-LICHT.

8.3.2 Template used for the design

The template for the gRNA was designed in a G(N19) configuration, where the "G" represents a guanosine at the 5’ end, followed by a 19-nucleotide guide sequence (N19).

We used this structure for several reasons. First, the guanosine at the 5' end is important because it ensures efficient transcription when using the U6 promoter, which drives gRNA expression. The U6 promoter requires a guanosine at the transcription start site to function properly.

The 19-nucleotide guide sequence that follows is crucial because it directs the Cas9 enzyme to the target gene, enabling precise gene editing. This sequence is specifically designed to match a unique region of the gene we want to target, ensuring that Cas9 makes a double-strand break at the desired location.

The full template sequence, including the constant regions, is carefully constructed to facilitate the cloning and expression of the gRNAs. The variable part (represented by "X" in the template) is the guide sequence unique to each gRNA. This flexibility allows us to generate a diverse gRNA library targeting different genes, such as T-DARK genes in our study.

gRNA library (X indicates gRNA sequences)	GCTAATCGAACTCACGTGAAGACCACACCGXXXXX XXXXXXXXXXXXXXXGTTTCTGTCTTCGTCTTGGTTT GTTCCTG

8.3.3 Amplification of Guide RNA Library

The amplification of a gRNA library using real-time PCR (qPCR), allowing for the monitoring of the reaction's progress in real time.

We performed four experimental reactions to amplify the gRNA library, each containing the gRNA oligos, forward and reverse primers, EvaGreen qPCR Supermix, AmpliTaq Gold 360 Mastermix, and water. Additionally, we conducted two control reactions with the same setup but without the gRNA template, allowing us to detect any non-specific amplification or contamination in the reaction mix.

The sequences of forward and reverse primers used are:

gRNA library amplification, for	GCTAATCGAACTCACGTGAAGACCA

gRNA library amplification, rev	CAGGAACAAACCAAGACGAAGACAG

In qPCR machine:

1. 95°C, 5 min;
2. 60°C, 10 min;
3. 95°C, 10s;
4. 60°C, 30s;
5. 72°C, 20s,
6. measure;
7. 72°C, 10s;
8. cycle back to 3. 39x; stop after measurement in logarithmic phase during step 7 at 72°C.

8.3.4 DNA Extraction

The rationale for DNA extraction from agarose gel after qPCR amplification of the oligos is to isolate and purify the specific amplified DNA fragment of interest. This step removes contaminants, such as non-specific amplification products, primers, and enzymes, ensuring that only the correctly sized DNA product is obtained.

We ran the DNA on a 3% agarose gel to separate small DNA fragments. This concentration was optimal for separating the target DNA from other potential fragments or contaminants.

After electrophoresis, we excised the band corresponding to the target DNA from the gel. We then used the Qiagen DNA extraction kit to purify the DNA from the gel slice.

After extraction, we ran the DNA on a 3% agarose gel again to confirm its size and purity. We compared the DNA to a DNA ladder (Lambda DNA/HindIII Marker).

9. Vector Preparation

9.1 Plasmid Characteristics

The plasmid encodes an enhanced version of Cas9 (eCas9), which has been modified to reduce off-target effects compared to the wild-type (WT) Cas9. This improved specificity is achieved by altering the enzyme's structure to enhance its DNA-binding accuracy, ensuring it more selectively targets the intended DNA sequence. The eCas9 expression is regulated by a tamoxifen-inducible Cre-lox system, allowing researchers to temporally control when gene editing occurs, crucial for studying developmental timing in organoids. Additionally, the vector includes a dTomato reporter gene, which serves as a visual marker for successful activation of the Cas9 system and subsequent gene editing. When Cas9 is activated, cells expressing dTomato can be easily identified and isolated, providing a clear indication of effective transduction and editing events.

Alongside these components, the plasmid carries gRNAs targeting specific genes of interest. By using multiple gRNAs, the system increases the likelihood of creating effective knockouts by targeting different regions of the gene, directing eCas9 to induce double-strand breaks at precise DNA sequences.

eCas9 was chosen over WT-Cas9 due to its enhanced precision, reducing the likelihood of off-target DNA cleavage. This is particularly important in complex tissue environments like cerebral organoids or in developing cells, where unintended gene edits could disrupt normal cellular processes and lead to erroneous conclusions.

By using eCas9 and targeting multiple regions of the gene via different gRNAs, the plasmid ensures a more effective and reliable gene knockout. This is crucial in loss-of-function (LOF) studies where the goal is to completely inactivate a gene to assess its role in various biological processes.

In summary, this setup enables systematic and controlled manipulation of genes to observe their effects on cell behavior and organoid development.

This plasmide was a kind gift from C. Esk.

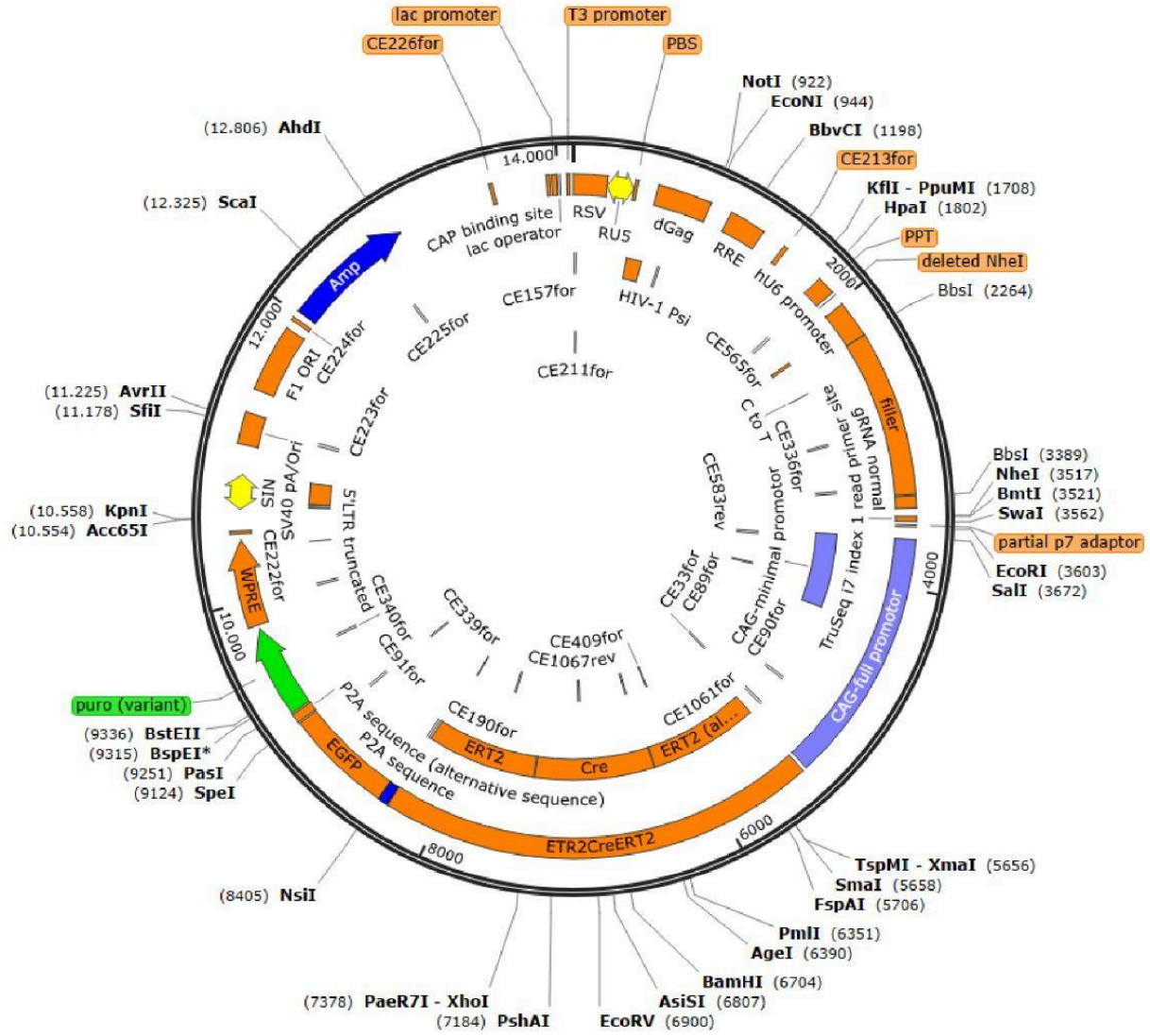

Figure 3. *Schematic representation of the plasmid vector (14,102 bp). The vector includes the hU6 promoter for gRNA expression and the CAG promoter driving the expression of ETR21CreERT2-P2A-EGFP-P2A-puromycin resistance. Enhanced green fluorescent protein (EGFP) is included for visual tracking, while the puromycin resistance gene allows for selection. The WPRE (woodchuck hepatitis virus posttranscriptional regulatory element) enhances gene expression. Restriction sites are annotated throughout the map to facilitate efficient cloning and modifications.*

9.2 Vector Digestion

9.2.1 Vector Digestion with NheI and EcoRI

The plasmid vector is digested with the same restriction enzymes,(of oligo 10N) NheI and EcoRI, for several important reasons. The first one is the creation of compatible sticky ends, both the DNA insert (in this case, the amplified 10N library) and the plasmid are digested

with NheI and EcoRI to generate compatible sticky ends. These sticky ends are single-stranded overhangs that facilitate the ligation of the insert into the plasmid in a precise and directional manner. By digesting both the insert and the vector with the same enzymes, you ensure that the ends of the insert match perfectly with the ends of the vector, allowing for efficient and specific ligation. Then using two different restriction enzymes (NheI and EcoRI) is important to ensure directional cloning. This means that the insert is ligated into the plasmid in a specific orientation. In addition digesting the vector with two different enzymes prevents it from recircularizing without the insert. When the plasmid is cut with both NheI and EcoRI, the two ends of the vector are not compatible with each other, meaning the plasmid cannot easily re-ligate to itself without an insert. This reduces the chances of obtaining empty plasmids after ligation, increasing the efficiency of obtaining the desired plasmid-insert combination.

We prepared two digestion reactions, each containing 5 µg of digested plasmid vector mixed with NheI HF and EcoRI HF enzymes, CutSmart Buffer and water.

We incubated both reactions at 37°C in a thermomixer set to 450 rpm for 1.5 hours. This allowed the restriction enzymes to cut the DNA at their specific recognition sites. After 1.5 hours, we added additional enzymes (NheI HF and EcoRI HF) to each reaction to ensure complete digestion, especially in case the initial enzyme activity had diminished over time. We then continued incubation at 37°C for an additional 1 hour

9.2.2 CIP Digestion

To prevent plasmid re-ligation following restriction digestion by treating the digested product with Calf Intestinal Phosphatase (CIP), followed by inactivation, phenol/chloroform purification, and DNA precipitation to prepare the plasmid for downstream applications.

We added CIP to the digested plasmid product then incubated the mixture at 37°C for 30 minutes. After the 30-minute incubation, the CIP enzyme must be inactivated by heating the reaction at 80°C for 4 minutes, which denatures the CIP enzyme, rendering it inactive while leaving the DNA intact.

9.2.3 Phenol/Chloroform Extraction and Ethanol Precipitation

This extraction step is essential to remove proteins, such as the inactive CIP enzyme, and other contaminants from the DNA solution. Phenol denatures proteins and partitions them into the organic phase, while the DNA remains in the aqueous phase. A subsequent chloroform wash ensures that any remaining phenol is removed, as residual phenol could interfere with downstream reactions like ligation or transformation.

Ethanol precipitation is a common technique to recover and concentrate DNA after purification. Sodium acetate helps to neutralize the negative charges on the DNA backbone, facilitating precipitation when ethanol is added. The 70% ethanol wash removes any remaining salts from the DNA pellet, ensuring that the DNA is clean for further manipulation. Air-drying the pellet ensures that residual ethanol is evaporated, as ethanol can inhibit enzymatic reactions.

We followed this steps:

1.Spin Phase Lock Gel tubes for 3 minutes at maximum speed to prepare for phase separation.

2.Work under a fume hood to ensure safety when handling phenol/chloroform.

3.If the sample volume is less than 300 µl, adjust the volume to 300 µl with Milli-Q water.

4.Add an equal volume of phenol/chloroform (1:1) to the sample.

5.Vortex the mixture thoroughly to ensure proper mixing.

6.Centrifuge at maximum speed for 10 minutes at room temperature to separate the phases.

7.Carefully collect the upper aqueous phase without disturbing the interphase.

8.Add 1/10th volume of sodium acetate (3M, pH 5.2) and mix gently by tapping or inversion.

9.Add 3 volumes of 100% ethanol, and mix thoroughly by inversion to precipitate the nucleic acids.

10.Incubate the mixture at -80°C for 30 minutes to enhance precipitation.

11.Centrifuge at maximum speed for 30 minutes at 4°C to pellet the DNA.

12.Carefully discard the supernatant without disturbing the pellet.

13.Add 200 µl of 70% ethanol, and mix gently to wash the pellet.

14.Centrifuge again for 10 minutes at room temperature at maximum speed.

15.Discard the supernatant carefully and allow the pellet to air dry—be cautious not to overdry the pellet to prevent cracking.

16.Resuspend the pellet in 30-50 µl of TE buffer to solubilize the DNA for further use.

10. Cloning of the 10N library into the vector

10.1 Ligation of 10N Insert and Vector

We designed this ligation setup to maximize the chances of successfully inserting the 10N barcode into the vector, while also incorporating proper controls to validate the efficiency and specificity of the reaction. To improve the likelihood of successful ligation, we used the insert in a three-fold molar excess. This strategy increased the chances of the vector ligating with the 10N insert by providing more opportunities for the insert to occupy the open vector, ultimately enhancing the overall efficiency of the reaction.

In addition, we ran multiple parallel ligation reactions (4-6 times) to ensure that we obtained a sufficient quantity of ligated products for downstream applications, such as transformation. Given that ligase can lose activity over time, especially in extended reactions, we added fresh ligase after two hours to keep the enzyme active throughout the entire process, thereby improving the likelihood of successful ligation.

We also included important controls in our setup. For instance, we used a vector-only control to check for any self-ligation or background re-ligation of the vector without the insert. Another control involved adding ligase to the vector but not the insert, to ensure that any resulting colonies were not due to non-specific effects of the ligase, such as random joining or recircularization of the vector.

These controls were essential to assess the specificity and success of our ligation. After allowing the reactions to incubate for two hours, we added fresh ligase to the ligation mixtures (excluding the controls) to maintain enzyme activity. Once the reactions were complete, we stopped them by placing the samples on ice, thereby preventing further ligation activity before moving on to the transformation step.

10.1.1 Ligation protocol

For the ligation of the 10N insert into the vector we followed this steps:

1. Use 200 ng of vector in a 20 µL reaction volume. Multiple reactions were prepared, specifically 6 reactions per ligation type, and a water control was included.

2. The 10N insert was added in a 3-fold molar excess relative to the vector.
3. Ligation reactions were carried out overnight (O/N).
4. Three types of ligation reactions were set up:
    - Ligation with vector + 10N insert + ligase (test ligation).
    - Ligation with vector only (no insert control).
    - Ligation with vector + ligase (control for the ligase effect without insert).

After ligation, all reactions were pooled.

10.1.2 Test-transformation into heat-shockable bacs

The rationale for this approach is centered around optimizing the transformation efficiency, ensuring accurate ligation results, and verifying the diversity of the 10N oligos:

1. Small-Scale Test of Ligation:
    - A small aliquot of both the control and sample ligation was tested before proceeding with the full-scale transformation. This step is crucial to assess the success of the ligation reaction. Only if the sample ligation resulted in at least 10-fold higher colony numbers than the controls would the full ligation be continued. This ensures that the transformation is efficient and that the ligation reaction is productive.
2. Pooling of Ligation Reactions:
    - By pooling the 6 ligations, we ensure sufficient product quantity for transformation. This pooling increases the overall likelihood of capturing a broad diversity of 10N oligos, maximizing the number of potential unique clones in the sample.
3. Transformation Conditions:
    - The transformation protocol, including the 15-minute incubation on ice, heat shock at 42°C, and recovery in LB medium without AMP, is designed to facilitate efficient uptake of the ligated DNA into Stbl3 cells. The recovery step in non-selective LB allows the cells time to express the antibiotic resistance gene before plating onto selective media.
    - The 300 µL final volume ensures enough cells for plating, with both 50 µL and 250 µL aliquots plated to optimize colony growth across varying concentrations.
4. Colony Verification via Sequencing:

- Picking 10 random colonies for sequencing provides an initial measure of ligation success and diversity. Confirming that all 10 colonies contained unique 10N oligos indicates the success of the ligation process and the efficient incorporation of a diverse set of inserts into the vector. This step is critical to ensure that the library is representative and not dominated by repeated or incorrect sequences.

By following this protocol, we confirmed the successful transformation of diverse 10N oligos, validating the ligation and transformation efficiency.

After ligation, transformations were performed using Stbl3 competent cells and after colony growth, 10 random colonies were selected from the sample ligation plates for sequencing. Sequencing confirmed that all 10 colonies contained different 10N oligos, indicating successful ligation of the insert into the vector **.**

10.1.3 Post-Ligation Purification and Concentration

The purpose of this post-ligation purification was to remove proteins, including ligase, and other contaminants that could interfere with downstream applications like transformation. Phenol extraction was used to partition proteins into the organic phase while retaining the DNA in the aqueous phase, ensuring a cleaner DNA sample. The use of Phaselock light tubes facilitated a clean separation of the phases, minimizing contamination.

Following extraction, salt and ethanol precipitation concentrated the DNA by neutralizing its charge (with NaCl) and precipitating it with ethanol. The overnight incubation at -20°C and subsequent washes with 75% ethanol ensured removal of residual salts and contaminants, yielding a purified DNA pellet. Resuspending this pellet in a small volume of 0.5x TE buffer.

This purification and concentration process ensures a high-quality, clean DNA sample for library transformation, ultimately improving cloning efficiency and reducing background noise.

10.1.4 Transformation into Ultracompetent Electrocompetent MEGA DH10 Bacteria

This protocol aims to efficiently transform and amplify the ligated DNA into MEGA DH10B bacteria using electroporation, which is a high-efficiency method suitable for large plasmid libraries. The use of ultracompetent cells increases transformation efficiency, which is crucial given the high complexity of the 10N library (approximately 4^10 diversity). By carefully plating the bacteria on LB-Amp plates and performing a dilution series. We ensured that we achieved a large enough colony count (over 3 million colonies) to cover the full diversity of the library. Subsequent pooling and expansion in liquid culture allowed for the amplification of the plasmid DNA, which was then purified using endotoxin-free maxiprep for downstream applications. This process maximized the representation of different inserts in the final library.

Protocol:

1. Plate Preparation:
    - Prepare and dry 5 large (15 cm) LB-Amp plates and 8 normal LB-Amp plates in a 37°C incubator for 2 hours.
2. Preparation of Bacteria and Medium:
    - Thaw the ultracompetent electrocompetent MEGA DH10B cells on ice.
    - Aliquot bacteria into 20 µL tubes, with each vial sufficient for 5 transformation reactions.
    - Thaw the recovery medium.
3. Electroporation:
    - Add 100 ng (maximum 1 µL) of the precipitated ligation reaction into 20 µL of bacteria and incubate on ice for 10 minutes.
    - Cool electroporation cuvettes on ice.
    - Transfer the bacteria-DNA mix into the cuvettes, ensuring no bubbles and that the bacteria are positioned correctly between the electrodes.
    - Perform electroporation using the Biorad electroporator.
4. Recovery:
    - Immediately add 1 mL of recovery medium to the cuvette and transfer to a new tube. Incubate for 1 hour at 37°C.
5. Plating:
    - Perform a dilution series of the bacteria (1:10^4, 1:10^5, 1:10^6, 1:10^7) and plate onto the prepared LB-Amp plates.
    - Plate the remainder of the transformation mix onto the 5 large 15 cm plates. Incubate all plates overnight at 37°C.

6. Colony Counting and Expansion:
    - After incubation, check colony numbers, aiming for 3+ million colonies (to cover the diversity of the 10N library, which is approximately 4^10 × 5).
    - Scrape the colonies from the plates into LB + Ampicillin and fill up to 200 mL.
    - Incubate the bacterial suspension for 1 hour at 37°C, then pellet the cells in 4 pre-weighed 50 mL tubes.
7. Plasmid Preparation:
    - Perform an endotoxin-free maxiprep following the Qiagen manual.
    - Measure DNA concentration using Nanodrop and store the purified plasmid at -20°C.

11. Cloning of gRNAs Library into the hSCR Vector with 10N Barcode Insert

11.1 Digestion of the vector using BbsI enzyme

The purpose of the BbsI digestion is to create specific overhangs required for the insertion of the gRNA sequences. BbsI is a Type IIS restriction enzyme that cuts outside of its recognition site, producing unique sticky ends. This feature is crucial for the directional cloning of the gRNA inserts, ensuring that they are inserted in the correct orientation for proper expression.

By digesting the vector with BbsI, we opened the cloning site for the integration of the gRNA sequences. We used a control reaction to confirm that the digestion was successful and the vector was ready for ligation with the gRNA insert. This step was essential for preparing the vector for the subsequent insertion of gRNAs, which we linked with the 10N barcode to enable lineage tracing in future experiments.

We mixed the hSCR vector with 10N Barcode with BbsI enzyme, CutSmart Buffer and water.

We incubated both reactions at 37°C in a thermomixer set to 450 rpm for 1.5 hours. This allowed the restriction enzymes to cut the DNA at their specific recognition sites. After 1.5 hours, we added additional enzyme (BbsI) to the reaction to ensure complete digestion, especially in case the initial enzyme activity had diminished over time. We then continued incubation at 37°C for an additional 1 hou.

Then for the same reasons as before we setted a CIP digestion and P/C extraction.

11.2 Golden Gate reaction for Cloning gRNAs into the hsCR-Vector-10N

11.2.1 Rationale for Golden Gate reaction

Golden Gate Assembly is a cloning technique that uses a Type IIS restriction enzyme (in this case, BbsI) and T4 DNA ligase to facilitate the precise and directional insertion of DNA fragments, such as gRNA sequences, into a vector. Unlike standard restriction enzymes, Type IIS enzymes cut outside their recognition site, creating unique sticky ends. This allows for seamless assembly of multiple DNA fragments in a one-pot reaction.

In this protocol, BbsI creates overhangs in both the vector and the insert, ensuring that they are ligated in the correct orientation. The cycling between 37°C and 20°C allows for repeated cutting and ligation steps. The iterative nature of the reaction improves the efficiency of the assembly, resulting in a higher likelihood of obtaining the desired recombinant plasmids. The final 75°C incubation inactivates the enzymes, stabilizing the newly formed recombinant plasmid.

This method is highly efficient for assembling complex constructs, such as inserting multiple gRNAs into the vector. Its precision, speed, and ability to handle multiple fragments in a single reaction make it ideal for constructing the diverse 10N barcode library in the hsCR vector.

Protocol:

1. Set Up Golden Gate Reaction:
    - Components for each reaction (6 reactions per library):
        - 1x Fast Digest Buffer: 2 µL
        - Vector: 100 ng
        - Insert: 3x molar ratio to the vector
        - ATP: 0.5 mM final concentration
        - H_2O: Add to a final volume of 19 µL
        - BbsI: 0.5 µL
        - T4 DNA Ligase: 0.5 µL
2. Reaction Cycling:
    - Incubate the reaction mixture using the following cycling conditions:
        - 37°C for 10 minutes (restriction enzyme activity)
        - 20°C for 10 minutes (ligation step)
        - 37°C for 5 minutes (restriction enzyme activity)

        - 20°C for 5 minutes (ligation step)
    - Repeat the above cycle 9 times to maximize the ligation efficiency.
    - Finally, incubate at 75°C for 15 minutes to inactivate the enzymes.
3. Post-Assembly Processing:
    - Combine the reactions into a single tube.
    - Perform a Plasmid Safe treatment following the manufacturer's protocol to degrade any unwanted linear DNA, if desired (optional, not strictly required).
    - Extract the DNA using phenol/chloroform purification to remove proteins and other contaminants.

11.2.2 Library Transformation

This step finalized the construction of the plasmid library by transforming the assembled gRNA constructs into bacteria, allowing for the amplification of the newly ligated vectors. The goal was to obtain a large pool of colonies to ensure that the full diversity of the gRNA library was represented. After the Golden Gate assembly, we transformed the final ligation product into competent bacterial cells using the previously described library transformation protocol (*according to protocol of paragraph 4.1.4*).

In addition to the clear band representing the plasmid with the expected gRNA and 10N barcode insert, a smaller band appeared on the gel (see Fig.3). This smaller band suggested the presence of an undesired recombinant product, indicating that not all of the plasmids in the library contained the correct insert.

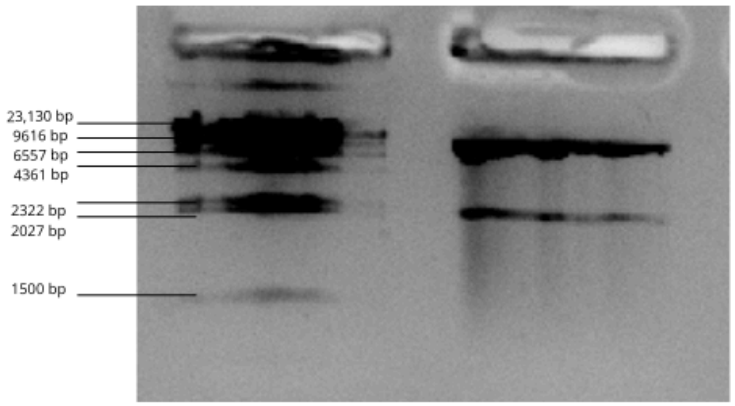

Figure 4. *Gel Electrophoresis Analysis of Plasmid Purity. The gel electrophoresis analysis shows the Lambda DNA/HindIII marker on the left and 300 ng of our plasmid on the right. The expected band*

corresponds to the plasmid containing the desired insert. However, a smaller additional band is visible below the main band, suggesting the presence of an undesired recombinant product.

11.2.3 Preventing and Managing Recombination in Lentiviral Vectors During Cloning.

Experience and literature have shown that recombination events frequently occur between the long terminal repeats (LTRs) in lentiviral vectors. In our specific vector, the 5' LTR and 3' LTR are approximately 300 base pairs apart, which corresponds to the size of the unexpected smaller band we observed on the gel, suggesting the presence of a recombinant plasmid. Recombination between these closely positioned LTRs can lead to the loss of essential vector components, resulting in smaller, undesired products.

To minimize these recombination events, we implemented specific precautions. One key approach was to conduct all bacterial growth steps at 30°C instead of the conventional 37°C. Lowering the temperature slows bacterial growth, thereby reducing the likelihood of recombination. Furthermore, the enzymes in E. coli responsible for mediating recombination are less efficient at 30°C, which further helps maintain the stability of the plasmid.

For large vectors (greater than 10 kb), like the ones we are working with, we also adjusted the antibiotic concentration. By reducing the ampicillin (AMP) concentration to 50 µg/mL and extending the growth period to up to 24 hours, we improved the selection pressure for bacteria carrying the correct plasmid. Higher antibiotic concentrations and shorter growth times can sometimes create stress conditions that inadvertently favor the selection of recombinants.

Additionally, colony size was used as an indicator of plasmid integrity. Among the colonies we obtained, the smaller colonies were more likely to contain the correct plasmid, while excessively larger colonies often contained the recombinant plasmid. The larger colonies likely grew more quickly due to the smaller size and reduced metabolic burden of the recombinant plasmid, providing a clue that they carried the undesired product.

By combining these strategies, lower growth temperatures, adjusting antibiotic concentration, and careful selection based on colony size, we aimed to enrich the correct plasmid construct and mitigate the impact of unwanted recombination events. These precautions are crucial for maintaining the integrity and diversity of our lentiviral library, especially given the inherent instability of lentiviral vectors during bacterial propagation.

After implementing the precautions, we successfully obtained a much cleaner plasmid preparation. As shown in the image (Fig. 4), the band corresponding to the suspected recombinant is now significantly reduced compared to the band of interest. This indicates that our efforts to minimize recombination, such as lowering the growth temperature to 30°C, adjusting antibiotic concentrations, and carefully selecting smaller colonies, were effective. The final result is a preparation with a higher purity of the desired plasmid, paving the way for more reliable downstream applications.

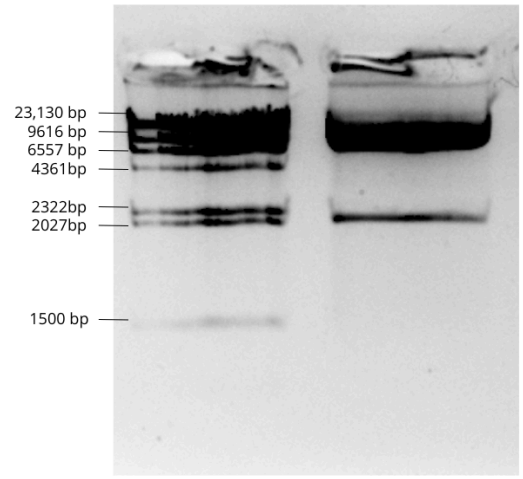

Figure 5. *Gel Electrophoresis Analysis of Plasmid Purity. Gel electrophoresis analysis showing the Lambda DNA/HindIII marker on the left and 300 ng of our plasmid on the right. The prominent band corresponds to our plasmid, while a smaller band, indicating a suspected contaminant, is visible in minimal quantity. This confirms that our optimization steps successfully minimized contamination.*

12. Lentiviral Production for Delivery of T-dark and Lineage Barcode Libraries

12.1 Overview

With both the T-dark library and the lineage barcode library successfully cloned, the next crucial step involves the production of lentiviral particles to facilitate the infection of the hESCs. The process starts with the transfection of HEK293 cells to generate viral particles that carry the desired libraries.

HEK293 cells are transfected with a mix of the plasmid containing the T-dark library and two essential packaging plasmids: psPAX2 and pMD2.G (VSV-G). The psPAX2 plasmid encodes the necessary viral proteins (gag, pol, and rev) required for assembling the viral particles, while the pMD2.G (VSV-G) provides the envelope glycoprotein, giving the virus the ability to infect a broad range of cells, including hESCs. This combination ensures the production of replication-deficient, pseudotyped lentiviral particles carrying the gRNA library.

Once transfection is complete, the viral particles are harvested from the HEK293 cell culture, followed by concentration and titration. Titration is a crucial step to determine the viral concentration, which allows us to optimize the multiplicity of infection (MOI) for the upcoming hESC transduction. Ensuring a controlled MOI is important to minimize the occurrence of multiple viral integrations in individual cells, which could complicate lineage tracing and downstream analysis.

The rationale for using a lentiviral system is its efficiency and stability in delivering genetic material to hESCs. Lentiviruses integrate the gRNA library and the lineage barcode directly into the host genome, allowing for long-term expression. The presence of the lineage barcode is vital for tracking individual cells and understanding the functional impact of gene knockouts during organoid formation.

12.2 Protocol

Day 1 (Late Afternoon): Preparing HEK293T Cells

1. Plate HEK293T cells in a 150 x 25 mm dish using DMEM with 10% FBS. Prepare 7 dishes for one virus preparation, plating approximately 4×10^6 cells per dish. This plating density should result in cells being 50% confluent by the time of transfection.

Day 2: Transfection of HEK293T Cells

1. Prepare the DNA Mix:
    - For each dish, use:
        - 6 µg of the envelope plasmid (VSVG).
        - 12 µg of the packaging plasmid (psPAX2).
        - The lentiviral plasmid encoding the insert of interest, in a 1:1:2 molar ratio with the other plasmids.
    - Add these components to 1 mL of a solution containing 0.1X TE buffer and 2.5M $CaCl_2$ (to a final concentration of 0.25M).
    - Prepare the DNA-$CaCl_2$ mix in a 50 mL tube.
2. Form the Precipitate:
    - Add an equal volume (1 mL) of 2X HBS solution to the DNA-$CaCl_2$ mix drop by drop while vortexing at full speed.
    - Allow the mixture to incubate at 25°C for 15 minutes, briefly vortexing occasionally.
3. Add the DNA Mix to Cells:
    - Carefully add the mixture drop by drop onto the HEK293T cells. High magnification microscopy should reveal a fine granular precipitate on the cell monolayer.
    - Let the CaPi-precipitated DNA remain on the cells for 6 hours.
    - After 6 hours, replace the medium with fresh DMEM (do not use chloroquine). Add 15-17 mL of fresh media per dish.

Day 5: Virus Collection

1. Collect the Supernatant:
    - Harvest the cell supernatant 72 hours post-transfection.
    - Centrifuge the supernatant at 1500 rpm for 5 minutes to remove cell debris.
    - Filter the supernatant through a 0.45 µm PVDF filter.
2. Concentrate the Virus:
    - Transfer the filtered supernatant to 50 mL tubes.
    - Centrifuge at 11000 rpm for 8 hours at 4°C to pellet the virus.

    - After centrifugation, carefully mark the area where the virus pellet is located.
3. Clean and Resuspend the Virus:
    - Carefully invert the tubes to empty the liquid.
    - Wash the tubes three times with 500 µL of PBS (avoid disturbing the pellet).
    - Resuspend the pellet in 60 µL of PBS per tube, ensuring all residues are collected.
    - Combine the resuspended virus into a 1.5 mL tube and mix gently.
    - Divide the virus into 40 µL aliquots.
4. Storage:
    - Freeze the aliquots in liquid nitrogen or on dry ice, then store them at -80°C for future use.

This optimized protocol ensures high-efficiency virus production and minimizes potential cytotoxic effects on the cells during transfection, improving the overall yield and quality of the lentiviral particles.

13. Validation of the Infection and Selection Strategy

13.1 Pilot Testing with Scramble Virus to Validate Selection and Cas9 Activation in organoids

Before proceeding with the main experiment involving the infection of hESCs with the virus containing the cloned libraries, we conducted a pilot experiment using a scramble virus. This scramble virus was designed to have the same characteristics as the final virus (as depicted in the Fig.3) but did not contain the gRNA library. The purpose of this test was to validate key aspects of the experimental workflow before introducing the actual libraries.

In the workflow shown in the image, starting at Day -3, we infected the hESCs with the scramble virus. On Day -1, we performed a selection of transduced cells based on GFP expression. This step was critical because it allowed us to confirm that our strategy for identifying transduced cells worked effectively. Successful GFP+ cell selection indicated that the viral integration and expression were functioning as intended.

The next phase involved culturing the cells into embryoid bodies (Day 0) and inducing Cas9 expression on Day 6 using 4-hydroxytamoxifen (4-OHT). This inducible system was essential for temporal control of Cas9 activation. In this test run, we aimed to ensure that the activation of Cas9 worked efficiently and that we could subsequently identify and select the cells expressing active Cas9.

Finally, around Days 30-35, we dissociated the organoids into single cells, and Fluorescence-Activated Cell Sorting (FACS) was performed to isolate TOM+ cells. The success of this step in the pilot experiment indicated that our method for activating and detecting Cas9 expression functioned as expected. This validation was crucial because, in the main experiment, the successful activation of Cas9 and the selection of the edited cells were key to linking the gRNA to specific cellular phenotypes.

In summary, the pilot experiment with the scramble virus served as a crucial quality check. It allowed us to confirm that our selection strategy for GFP+ cells was effective and that the subsequent induction and selection of Cas9-expressing cells worked smoothly. Only after this validation did we proceed with the final infection using the virus containing the gRNA library, confident that our system was correctly set up for downstream lineage tracing and functional analysis.

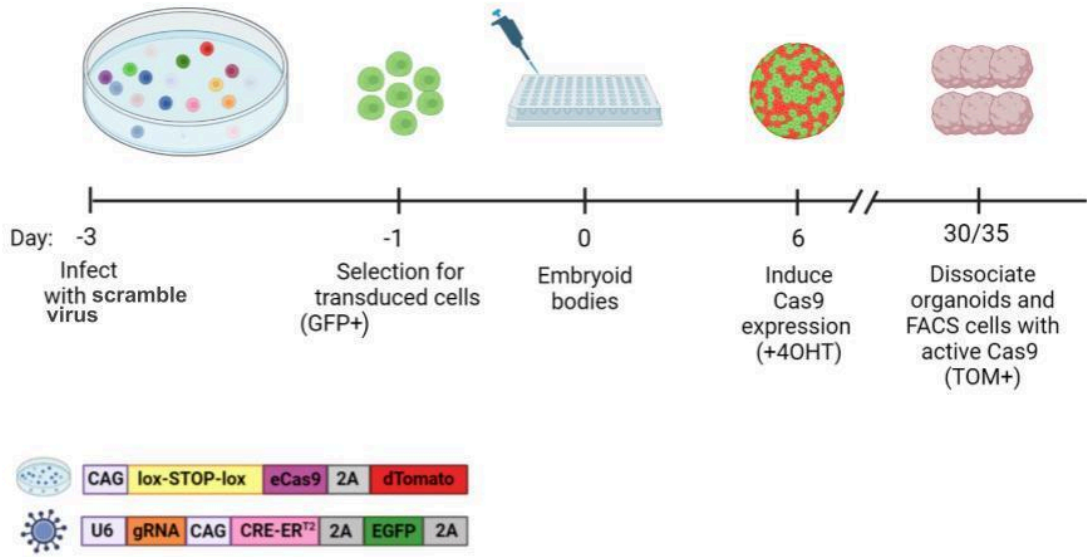

Figure 6. *The image illustrates the step-by-step process starting from Day -3, where hESCs are infected with the scramble virus. On Day -1, transduced cells are selected based on GFP expression, followed by the formation of embryoid bodies on Day 0. Cas9 expression is induced on Day 6 using 4-hydroxytamoxifen (4-OHT). Around Days 30-35, organoids are dissociated, and Fluorescence-Activated Cell Sorting (FACS) is used to isolate TOM+ cells, confirming successful Cas9 activation and workflow validation.*

13.2 Experimental Timeline and Protocol

Day -3: Infection of H9-Cas9-Tomato Cells

1. 1.5-2 hours before infection: Treat cells with Accutase and split them 1:2.5 onto the desired plates to achieve dense coverage, ensuring a large surface area for infection.
2. Infection: Add the appropriate amount of scramble virus to the infection mix prepared as per the table below.

Component	Initial Concentration	Final Concentration
Revitacell	100x	1x
P/S Penicillin/Streptomycin	100x	1x
Polybrene	1000x	1x
Synperonic	1000x	1x
E8 Medium	-	Diluent

    - Infection mix is added to 10 cm H9 plates.
    - Allow the cells to incubate with the virus for 2-3 hours.
3. Post-infection: Wash the cells 2 times with DMEM/F12 and then add E8 medium + P/S + Revitacell for overnight incubation. The plates can then be transferred to the hESC culture room.

Day -1: Selection of GFP+ Cells by Sorting

1. FACS for positive cells: After 48 hours, use FACS to sort GFP+ cells.
2. FACS Settings: Use high pressure if needed for the required number of cells. Use PBS (not FACS flow) and collect cells in 0.25-1 mL of E8 + P/S + Revitacell. Plate cells on an appropriately sized dish to achieve a fairly dense population, allowing for better recovery.
3. After attachment: After cells attach (approximately 1 hour), replace the medium with fresh E8 + P/S + Revitacell to remove residual PBS.

Day 0: Culturing Sorted Cells into Embryoid Bodies

- Plate 15,000 cells per organoid to form embryoid bodies.

Day 6: Induction of Cas9 Expression

- Add Tamoxifen to the culture medium at a final concentration of 0.3 µg/mL to induce Cas9 expression.

Day 7: Neural Induction

- Embryoid bodies (EBs) should be 500-600 µm in diameter and begin to show smooth, bright edges.
- Switch to NI medium for further neural induction.

Day 11: Embedding of Embryoid Bodies

1. Transfer the aggregates into Matrigel droplets as follows:
    - Rinse a 10 cm dish with an anti-adherence solution.
    - Use a cut P200 tip to transfer aggregates to dimples made on a sheet of parafilm.
    - Remove excess media and add Matrigel to each aggregate, centering them in the droplet using a pipette tip.
    - Incubate the parafilm sheet in a 37°C incubator for 30 minutes to polymerize the Matrigel.
    - Add NI medium + P/S and gently agitate to release the Matrigel droplets from the parafilm.
    - Place the organoids on a shelf for further incubation.

Days 13 & 14: Differentiation with Improved Medium (IMP-A) + CHIR

- Change the medium to IMP-A supplemented with CHIR (1:3300, final concentration of 3 µM).

Day 16: Differentiation with Improved Medium (IMP-A)

- Replace the medium with IMP-A (without CHIR).

Day 21: Culturing Organoids on Shaker

- Replace the medium with IMP-A and put the plate on the shaker.

Dat 25: Differentiation with Improved Medium (IMP+A)

- Change to IMP+A medium and place the organoids on a shaker set to 52 rpm.

Day 35: Dissociation for Sorting

- Dissociate organoids to obtain single cells for sorting.

Summary table

Day	**Step/Media**
Day-3	Infection with scramble virus using E8 + P/S + Revitacell.
Day-1	GFP+ selection via FACS; E8 + P/S + Revitacell for cell maintenance.
Day 0	Culture sorted cells into embryoid bodies (15,000 cells per organoid).
Day 6	Induce Cas9 with Tamoxifen (0.3 µg/mL) to activate gene editing.
Day 7	Neural induction with NI medium; EBs should be 500-600 µm in diameter.
Day 11	Embedding in Matrigel with NI medium + P/S.
Days 13&14	Differentiation with IMP-A medium supplemented with CHIR (3 µM).
Day 16	Differentiation with IMP-A medium (no CHIR).
Day 21	Culture organoids on a shaker with IMP+A medium (52 rpm).
Day 25	Change medium to IMP+A and continue culturing on a shaker.
Day 35	Organoid dissociation for sorting to obtain single cells.

14. Testing Selection Strategies for Transduced Cells on day -1

To ensure the successful selection of transduced cells, we tested two different approaches: GFP+ cell sorting and puromycin selection. Each method has its own advantages and limitations, and by comparing both, we aimed to determine the most efficient and accessible strategy for use in our experiments (see fig.2 for the workflow).

14.1 Approach 1: GFP+ Cell Sorting

In the first approach, the transduced cells expressed GFP (Green Fluorescent Protein) as a marker for successful viral transduction. We used Fluorescence-Activated Cell Sorting (FACS) to isolate the GFP+ cells, ensuring that only the successfully transduced cells were selected for further culturing and differentiation.

This method was highly specific, providing us with a purified population of GFP+ cells, but it required access to specialized sorting equipment, which was not always available.

14.2 Approach 2: Puromycin Selection

In the second approach, we selected cells based on their resistance to puromycin. The lentiviral vector carries a puromycin resistance gene, allowing only the transduced cells to survive when cultured in the presence of puromycin (0.75 µg/mL).

Advantage: This approach was simpler and did not require specialized equipment, making it easier to replicate across different laboratories. It was also less labor-intensive as it relied on antibiotic selection rather than manual sorting.

14.3 Results from the Two Different Approaches

14.3.1 Results from GFP+ Cell Sorting

Fig. 4 shows the results of Fluorescence-Activated Cell Sorting (FACS) analysis following the GFP+ selection approach used in the CRISPR-LICHT workflow. The three panels (A, B, C) correspond to different cell conditions: H9 cells not transduced (A), H9 cells transduced but not treated with tamoxifen (B), and H9 cells transduced and treated with tamoxifen (C).

Panel A (H9 non-transduced): This served as the negative control for the analysis. The gray color represented the non-fluorescent reference population, and we observed no significant GFP+ (green) or tdTomato+ (red) cells. The P2 gate showed only 0.08% of the population falling into the GFP+ region, indicating minimal background fluorescence.

Panel B (H9 transduced, no tamoxifen): In this panel, the transduced cells (without tamoxifen) showed 51.67% of GFP+ cells (green), suggesting successful transduction of the GFP reporter. There were still no tdTomato+ cells (0%) in the red P3 gate, as Cas9 had not yet been activated to induce gene editing and lineage tracing.

Panel C (H9 transduced, with tamoxifen): In this condition, the addition of 4-hydroxytamoxifen (4-OHT) activated the inducible Cas9 system, triggering gene editing and expression of the tdTomato lineage marker. As a result, we observed a population shift with 46.12% GFP+ cells and 28.06% tdTomato+ cells (red). The presence of the red tdTomato+ cells confirmed that gene editing and lineage tracing had occurred, while GFP+ cells were still present, indicating successful transduction.

This data highlights that the tamoxifen-inducible Cas9 system effectively activates gene editing, as seen by the emergence of the tdTomato+ cells in the tamoxifen-treated population.

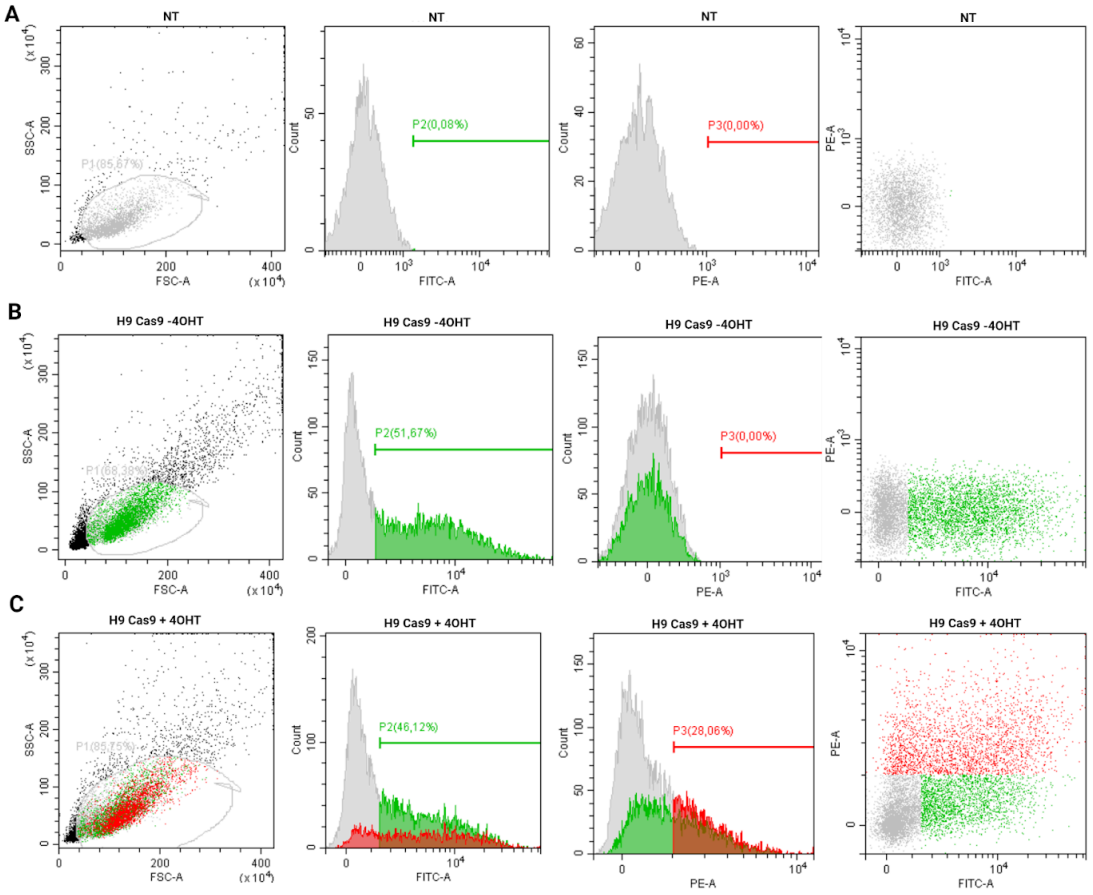

Figure 7. *FACS Analysis of H9 cas9 Cells for GFP+ and Tomato+ signal*

(A) *Non-transduced H9 cells serve as a negative control, showing no GFP+ or tdTomato+ populations.*

(B) *H9 cells transduced with the CRISPR-LICHT lentiviral vector but not treated with tamoxifen display 51.67% GFP+ cells, confirming successful transduction. No tdTomato+ cells are observed as Cas9 is not activated.*

(C) *H9 cells transduced and treated with tamoxifen show 46.12% GFP+ cells and 28.06% tdTomato+ cells, confirming both successful transduction and activation of Cas9-mediated gene editing through the tamoxifen-inducible system.*

14.3.2 Results from Puromycin Selection

Fig 5. show the FACS analysis following the puromycin-based selection approach for transduced cells in the CRISPR-LICHT workflow.

First Panel (Left) - General Population Reference: This panel shows the overall reference population of the sample. It displays the scatter plot of all cells based on FSC (Forward Scatter) and SSC (Side Scatter), which are indicators of cell size and granularity. The cells within the gated region represent the total cell population used in the subsequent analyses.

Second Panel - Singlets: This panel focuses on the singlets, which are single, intact cells as opposed to cell doublets or clusters. Singlets are a crucial indicator of the quality of the cell dissociation process. The fact that 94.29% of the cells are singlets suggests a good-quality dissociation, with minimal cell clumping or damage.

Third Panel - Live Cells: This panel highlights the live cell population. Cells that are alive and healthy are gated based on their ability to exclude dead cell markers (e.g., DAPI). Here, 83% of the cells are viable, showing that the puromycin selection and dissociation processes preserve a substantial proportion of live cells for downstream analysis.

Fourth Panel - tdTomato+ Cells: This final panel focuses on the population of tdTomato+ cells, which are indicative of successful Cas9 activation and gene editing. The panel shows that 53.57% of the live, singlet cells express the tdTomato marker.

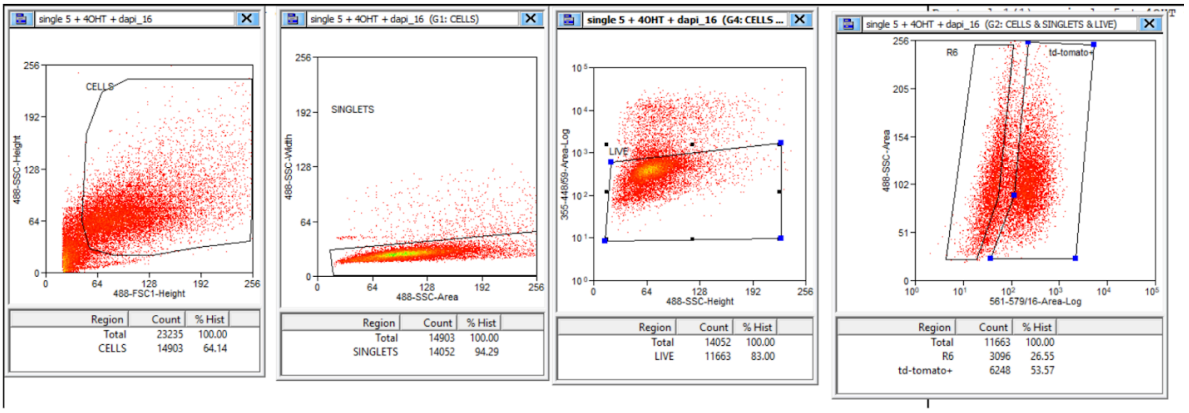

Figure 8. *FACS Analysis of H9 Cells with Puromycin Selection Approach*

First panel*: The overall reference population of the sample, showing all cells based on forward and side scatter characteristics.*

Second panel*: The singlet population, representing 94.29% of the total cells, indicating good dissociation quality.*

Third panel*: The live cell population, comprising 83% of the sample, confirming the viability of the cells after puromycin selection.*

Fourth panel*: The tdTomato+ population, where 53.57% of the live, singlet cells express the tdTomato marker.*

15. Evaluating Organoid Dissociation Strategies for Cell Sorting

15.1 Organoid Pooling and Dissociation

At days 30-35 of organoid differentiation, we tested two different dissociation strategies in parallel to evaluate their efficiency and impact on cell recovery. The first strategy involved pooling organoids at an 8:1 ratio and dissociating them using GentleMACS, a mechanical dissociation device. This method allowed us to rapidly generate a single-cell suspension from multiple organoids at once, significantly speeding up the process. The second strategy involved dissociating individual organoids mechanically, without pooling. Although this method was more time-consuming, it potentially preserved the integrity and specific characteristics of each organoid.

The rationale behind testing these two approaches was to compare the faster pooled method with the more meticulous, manual dissociation to determine which yielded better results for subsequent tdTomato+ cell sorting and lineage analysis. While the pooled dissociation was quicker, the individual dissociation method offered more precision, potentially minimizing cell loss or changes in cell characteristics that could occur during pooling. This comparison helped us decide which strategy was more suitable for high-resolution, large-scale studies in cerebral organoids.

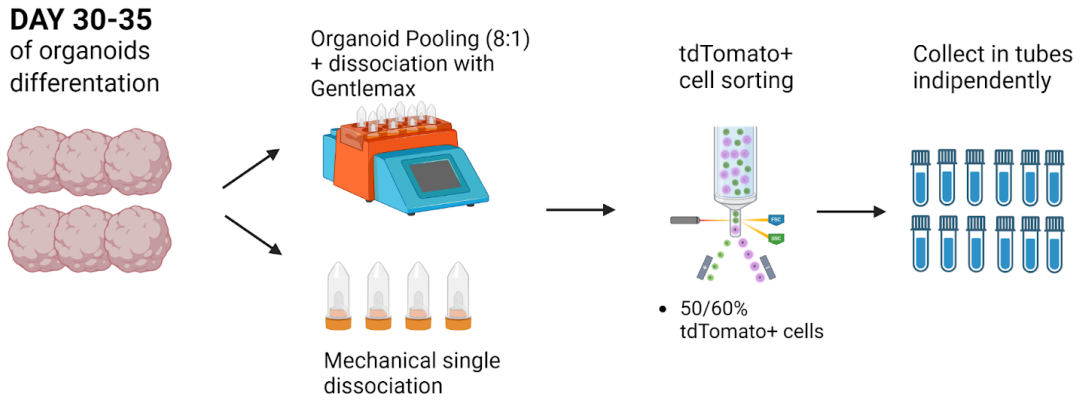

Figure 9. *Dissociation and sorting workflow for cerebral organoids on Days 30-35. Organoids were processed using two different dissociation strategies: pooling organoids at an 8:1 ratio followed by GentleMACS dissociation, or mechanically dissociating individual organoids. After dissociation, tdTomato+ cells were sorted and collected independently for further analysis.*

15.2 tdTomato+ Cell Sorting

After dissociation, we isolated the cells expressing tdTomato, a fluorescent protein marker, using fluorescence-activated cell sorting (FACS). This step allowed us to collect cells that were successfully edited and actively expressing the reporter gene linked to the lineage tracing system.

Approximately 50-60% of the sorted cells expressed tdTomato+, which we collected for further analysis.

15.3 Collection in Tubes for Independent Analysis

The sorted cells were collected into individual tubes for downstream processing, including next-generation sequencing (NGS). This step ensured that we could analyze the impact of specific gene knockouts on cell fate, differentiation, and survival with high precision. By sequencing the lineage and cell barcodes, we will be able to trace the lineage of each cell and determine the effect of gene knockouts on various aspects of brain development.

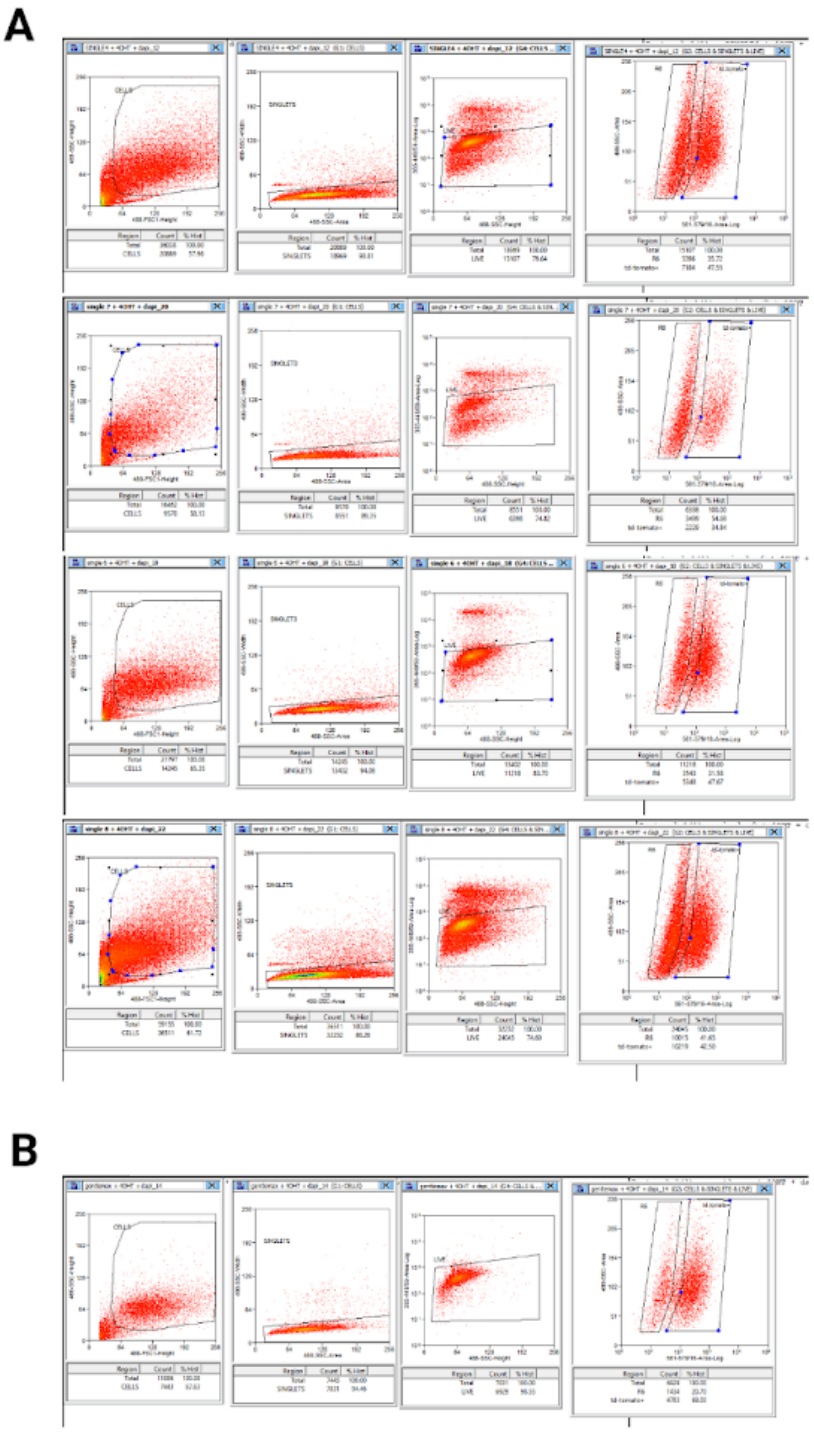

Figure 10. *FACS Analysis Comparing Individual and Pooled Organoid Dissociation.*

(A) *FACS results for four individual organoids dissociated separately. The panels show the general population, singlet population, live cells, and tdTomato+ cells, with tdTomato+ percentages ranging from 45.3% to 58.7%.*

(B) *FACS results for pooled dissociation of eight organoids, showing similar gating for the general population, singlet population, live cells, and tdTomato+ cells. The pooled dissociation resulted in*

55.6% tdTomato+ cells, consistent with the individually dissociated organoids, indicating effective gene editing and lineage tracing.

15.4 Comparison of Individual vs. Pooled Organoid Dissociation for Cell Sorting

15.4.1 Figure A - Individual Organoid Dissociation (Four Separate Organoids):

Each row in Figure A represents FACS analysis for a single organoid that was dissociated manually without pooling. Across the four organoids, the following panels are presented:

First Panel (Left) - General Population: This panel displays the scatter plot of all cells within the organoid based on FSC (Forward Scatter) and SSC (Side Scatter), which measure cell size and granularity. Cells within the gated region (highlighted in the scatter plot) represent the population used for subsequent analysis. Across the four organoids, the number of cells within the gated region is fairly consistent.

Second Panel - Singlets: The second panel isolates the singlet population of each organoid, which consists of single cells as opposed to cell clumps or doublets. The percentage of singlets is consistent across the organoids, showing good quality dissociation with minimal cell aggregation. For the four organoids, the percentage of singlets ranges from 94.7% to 96.2%, indicating high-quality dissociation.

Third Panel - Live Cells: This panel highlights the live cell population for each organoid, excluding dead cells based on their ability to exclude viability stains (e.g., DAPI). The live cell percentages range from 80.5% to 84.2%, showing that most of the cells remain viable post-dissociation.

Fourth Panel - tdTomato+ Cells: The final panel displays the tdTomato+ cells, which are indicative of successful Cas9 activation and gene editing. The percentage of tdTomato+ cells in the live singlet population ranges from 45.3% to 58.7% across the four organoids. This shows variability between individual organoids, but all have a significant population of tdTomato+ cells.

15.4.2 Figure B - Pooled Organoid Dissociation (Eight Organoids Pooled Together):

First Panel (Left) - General Population: This panel shows the scatter plot for the entire cell population from the pooled organoids, again based on FSC and SSC. The pooled organoid dissociation yields a similar scatter profile to the individual dissociations, with cells gated for further analysis.

Second Panel - Singlets: The singlet population in the pooled sample accounts for 94.8% of the total cells, indicating good dissociation quality with minimal cell clumping, comparable to the individual dissociations.

Third Panel - Live Cells: The live cell population comprises 81.9% of the total gated cells, showing that the majority of cells remain viable after pooled dissociation. This percentage is consistent with the viability seen in the individual dissociation strategy.

Fourth Panel - tdTomato+ Cells: The pooled dissociation resulted in 55.6% tdTomato+ cells, which falls within the range observed in the individually dissociated organoids. This indicates that pooled dissociation maintains efficient gene editing and lineage tracing.

DISCUSSION

16. Human Specificity in Organoid Models

One of the most significant advantages of human organoid models over traditional mouse models is their ability to replicate human-specific developmental processes. This characteristic, referred to as human specificity, is crucial for understanding diseases and biological mechanisms that are unique to human development, which cannot be effectively modeled in other species.

Human-Specific Developmental Timing and Complexity

The human body and particularly the human brain develops over a much longer timeframe compared to mice. In humans, neurogenesis and brain maturation span several months, allowing for the formation of complex structures like the cerebral cortex, which is responsible for higher cognitive functions. Human organoids, derived from pluripotent stem cells, can accurately recapitulate this prolonged development. This makes them particularly useful for studying human-specific conditions, where the slower developmental timeline is essential for observing critical processes like synapse formation, neuronal migration, and cortical layer organization.

Mouse models, on the other hand, exhibit much faster development and lack the same degree of cortical complexity. The shorter development period in mice (about five weeks for cortical neuron maturation) means that many aspects of human-specific brain development, such as the expansion of progenitor cells and the formation of the layered cortex, are poorly represented. Organoids, by contrast, provide a model that is better aligned with human biology, allowing researchers to study how certain genetic or environmental factors influence long-term brain development.(15)

Cellular Diversity and Tissue Architecture

Human organoids also possess a higher level of cellular diversity compared to mouse models. The human brain contains a more complex array of neuronal subtypes, particularly in the neocortex. These neurons are critical for intricate processes such as learning, memory, and cognition. Human organoids can model this complexity by generating the different types of neurons and glial cells found in the human brain. This is important because many neurodevelopmental diseases involve disruptions in specific neuronal subtypes, which are not fully represented in mouse models.

Additionally, the tissue architecture of human organoids closely mimics the three-dimensional structure of human organs. This includes the layering of the cortex, the organization of progenitor zones, and the formation of intricate networks of neural connections. Mouse models, while useful for studying some basic biological processes, do not exhibit the same organizational complexity, limiting their ability to model human-specific diseases that depend on proper tissue architecture and cell-cell interactions.(15)

Species-Specific Disease Mechanisms

Certain diseases, especially those related to neurodevelopment, manifest differently in humans than in mice due to species-specific differences in genetic regulation and metabolic processes. For example, some mutations or gene regulatory networks that cause severe developmental disorders in humans may not have the same effect in mice, or may not manifest at all. This makes it challenging to use mouse models to study human diseases that are influenced by these species-specific genetic factors.

Human organoids provide a platform that more accurately reflects human biology, allowing researchers to investigate disease mechanisms in a way that is directly relevant to human health. This is particularly important in the field of personalized medicine, where understanding how a specific genetic mutation affects human tissue development can guide the design of targeted therapies.(15)

In summary, human organoid models offer a highly valuable tool for studying human biology and disease due to their ability to replicate human-specific developmental processes. The extended developmental timeline, increased cellular diversity, and complex tissue architecture make them a superior alternative to mouse models, particularly when investigating diseases that are unique to humans or that involve intricate tissue structures. By providing a more accurate representation of human physiology, organoids allow researchers to study diseases in a more relevant context, leading to better insights and potential therapeutic interventions.

The study of SOX2-related diseases is significantly improved by the use of human organoid models rather than mouse models due to several species-specific differences, because there are notable limitations in the mouse models that underscore the importance of human organoids for studying diseases involving SOX2 heterozygous mutations.

Species-Specific Differences in SOX2 Functionality

In both humans and mice, SOX2 is essential for the maintenance and proliferation of neural stem cells. However, SOX2 heterozygous mutations in mice generally result in milder

phenotypes compared to humans. In human patients, SOX2 haploinsufficiency is associated with severe neurodevelopmental disorders, such as microphthalmia, brain malformations, and intellectual disabilities, which are typically not seen in mouse models carrying similar mutations. According to the study of Nicolis et al., 2004 Sox2β-geo/ΔENH heterozygous mice exhibit some abnormalities, including neurodegeneration and impaired neurogenesis, but these defects are less severe and occur in a more limited scope than in humans(16). Furthermore, it must be noted that these animals carry, in addition to the deletion of one SOX2 copy, an enhancer deletion (ΔENH). Without the latter, no phenotype was observed.

Neurodevelopmental Differences

One of the key issues when using mouse models is the difference in neurodevelopmental timing between humans and mice. The human brain undergoes a prolonged period of development, which allows for greater cellular diversity and complexity, particularly in areas like the cortex. The paper of Ferri et al., 2004 shows mouse models do not adequately replicate the prolonged expansion of neural progenitor cells that is seen in human brain development. As a result, mouse models may fail to fully capture the developmental abnormalities caused by SOX2 mutations in humans.(16)

Cellular and Structural Defects in Mice

The study (16) describes that Sox2β-geo/ΔENH heterozygous mice show structural defects, such as ventricle enlargement and neurodegeneration in specific brain regions like the thalamus and striatum. However, these defects are not as pronounced as those observed in humans with SOX2-related disorders. Additionally, the defective proliferation of neural stem cells observed in these mice is less severe than in human conditions. This suggests that while mouse models can offer some insight into SOX2 function, they do not fully replicate the human disease phenotype.(16)

Human organoid models provide a more accurate system for studying SOX2-related diseases due to their ability to mimic human-specific developmental processes. Mouse models, although useful, have significant limitations due to species-specific differences in gene function, developmental timing, and cellular complexity. These limitations make human organoids a superior model for understanding the full spectrum of neurodevelopmental abnormalities caused by SOX2 heterozygous mutations.

17. Comparison of Dissociation Strategies

In this study, we compared two dissociation strategies, individual dissociation of single organoids (Figure A) and pooled dissociation of eight organoids (Figure B), to evaluate their efficiency and impact on cell recovery and tdTomato+ cell sorting.

- **Efficiency and Time**: **Pooled dissociation** (Figure B) was significantly faster, as it allowed us to dissociate and process eight organoids simultaneously. This approach is ideal for large-scale studies where time and efficiency are critical. By comparison, the **individual dissociation** of each organoid (Figure A) was more time-consuming, as each organoid had to be processed separately.
- **Singlet Population**: Both strategies yielded a high proportion of singlet cells (94.7% to 96.2% for individual dissociation vs. 94.8% for pooled dissociation), indicating that both methods resulted in effective dissociation with minimal cell aggregation. The GentleMACS device used for pooled dissociation was able to maintain dissociation quality comparable to manual dissociation.
- **Cell Viability**: The live cell population was similarly high in both strategies, with 80.5% to 84.2% viability in the individually dissociated organoids and 81.9% viability in the pooled dissociation. This suggests that neither method caused significant cell death, and both are suitable for maintaining cell viability during the dissociation process.
- **tdTomato+ Cells**: The percentage of tdTomato+ cells was variable across the individually dissociated organoids, ranging from 45.3% to 58.7%, while the pooled dissociation resulted in 55.6% tdTomato+ cells. Despite slight variability in the individual organoids, the pooled method produced a consistent and comparable percentage of tdTomato+ cells, suggesting that it effectively preserves the outcome of Cas9-mediated gene editing and lineage tracing.

Both dissociation strategies proved effective, with the pooled dissociation method providing a faster, more efficient workflow without compromising cell viability, singlet population quality, or the proportion of tdTomato+ cells. However, the individual dissociation method offered slightly more variability between organoids, potentially preserving unique characteristics of each organoid. For large-scale studies where time and efficiency are essential, the pooled dissociation method is more practical, while individual dissociation may be preferable for more detailed studies requiring precision at the level of single organoids.

18. Comparison of GFP+ Sorting and Puromycin Selection: Specificity vs. Practicality

In comparing the two selection strategies, GFP+ sorting (Approach 1) and puromycin selection (Approach 2), we observed that both methods were effective in isolating transduced cells, but each had its own distinct advantages.

For Approach 1, the GFP+ sorting provided a very specific way to select transduced cells, as seen through the clear detection of GFP+ and tdTomato+ cells after tamoxifen induction. This approach allowed us to obtain a purified population of transduced cells and visually confirm successful gene editing through the appearance of tdTomato+ cells. However, this method requires specialized equipment, making it more complex and less accessible in some laboratory settings.

In contrast, Approach 2 with puromycin selection offered a simpler and more practical approach, as it did not rely on specialized sorting machinery. By using a resistance gene in the vector, only successfully transduced cells survived in the presence of puromycin, resulting in an enriched population of viable, edited cells. The percentage of tdTomato+ cells obtained through this method was consistent with our expectations and comparable to the GFP+ sorting results. This suggests that puromycin selection can effectively isolate transduced cells while being easier to implement and reproduce.

Overall, both strategies achieved the goal of selecting transduced cells for downstream analysis. GFP+ sorting provided high specificity, while puromycin selection offered a more straightforward, equipment-free approach. The choice between the two depends on the available resources and the specific requirements of the experimental setup.

Materials And Methods

Cell Culture

The feeder-free human embryonic stem cell (hESC) line WA09 (H9) was obtained from WiCell and confirmed to have a normal karyotype and to be free of contamination. The cells were regularly screened for mycoplasma. H9 cells were cultured on plates coated with hESC-qualified Matrigel (Corning, cat. no. 354277) using Essential 8 Medium (Thermo Fisher Scientific, Catalog # A1517001). All stem cells were kept in a 5% CO2 incubator at 37 °C. The cells were passaged by washing once with DPBS -/- (Gibco, Catalog # 14190-250), incubating them in the incubator at 37 °C for 5 minutes, then detaching and replating them in Essential 8 Medium with RevitaCell Supplement (Thermo Fisher Scientific, Catalog # A2644501).

Cloning, molecular biology and generating transgenic and KO hESCs lines

For the inducible eCas9 cell line, a transgene carrying a CAG-lox-TagBFP-loxeCas9-P2A-dTomato cassette was inserted into the AAVS1 locus in H9 cells using specific TALENs. Nucleofection was carried out using the Amaxa nucleofector (Lonza) with the Stem Cell Kit. Cells were grown for 7 days, sorted for TagBFP expression and plated. Cells were subsequently nucleofected with pSpCas9(BB)-2A-GFP (PX458, Addgene: Catalog # 48138), modified to express eCas9 instead of WT-Cas9 expressing gRNAs targeting TagBFP to abolish TagBFP signal, sorted for absence of TagBFP signal and plated as 3 single cells. Colonies were picked and genotyped to confirm transgene presence. For knock-out cell line generation, the pSpCas9(BB)-2A-GFP (PX458, Addgene: Catalog # 48138) plasmid, modified to express eCas9 instead of WT-Cas9 and dTomato instead of GFP, was used carrying a gRNA targeting IER3IP1. H9 cells were nucleofected. The modified hESCs lines were a kind gift from C. Esk (Institute of Molecular Biology, University of Innsbruck).

Cerebral organoid generation

Cerebral organoids were generated according to the Lancaster protocol (9). In brief, hESCs were cultured until they reached 60-80% confluency, and single-cell suspensions were

prepared using Accutase. The pellet cells were then resuspended in Essential 8 media with RevitaCell and counted. A total of 12,000 cells were seeded to form embryoid bodies in a 96-well ultra-low-attachment U-bottom plate (Sigma, cat. #CLS7007) in 150 µl of Essential 8 media supplemented with RevitaCell. On day 3, the media was replaced with Essential 8, and from day 6 onward, the organoids were grown as per protocol (9). In experiments where eCas9 induction was required, the media was changed on day 5 to include 0.3 µg/ml 4-Hydroxytamoxifen (Sigma-Aldrich, Catalog # H7904) for an overnight incubation.

Virus construction and viral library preparation

Two retroviral libraries based on pRSF retrovirus expressing TagBFP and containing one of two semi-random barcode libraries were used for lineage tracing. To generate the random barcode libraries, library oligos were amplified until they reached logarithmic phase (6-8 cycles, monitored in qPCR machine, stopped in extension phase). The purified PCR product was cloned into a modified pRSF retroviral vector using XhoI and EcoRI restriction enzymes. A lentivirus based on pRRL lentivirus was constructed for all gRNA deliveries using standard molecular techniques. The construct contained a U6-gRNA expression cassette along with a CAG promoter driven ERT2-Cre-ERT2-P2A-EGFP-P2A-puro cassette. For CRISPR-LICHT a lineage barcode containing random 10N bases was cloned as described above using NheI and EcoRI. Into this library a gRNA library was cloned based on an oligo library generated by TWIST Biosciences, again as described above using BbsI. For library preparations, purified ligation reactions were transformed into MegaX DH10 T1R Electrocomp Cells (Thermo Fisher Scientific, catalog # C640003) according to the manufacturer's instructions and plated onto LB media plates containing ampicillin. Dilution series were performed, and libraries were determined to contain at least 8x106 colonies. Overnight colonies were scraped off and recovered for 1 h in LB media containing ampicillin. Plasmid DNA was extracted using QIAGEN's EndoFree Plasmid Maxi Kit (Cat No./ID: 12362).

Virus production and hESC infection

Retrovirus and lentivirus were produced in HEK293 cells cultured in DMEM with 10% FBS, 2 mM L-Glutamine, 100 units/ml Penicillin, and 0.1 mg/ml Streptomycin, and coated with VSV-G using standard procedures. The supernatant was collected 72 hours post-transfection, spun at 1500 rpm for 5 minutes to remove large cell clumps, filtered through a 0.45 µm PVDF filter (Millipore), and centrifuged at 11,000 rpm for 8 hours at 4°C.

The resulting pellet was resuspended in PBS. For infection, the virus was mixed in stem cell media containing 4 µg/ml polybrene (Sigma-Aldrich, H9268) and 100 µg/ml Synperonic F108 (Sigma-Aldrich, 07579). hESCs were split at a 1:2.5 ratio one hour prior to infection using Accutase. The infections were carried out over 3 hours, followed by two washes with DMEM/F12, and fresh Essential 8 media supplemented with RevitaCell was added.

After infection

One day post-infection, single-cell suspensions of stem cells were prepared and sorted by flow cytometry for TagBFP (4% TagBFP-positive, 1.8 x 10^6 total cells sorted). The sorted cells were then plated in Essential 8 media supplemented with RevitaCell and allowed to recover overnight before initiating the organoid protocol. The time between viral library infection, sorting, and organoid setup was minimized to prevent excessive clonal expansion. Organoids and stem cells were collected at various time points from day 0 to day 35.

Dissociating organoids for flow cytometry analysis

Between days 30 and 35 of organoid differentiation, we assessed two distinct dissociation methods in parallel to determine their effectiveness and impact on cell recovery. In the first approach, we combined organoids at an 8:1 ratio and used the GentleMACS system, a device designed for mechanical dissociation. The second approach involved dissociating each organoid separately, without pooling them, using a mechanical method.

For the dissociation of organoids at days 30-35, we utilized the GentleMACS™ system alongside the Neural Tissue Dissociation Kit (P) from Milteny (Order No.: 130-093-237). The process began by preparing Mix 1, which consisted of 1900 µl of Buffer X combined with enzyme P, and preheated at 37°C. Simultaneously, Mix 2 was prepared by adding 30 µl of Buffer Y to enzyme A. The organoids were washed three times with DPBS (-/-) to ensure they were free from any residual media. Following the washing steps, the organoids were transferred to a violet C tube, and the prepared mixtures were added. The tube was sealed, gently inverted, and mounted upside down onto the GentleMACS™ system. The program B03 was initiated, with progress checked after 10 minutes, and the entire dissociation process lasted approximately 25 minutes. Once the dissociation was complete, the cells were sorted using the MoFlo Astrios system for downstream analysis.

For manual organoid dissociation, we began by preparing a solution with Trypsin 10X and Accutase at a 1:10 ratio. Specifically, 250 µl of this Trypsin/Accutase mix was added to a tube. One organoid was collected and placed into the tube containing the Trypsin/Accutase mix. The organoid was allowed to dissociate at 37°C for 10-15 minutes, either in a thermomixer set at 450 rpm or an incubator. During dissociation, we gently agitated the sample by scratching the tube on a rack.

After dissociation, we neutralized the enzymes by adding fetal bovine serum (FBS) to a final concentration of 10% in PBS. The sample was then centrifuged at 200g for 4 minutes, and the pellet was resuspended in 300 µl of the solution of interest. Finally, the dissociated cells were filtered using a cell strainer to remove any undissociated tissue fragments.

This manual dissociation protocol was used in parallel with the GentleMACS system to evaluate the differences in cell recovery and efficiency.

BIBLIOGRAPHY

1. Mercurio, S., Serra, L., & Nicolis, S. K. (2019). More than just stem cells: Functional roles of the transcription factor Sox2 in differentiated glia and neurons. *International Journal of Molecular Sciences, 20*(18), 4540. https://doi.org/10.3390/ijms20184540
2. Mercurio, S., Serra, L., Pagin, M., & Nicolis, S. K. (2022). Deconstructing Sox2 function in brain development and disease. *Cells, 11*(1604). https://doi.org/10.3390/cells11101604
3. Pevny, L. H., & Nicolis, S. K. (2010). Sox2 roles in neural stem cells. *The International Journal of Biochemistry & Cell Biology, 42*(3), 421–424. https://doi.org/10.1016/j.biocel.2009.08.018
4. Pagin, M., Pernebrink, M., Pitasi, M., Malighetti, F., Ngan, C.-Y., Ottolenghi, S., Pavesi, G., Cantù, C., & Nicolis, S. K. (2021). FOS rescues neuronal differentiation of Sox2-deleted neural stem cells by genome-wide regulation of common SOX2 and AP1 (FOS-JUN) target genes. *Cells, 10*(7), 1757. https://doi.org/10.3390/cells10071757
5. Shalem, O., Sanjana, N. E., & Zhang, F. (2015). High-throughput functional genomics using CRISPR–Cas9. *Nature Reviews Genetics, 16*(5), 299-311. https://doi.org/10.1038/nrg3899
6. Wang, T., Wei, J. J., Sabatini, D. M., & Lander, E. S. (2014). Genetic screens in human cells using the CRISPR-Cas9 system. *Science, 343*(6166), 80-84. https://doi.org/10.1126/science.1246981
7. Doudna, J. A., & Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science, 346*(6213), 1258096. https://doi.org/10.1126/science.1258096
8. Esk, C., Lindenhofer, D., Haendeler, S., Wester, R. A., Pflug, F., Schroeder, B., Bagley, J. A., Elling, U., Zuber, J., von Haeseler, A., & Knoblich, J. A. (2020). A

human tissue screen identifies a regulator of ER secretion as a brain-size determinant. *Science, 370*(6519), 935–941. https://doi.org/10.1126/science.abb5390

9. Lancaster, M. A., & Knoblich, J. A. (2014). Organogenesis in a dish: modeling development and disease using organoid technologies. *Science, 345*(6194), 1247125. https://doi.org/10.1126/science.1247125
10. Clevers, H. (2016). Modeling development and disease with organoids. *Cell, 165*(7), 1586–1597. https://doi.org/10.1016/j.cell.2016.05.082
11. Lancaster, M. A., Renner, M., Martin, C. A., Wenzel, D., Bicknell, L. S., Hurles, M. E., Homfray, T., Penninger, J. M., Jackson, A. P., & Knoblich, J. A. (2013). Cerebral organoids model human brain development and microcephaly. *Nature, 501*(7467), 373–379. https://doi.org/10.1038/nature12517
12. Kim, J., Koo, B. K., & Knoblich, J. A. (2020). Human organoids: model systems for human biology and medicine. *Nature Reviews Molecular Cell Biology, 21*(10), 571–584. https://doi.org/10.1038/s41580-020-0259-3
13. Rossi, G., Manfrin, A., & Lutolf, M. P. (2018). Progress and potential in organoid research. *Nature Reviews Genetics, 19*(11), 671-687. https://doi.org/10.1038/s41576-018-0051-9
14. Shi, Y., Sun, L., Wang, M., Liu, J., Wu, J., & Zhou, Q. (2020). Approaches to vascularizing human brain organoids. *PLOS Biology, 18*(6), e3000702. https://doi.org/10.1371/journal.pbio.3000702
15. Knoblich, J. A. (2022). Human-specific features of organoids and their relevance in disease modeling. *Cell, 185*(14), 2756-2765. https://doi.org/10.1016/j.cell.2022.06.051
16. Ferri, A. L. M., Cavallaro, M., Braida, D., Di Cristofano, A., Canta, A., Vezzani, A., Ottolenghi, S., Pandolfi, P. P., Sala, M., De Biasi, S., & Nicolis, S. K. (2004). Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development, 131*(15), 3805-3819. https://doi.org/10.1242/dev.01204