Convergence
A Journal for Young Researchers

ORIGINAL RESEARCH

Received: 9-22-26; Accepted: 10-18-26;
Managing Editor: Yehuda Silverman;
Action Editor: Yehuda Silverman

A GFP Reporter Workflow for Evaluating CRISPR Prime Editing in E. coli: A Feasibility Study Toward NR2B-Targeted Editing

Azrey Auranugraha Supriadi

Husnul Khotimah Boarding School, Kuningan Regency, West Java, Indonesia

Abstract

Ischemic stroke damages neurons in part through overactivation of NMDA receptors that contain the GluN2B (NR2B) subunit. Drugs that block the whole receptor have failed in clinical trials, so selective NR2B inactivation remains an unmet goal. CRISPR prime editing could install a loss-of-function edit in GRIN2B, but only if that edit can be reliably detected. This study tested such a verification workflow in a simple bacterial reporter system. A pegRNA was designed to change codon 5 of GFP in Escherichia coli from GAA to the stop codon TAG. Loss of fluorescence was used as a candidate screening proxy. Growth, bulk and colony fluorescence, and Sanger sequencing at the target site were assessed. All three matched the wild-type sequence, so no edit was detected in the samples tested. The fluorescence data were confounded by culture density and by colony size. No unedited GFP-expressing control was included, and sequencing was performed on pooled colonies rather than isolated ones. The result is therefore inconclusive rather than negative, and identifies detection sensitivity as the problem to solve before this approach is extended to GRIN2B.

Keywords: CRISPR prime editing, prime editing, Escherichia coli, GFP reporter, gene editing verification, Sanger sequencing, ischemic stroke, GRIN2B, stop codon insertion

1. Introduction

Ischemic stroke remains one of the leading causes of mortality and disability worldwide, with around 7.8 million new cases

and 3.6 million deaths in 2021 alone (Feigin et al., 2025). While reperfusion therapies such as mechanical thrombectomy restore blood flow, they do not directly address neuronal damage that follows the initial ischemic insult, leaving many recovered patients with permanent side effects (Dammavalam et al., 2024). This has motivated years of research into neuroprotective strategies that target cellular injury mechanisms rather than blood flow alone.

A central driver of this secondary damage is glutamate excitotoxicity. During ischemic stroke, energy failure causes excessive glutamate release and reuptake dysfunction, leading to sustained overactivation of N-methyl-D-aspartate receptors (NMDARs) on postsynaptic neurons (Yu et al., 2023). Critically, not all NMDAR subtypes contribute equally to this damage. Receptors containing the GluN2B (NR2B) subunit are strongly associated with pro-death signaling, whereas GluN2A-containing receptors tend to promote neuronal survival. This functional divergence makes NR2B a more appealing therapeutic target than the receptor as a whole. It may also explain why pan-NMDAR antagonists have repeatedly failed in clinical trials, since blocking the receptor indiscriminately inhibits the protective GluN2A pathway alongside the harmful GluN2B one (Wu et al., 2023).

Achieving pharmacological NR2B inhibition has proven difficult, motivating interest in precision genome-editing tools that can mutate a single target gene with minimal off-target effects. CRISPR prime editing is one such tool since it uses a Cas9 nickase (Cas9n) fused to an engineered reverse transcriptase, guided by a prime editing guide RNA (pegRNA) without generating double-strand breaks or requiring a donor template (Anzalone et al., 2019). Because prime editing can make precise insertions, such as premature stop codons that shorten a protein and abolish its function, it offers a candidate route to targeted, functional inactivation of NR2B (*GRIN2B*), an approach that has not been tested in a neural or disease context.

Before such an approach can be tested in a mammalian or clinical context, it is necessary to establish whether prime editing can install a loss-of-function edit. It is equally necessary to establish whether a given workflow is sensitive enough to detect that edit when it occurs. Prior work has demonstrated this principle in *Escherichia coli* by using prime editing to insert premature stop codons into the gene encoding green fluorescent protein (GFP), with loss of fluorescence functioning as an easily scored proxy for successful gene disruption (Tong et al., 2021). The present study adapts this bacterial GFP-based design as a first step toward the longer-term goal of NR2B-targeted prime editing. Using *E. coli* expressing GFP as a substitute target, we designed a pegRNA to insert a premature stop codon into the GFP coding sequence, hypothesizing that successful editing would silence GFP expression and be visually detectable as loss of fluorescence. Evaluating this approach in a simple, well-characterized system allows the verification workflow itself to be assessed before it is applied to a more demanding target. Extension to NR2B (*GRIN2B*) silencing in models of ischemic stroke remains a future objective rather than an outcome of the present work. The experimental design and the verification workflow are summarized in Figure 1.

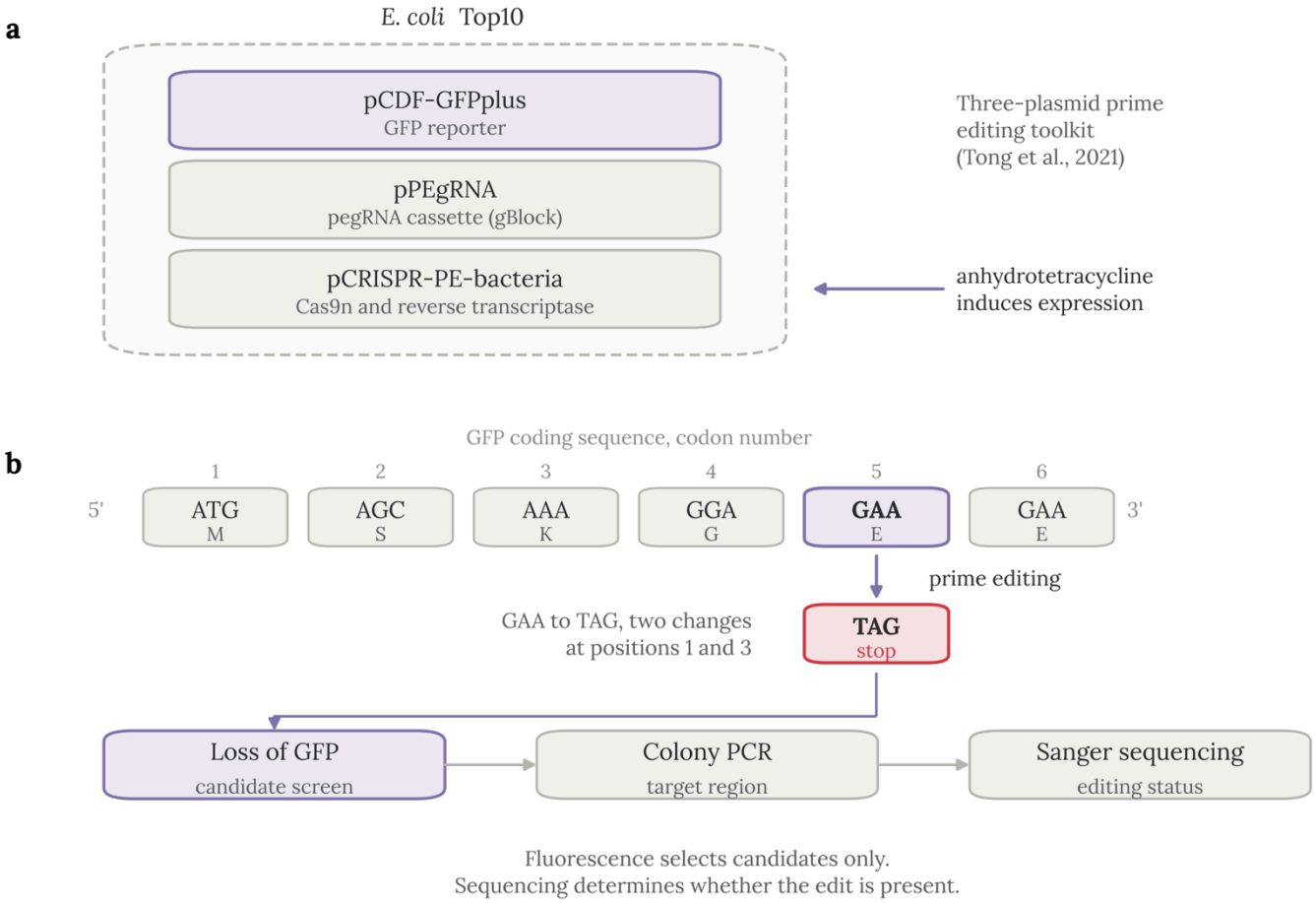


Figure 1: *Experimental design and verification workflow.*

Note. (a) The three-plasmid prime editing system in *Escherichia coli* Top10, comprising the pCDF-GFPplus reporter, the pPEgRNA plasmid carrying the pegRNA cassette, and pCRISPR-PE-bacteria encoding the Cas9 nickase and reverse transcriptase fusion, adapted from Tong et al. (2021). Expression of the editing machinery is induced with anhydrotetracycline. (b) The intended edit converts codon 5 of the GFP coding sequence from GAA to the stop codon TAG, which requires two nucleotide changes at the first and third positions. Candidate clones were screened by loss of GFP fluorescence, then assessed by PCR on isolated plasmid DNA and Sanger sequencing across the target region. Created with BioRender.

2. Methods

2.1. Strains, Plasmids, Media, and Growth Conditions

Escherichia coli Top10 cells and the plasmids used in this study were adapted from the CRISPR-Prime Editing toolkit for E. coli (Tong et al., 2021). Cultures were grown at 37°C in Luria-Bertani (LB) broth or on LB agar (Fisher Scientific). Appropriate antibiotics were supplemented at the following working concentrations: spectinomycin (50 µg/mL), ampicillin (100 µg/mL), and chloramphenicol (25 µg/mL). Anhydrotetracycline was used to induce expression of the Cas9n–reverse transcriptase fusion. For the final experiment, the working concentration was 100 ng/mL. A separate experiment set was run as an inducer dilution series covering 200, 100, and 10 ng/mL. The inducer concentration used for Experiment Set A and for the first sequencing batch was not recorded in the laboratory records available to the author. Induced cultures were grown either overnight or up to 72 hours, with the selective antibiotics present throughout. Two percent agar was used for agar plates. Antibiotics were purchased from Sigma-Aldrich (USA). Enzymes and cloning reagents were purchased from New England Biolabs (USA). Plasmid purification kits were purchased from Qiagen (Germany).

2.2. pegRNA Design and Plasmid Construction

A pegRNA was designed to insert a premature stop codon at the fifth codon of the GFP coding sequence. The cassette followed the standard prime editing guide RNA architecture. It contained a 20-nucleotide protospacer, guide RNA scaffold, a 16-nucleotide reverse-transcription template, and a 13-nucleotide primer-binding site. The protospacer was 5'-GAGAAATACTAGATGAGCAA-3', taken from the GFP coding strand. The reverse-transcription template was 5'-AAGTTCCTATCCTTTG-3 ', and the primer-binding site is the reverse complement of the 3’ end of the protospacer. The reverse-transcription template is the reverse complement of the edited coding strand, so in the coding orientation it reads CAAAGGATAGGAACTT against CAAAGGAGAAGAACTT in the unedited sequence. Converting codon 5 from GAA to TAG requires two substitutions, G to T at the first position and A to G at the third, and the template carries both. The complete 180-nucleotide cassette, including the homology arms used for assembly, is given in the Appendix.

The pegRNA cassette was synthesized as a gBlock and assembled into the empty pPEgRNA backbone by Gibson Assembly (NEB Gibson Assembly Master Mix, E2611S) using NEBuilder-compatible overhangs. The backbone was amplified by PCR using Q5 High-Fidelity 2X Master Mix (NEB, M0492S), using the primers 5'-TTTTTTTGAAGCTTGGGCCC-3' and 5'-ACTAGTATTATACCTAGGAC-3'. Each 25 µL reaction contained 12.5 µL of master mix, 1.5 µL of each primer, template DNA, and water to volume. The amplified backbone was checked by agarose gel electrophoresis against a 1 Kb Plus DNA Ladder before assembly and transformation.

2.3. Transformation and Clone Selection

Assembled plasmid DNA was introduced into chemically competent Top10 cells by heat-shock transformation; cells were equilibrated on ice with the Gibson Assembly product for 5 minutes, heat-shocked at 42°C for 30 seconds, returned to ice, recovered in SOC medium at 37°C for 1 hour, and plated on LB agar containing the appropriate antibiotic. Individual colonies were picked from each plate and combined into a single liquid culture for downstream screening.

2.4. Plasmid Isolation and PCR Screening

Plasmid DNA was isolated from overnight cultures using the QIAprep Spin Miniprep Kit (Qiagen, 27104). PCR products were purified using the QIAquick PCR Purification Kit (Qiagen, 28104) and resolved on agarose gels (Sigma, A6013) stained with SYBR Safe (Thermo Fisher, S33102) in 1X TAE buffer, alongside a 1 Kb Plus DNA Ladder (Thermo Fisher, 10787018).

2.5. Fluorescence and Growth Measurements

Bacterial growth (OD_{600}) and GFP fluorescence intensity were measured using a SPARK multimode microplate reader in a 96-well plate format. One column of wells containing LB medium only (no bacteria) was included on each plate as a medium blank and read alongside the samples. The spreadsheet in which any blank subtraction would have been applied is no longer available, so the OD_{600} and fluorescence values reported here are the readings as exported by the instrument, without blank subtraction. Fluorescence is reported in arbitrary units (a.u.). Colonies were additionally imaged under UV illumination using a gel documentation system (SGC-W5000 Plus, Servicebio) to visually assess fluorescence.

2.6. Fluorescence Image Analysis

Fluorescence images were quantified using Fiji (ImageJ). A region of interest was drawn around each of six individual colonies per plate using the ROI Manager. A seventh region of interest was placed on an area of agar with no colonies, to serve as a background measurement. For every region, the Area, Mean, Min, Max, IntDen, and RawIntDen measures were recorded. Background-corrected RawIntDen was calculated by subtracting the background region's RawIntDen from each colony value. The background region was not matched in area to the colony regions. It covered 289 pixels against 306 to 484 pixels for Experiment Set A, and 81 pixels against 96 to 154 pixels for Sample C_f. A fixed total was therefore subtracted from regions of differing size, which under-corrects larger regions and over-corrects smaller ones. A size-independent measure was calculated for this reason, as the colony mean gray value minus the background mean gray value. The six colonies on each plate were picked without reference to their brightness, and no formal randomization procedure was used. Measurements were exported to Microsoft Excel for analysis. Per-colony values are given in Supplementary Table S1.

Replication was as follows. Bulk growth and fluorescence were measured once per condition at each time point, so Figure 2 represents a single biological replicate with no technical replicates. Colony fluorescence was quantified for six colonies on each of two independent platings. No inferential statistical tests were performed, because the design does not support them. Summary values are reported as means, ranges, and coefficients of variation.

2.7. Assessment of Editing Status by Sanger Sequencing

Candidate clones were selected on the basis of reduced GFP fluorescence and assessed at the DNA level by PCR on plasmid DNA isolated from the pooled cultures, followed by Sanger sequencing across the target region, in order to determine whether the intended stop codon was present at codon 5 of the GFP coding sequence. Selection on fluorescence identifies candidates only. It does not establish that an edit occurred.

2.8. Public Expression Data

The dataset GSE38037 (Wang & Zhang, 2013) was retrieved from the NCBI Gene Expression Omnibus (platform GPL3631). Public expression data were used to establish whether the intended target gene is altered in ischemic brain tissue. The dataset GSE38037 was retrieved from the NCBI Gene Expression Omnibus (platform GPL3631). It comprises cerebral cortex from Sprague Dawley rats after transient middle cerebral artery occlusion (tMCAO, n = 4, GSM932851 to GSM932854) and from sham-operated controls (n = 4, GSM932847 to GSM932850). The array is two colors. Every sample, tMCAO and sham alike, was co-hybridized against a common reference pool of sham-operated cortex, so each value is a ratio against that shared reference. Differential expression between the two groups was computed with GEO2R on October 7, 2026, using its default settings. GEO2R fits a limma empirical-Bayes linear model to the series matrix and corrects for multiple testing across all 23,376 probes by the Benjamini-Hochberg procedure. No new experimental data were generated for this analysis.

3. Results

3.1. Growth and Fluorescence Measurements

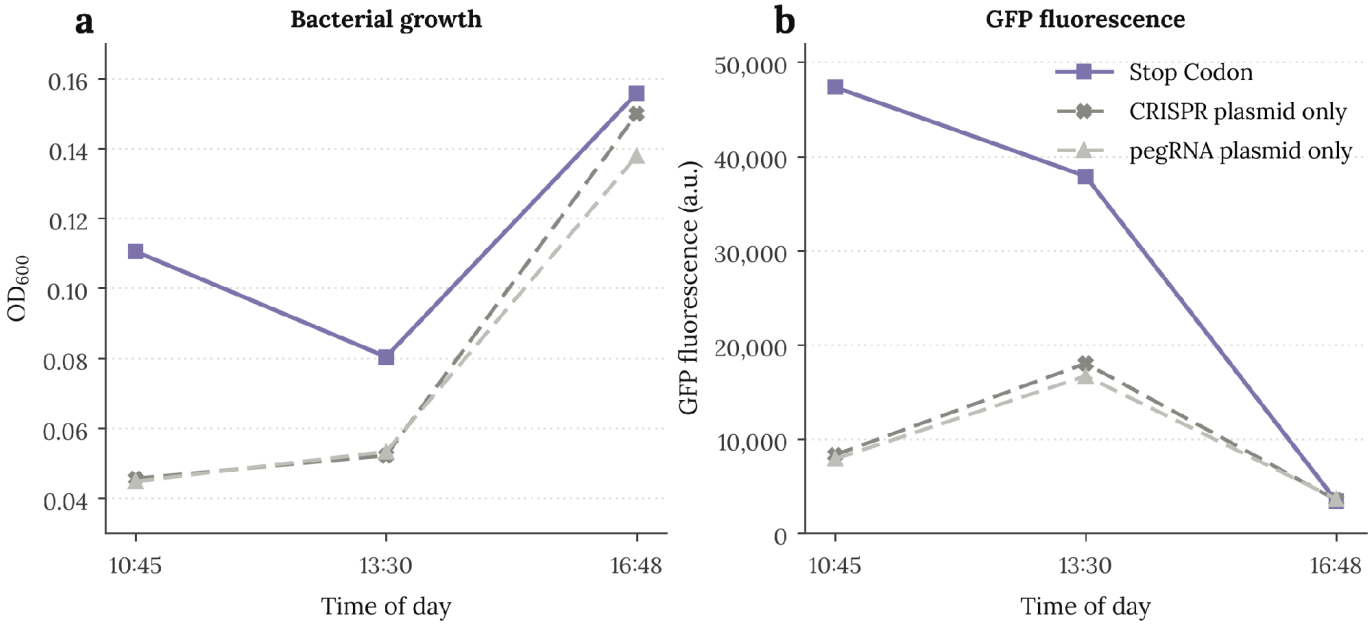


Figure 2: *Bacterial growth and GFP fluorescence for the Stop Codon construct and the plasmid-only controls.*

Note. (a) OD_{600} and (b) raw GFP fluorescence at three time points (10:45, 13:30, and 16:48) for the Stop Codon construct, the CRISPR plasmid-only control, and the pegRNA plasmid-only control. Each condition is a single culture measured at the three time points, so n = 1 biological replicate with no technical replicates. Neither control carries the GFP reporter plasmid, so the control signal reflects background autofluorescence rather than unedited GFP expression. Fluorescence

normalized to culture density is given in Table 1. Because the medium blank was not subtracted, the OD_{600} values include a contribution from the medium itself, so the normalized values in Table 1 are lower than they would be after correction.

Bacterial growth (OD_{600}) was monitored at three time points (10:45, 13:30, and 16:48) for the Stop Codon construct and the two plasmid-only controls (Fig. 2a). All three conditions grew comparably and reached similar final OD_{600} values of 0.14 to 0.16. The Stop Codon condition was a single culture measured at these three time points, so n = 1 biological replicate.

Raw GFP fluorescence in the Stop Codon construct fell across the day, from 47,344 a.u. at 10:45 to 37,891 a.u. at 13:30 and 3,417 a.u. at 16:48 (Fig. 2b). Both plasmid-only controls stayed low throughout, between 7,879 and 18,018 a.u. By the final timepoint, all three samples had converted to a low range of 3,417 to 3,631 a.u. This convergence occurred in the Stop Codon construct and in both controls alike, so it cannot be a signature of editing.

Neither control carries the GFP reporter plasmid. Their signal therefore reflects background autofluorescence and not unedited GFP expression. They serve as plasmid-only controls rather than as unedited-GFP comparators. The study includes no GFP-expressing condition that received no pegRNA. Without that condition, there is no baseline from which fluorescence could be described as lost, so the fluorescence data cannot speak to editing status.

Fluorescence scales with the number of cells present, so each reading was divided by the OD_{600} of the same sample (Table 1). Normalized values for the Stop Codon construct were 428,452, 471,281, and 21,946 a.u./OD_{600} across the three time points. Relative to the mean of the two controls, the Stop Codon construct was 2.39 times higher at 10:45, 1.43 times higher at 13:30, and 0.88 times as high at 16:48. The equivalent raw ratios were 5.85, 2.18, and 0.96. Normalization therefore removes more than half of the early difference. It also places the Stop Codon construct slightly below both controls at the final timepoint, where the raw data had the three conditions roughly equal. Because the controls carry no GFP, the difference that remains confirms GFP expression in the Stop Codon construct and does not bear on editing status.

Table 1: *GFP fluorescence normalized to culture density (a.u./OD_{600})*

Condition	10:45	13:30	16:48
Stop Codon	428,452	471,281	21,946
CRISPR plasmid only	182,259	345,172	23,485
pegRNA plasmid only	175,871	312,964	26,350

Note. Fluorescence divided by the OD_{600} of the same sample at the same timepoint, for the Stop Codon construct and the two plasmid-only controls at each of the three timepoints.

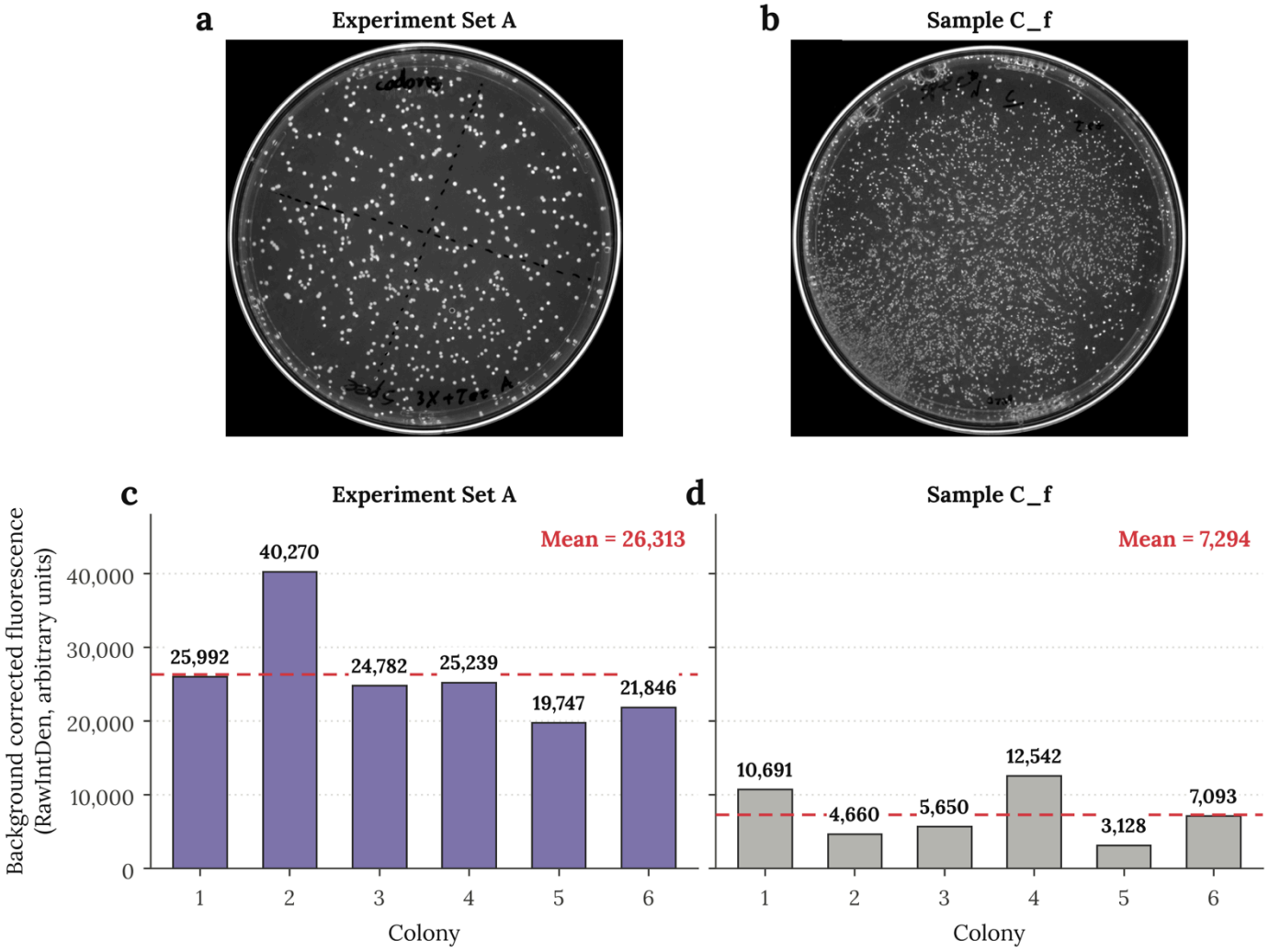


Figure 3: *Plate density and colony fluorescence for two independent platings of the Stop Codon construct under anhydrotetracycline induction.*

Note. (a) Plate image for Experiment Set A, showing colonies at moderate density. (b) Plate image for Sample C_f, showing markedly higher density with regions approaching confluence. (c, d) Background-corrected RawIntDen for six colonies from each plate, Experiment Set A in (c) and Sample C_f in (d), plotted on a common y-axis. RawIntDen is the sum of gray values within a region of interest and therefore scales with colony area, which differs 2.85-fold between the two plates. The size-independent per-pixel measure differs only 1.55-fold. Per-colony values, region areas and per-pixel intensities are given in Supplementary Table S1. The two plates are separate platings of the same construct and differ substantially in colony density.

Plate images and colony fluorescence for two independent platings are shown in Fig. 3. The plate for Experiment Set A (Fig. 3a) carried colonies at moderate density that were individually distinguishable. The plate for Sample C_f (Fig. 3b) was markedly denser, with several regions approaching confluence.

Background-corrected RawIntDen was then measured for six colonies on each plate. In Experiment Set A (Fig. 3c), values ranged from 19,747 to 40,270, with a mean of 26,313. In Sample C_f (Fig. 3d), values ranged from 3,128 to 12,542, with a mean of 7,294. The two means differ by a factor of about 3.6. RawIntDen is the sum of gray values inside a region of interest, so it scales with the area of that region. Dividing RawIntDen by the mean gray value recovers the pixel count of each region. Colonies in Set A averaged 355 pixels, and those in C_f averaged 124 pixels, a difference of 2.85 times in area. Most of the 3.6-fold difference in RawIntDen is therefore a difference in colony size rather than in brightness.

A size-independent measure was calculated as the colony's mean gray value minus the background mean gray value. On this measure, Set A averaged 65.1, and C_f averaged 41.9, a difference of 1.55 times. The two groups also overlap. C_f colonies 4 and 6 (66.9 and 62.1) were brighter than the two dimmest colonies in Set A (56.8 and 56.0). Variability was much higher in C_f, with a coefficient of variation of 48.2% compared with 12.9% for Set A. Per-colony values and pixel counts are given in Supplementary Table S1.

These two plates are two separate platings of the same construct that differed substantially in colony density, so plate-to-plate technical variation cannot be separated from any real difference in brightness. The six colonies on each plate were picked without reference to their brightness, and no formal randomization procedure was used. This comparison is therefore descriptive and cannot be read as a difference in editing status.

3.2. Editing Status at the Target Site

Three Stop Codon samples were sequenced across the codon 5 target site to test directly whether the intended edit was present. Each was compared with an unedited no pegRNA reference (Fig. 4a). The three test samples, C_t_1 (well B08, forward read, Fig. 4b), C_f (well H04, reverse read, Fig. 4c), and C_d (well D05, reverse read, Fig. 4d), read GAA, TTC, and TTC at codon 5, respectively (TTC being the reverse complement of GAA on the reverse strand), matching the sequence in Fig. 4a. None of the three samples shows the TAG sequence. In addition, average Sanger sequencing quality scores across the 21 bp window centered on codon 5 (Fig. 4e) were 57.4 for C_t_1, 54.2 for C_f, and 59.4 for C_d, compared with 59.0 for the unedited control (dashed red line). C_t_1, C_f, and C_d were sequenced from pooled-colony liquid cultures, whereas Tong et al. (2021) sequenced isolated single colonies.

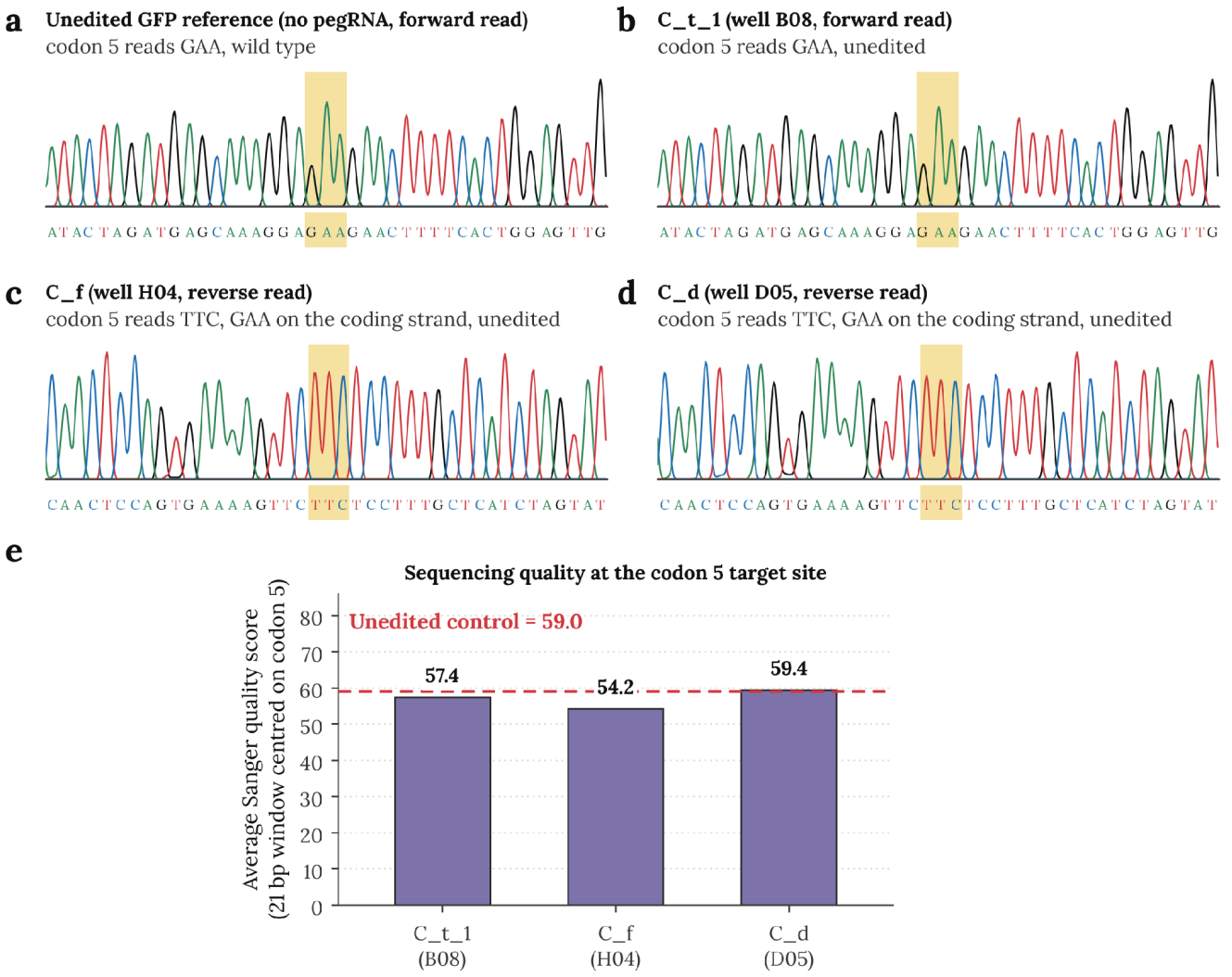


Figure 4: *Sanger sequencing at the GFP codon-5 target site.*

Note. (a) Chromatogram of the unedited GFP reference (no pegRNA control, forward read), showing wild-type codon 5 (GAA). (b) Chromatogram of sample C_t_1 (well B08, forward read), showing unedited codon 5 (GAA). (c) Chromatogram of sample C_f (well H04, reverse read), showing unedited codon 5 as TTC on the reverse strand, corresponding to GAA on the coding strand. (d) Chromatogram of sample C_d (well D05, reverse read), showing unedited codon 5 as TTC on the reverse strand. (e) Mean Sanger sequencing quality score across a 21 bp window centered on codon 5 for each sample, with the unedited control shown as a dashed line at 59.0. None of the three samples carries the intended TAG codon. All three were sequenced from pooled-colony liquid cultures rather than isolated single colonies. Created with BioRender.

3.3. *Grin2b* Expression in Ischemic Cortex

In GSE38037, *Grin2b* was modestly higher in ischemic cortex than in sham cortex (probe 10020292529, $\log_2$ fold change +0.18, equivalent to a 1.13-fold increase, $p = 0.014$, adjusted $p = 0.049$). Other NMDA receptor subunit transcripts rose to a similar or greater degree. These were *Grin1* (+0.27, adjusted $p = 0.005$), *Grin2a* (+0.34, adjusted $p = 0.009$), *Grin2c* (+0.21, adjusted $p = 0.006$), and *Grin3a* (+0.29, adjusted $p = 0.020$). *Grin2d* and *Grin3b* were unchanged (Fig. 5). In this dataset, 6,919 of 23,376 probes (29.6%) reached an adjusted p-value below 0.05. That high proportion is consistent with the low within-group variance expected of a common-reference two-color design. Statistical significance alone therefore provides limited evidence of specificity, and *Grin2b* ranked 6,850th by p-value. These data are consistent with coordinated engagement of NMDA receptor subunit expression after cerebral ischemia. They do not show selective *Grin2b* dysregulation.

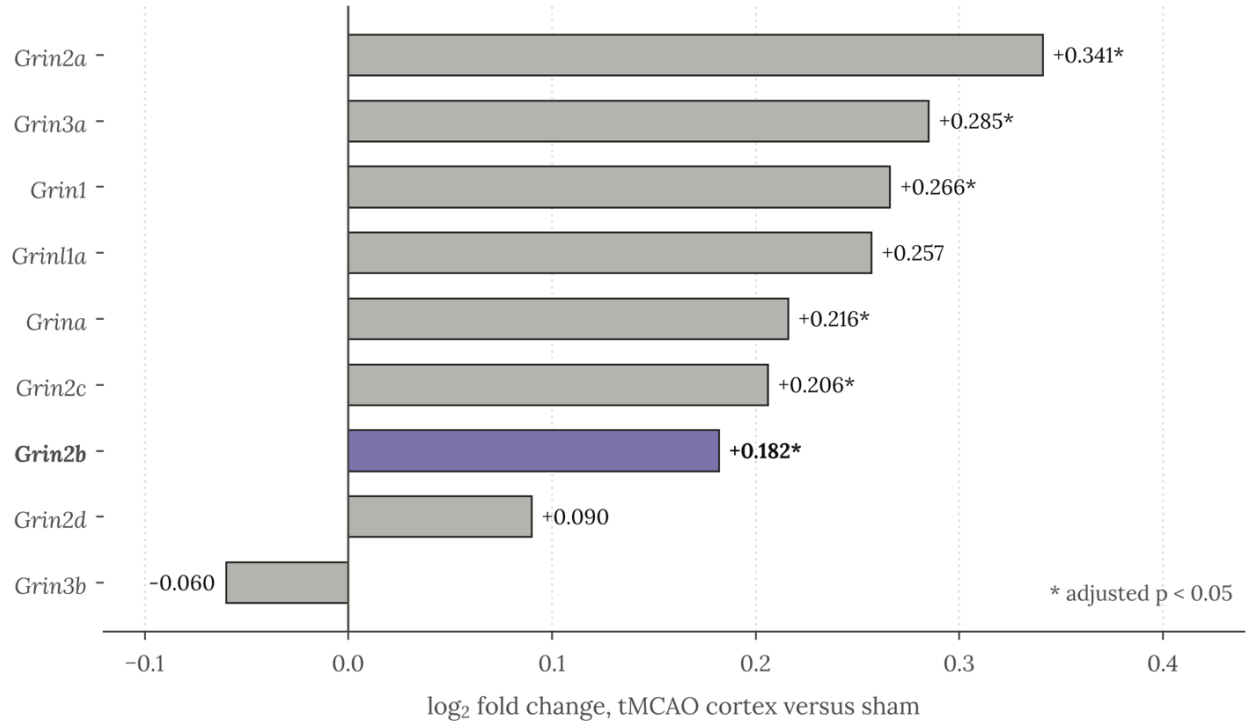


Figure 5: NMDA *receptor subunit transcript changes in rat ischemic cortex.*

Note. Differential expression of NMDA receptor subunits and subunit-associated transcripts in GSE38037 (transient middle cerebral artery occlusion versus sham, rat cerebral cortex, four samples per group), computed with GEO2R. Bars show $\log_2$ fold change. *Grin2b*, the target of the editing strategy evaluated in this study, is highlighted. Asterisks mark an adjusted p-value below 0.05 after Benjamini-Hochberg correction. *Grin2b* shows the smallest significant increase among the subunit transcripts measured.

4. Discussion

Public expression data indicate that *Grin2b* is transcriptionally higher in the ischemic cortex. This confirms that the gene is altered in the disease context that motivates this work. The effect is small, and comparable or larger increases occur in the other NMDA receptor subunits. These data therefore do not identify *GRIN2B* as uniquely dysregulated. The case for subunit-selective intervention rests instead on the divergent downstream signaling of GluN2A-containing and GluN2B-containing receptors (Wu et al., 2023; Yu et al., 2023). Transcript abundance does not address that question. The relevance of this analysis to the present study is narrower still. It establishes that the eventual target is disease-relevant, while the editing and verification questions examined here remain unresolved. The public expression analysis used a rodent model of ischemia rather than human stroke tissue, and the gene is *Grin2b* in rats rather than *GRIN2B* in humans. It is also a transcript-level observation, which does not establish that receptor-level signaling changes in proportion.

Three lines of evidence were presented above. These were growth and bulk fluorescence over time (Fig. 2), colony fluorescence on two plates (Fig. 3), and Sanger sequencing at the codon 5 target site (Fig. 4). None of them provides molecular evidence of the intended GAA-to-TAG edit. Growth was comparable across the Stop Codon construct and both plasmid-only controls. The fall in fluorescence at the final timepoint occurred in all three conditions, including the two that carry no GFP. Sequencing showed the wild-type codon in all three samples tested. Each line of evidence is taken in turn below.

Considered on its own, the difference in mean colony brightness between Experiment Set A and Sample C_f could be read as evidence of differing editing status between the two plates. However, this interpretation is complicated by colony density, which also differed markedly between the two plates: colonies in Set A were moderately spaced and individually distinguishable, while those in C_f were considerably denser and, in places, near confluent (Fig. 3a, b). Because fluorescence was quantified from manually drawn regions of interest around individual colonies, crowding on the C_f plate provides a plausible technical confound. Overlapping regions of interest, competition for nutrients or oxygen between adjacent colonies, and reduced accuracy of background correction could each lower or add noise to individual colony readings. These effects would occur independent of any underlying genotype (Chacón et al., 2018). The comparatively higher variability observed in C_f is more consistent with this kind of density-related technical variation than with a straightforward genotype-driven reduction in brightness. Because the two plates are separate platings of the same construct and were not matched for density, the present data cannot separate a genuine biological difference from plate-to-plate technical variation.

A related limitation affects Fig. 2. At the first timepoint, the Stop Codon construct already showed both higher OD_{600} (0.11) and higher fluorescence (47,344 a.u.) than either control. Control OD_{600} values were roughly 2.4 times lower (0.045) at the same time point. Because the fluorescence plotted in Fig. 2b is raw rather than normalized, part of the early difference in that panel is attributable to differences in culture density rather than to per-cell GFP expression. The normalized values in Table 1 remove most of it. By the final timepoint, fluorescence in all three conditions, including both plasmid-only controls, which carry no GFP reporter, converged to a similarly low range (3,417-3,631 a.u.) even as OD_{600} reached its highest value in every condition. This decline occurred uniformly across edited and unedited conditions alike. This pattern is more consistent with a general reduction in GFP signal as cultures approach stationary phase than with a signature specific to stop-codon-mediated loss of function. Possible causes of this general decline include reduced GFP synthesis, impaired chromophore maturation, or increased culture turbidity (El Itawi et al., 2026). Taken together, the OD confounds at the first

timepoint and the shared decline at the last limits how much Fig. 2 alone can indicate about editing status.

Sanger sequencing provided the only nucleotide-level test of editing status in this study, and none of the three samples sequenced carried the intended TAG stop codon at position 5. All three instead matched the wild-type GAA sequence of the unedited control (Fig. 4a-d). Average sequencing quality across the 21 bp window centered on codon 5 remained within five points of the control for all three samples (Fig. 4e). This indicates that these were high-confidence base calls rather than low-quality reads. It also rules out poor sequencing as an explanation for the absence of a detected edit. This null result should, however, be interpreted in light of an important methodological difference from Tong et al. (2021): that study verified editing outcomes by sequencing 24 isolated single colonies per condition, whereas the present study sequenced pooled-colony liquid cultures. A Sanger read from a pooled culture reports only the dominant sequence present across the sampled population. This approach is characteristically far less sensitive to a minority edited subpopulation than sequencing of isolated clones. An edited variant generally must approach a substantial fraction of the population, rather than being present in only a small number of cells, before it becomes visible above background in the chromatogram trace. Accordingly, the present result indicates that the wild-type sequence was the dominant outcome among the cells sampled in each pooled culture, but it does not by itself establish that no cell within either culture carried the intended edit.

The nature of the edit attempted here makes that sensitivity gap more important. Converting codon 5 from GAA to TAG requires two nucleotide changes within one codon, at the first and third positions. Tong et al. (2021) characterized a single-nucleotide substitution across a full anhydrotetracycline dose-response series, but reported no comparable series for a two-base substitution, so a directly matched efficiency figure is not available. Among the edit types they did characterize, substitution was consistently the least efficient. It reached only about 2% at the highest inducer concentration and about 1% at a moderate one. Their combinatorial category, which requires more than one simultaneous base change, reached about 20% and 13% under the same two conditions, against roughly 44% and 29% for a 1 bp deletion. If the efficiency of the present edit falls anywhere in that range, a pooled culture would contain only a small minority of edited cells, probably below the level at which a Sanger consensus would show anything but the wild-type sequence. A further possibility is that edited cells were present but lost during the outgrowth. Cells carrying a disrupted reporter were grown for up to 72 hours alongside unedited cells in the same culture, so any growth disadvantage would reduce their share of the population before sequencing. The absence of a detected edit is therefore as consistent with low editing efficiency, or with loss of edited cells during outgrowth, as with no editing at all.

Several limitations qualify these findings. Most centrally, and consistent with the detection-sensitivity argument above, editing status was assessed by Sanger sequencing of only three pooled-culture samples. This falls short of isolated single-colony sequencing at a scale comparable to Tong et al.'s 24 clones per condition. It also falls short of amplicon-based deep sequencing capable of resolving low-frequency variants within a population. Second, the dataset does not include a matched no-tetracycline induction condition for the Stop Codon construct, so a paired induced versus uninduced comparison, of the kind used by Tong et al. (2021) to establish dose-dependent editing efficiency, could not be performed. Additional confounds affect the fluorescence-based measurements discussed above. The two plates compared in Fig. 3 also differed substantially in colony density and were not matched replicates. Sample sizes were modest throughout. Six colonies were quantified per plate. Only two genuinely distinct Codon-plus-TET plate images were available in the dataset (an additional candidate pair was found to be a byte-identical duplicate file rather than an independent replicate). Only three colonies were carried forward to sequencing. No positive control was included to confirm that the fluorescence and sequencing workflows used here were capable of detecting a known successful edit. Fluorescence and sequencing data also

came from different cells, plate-grown colonies for fluorescence and pooled liquid culture for sequencing, so colony brightness cannot be directly linked to the sequenced genotype. Finally, as with any GFP-based reporter assay, loss of fluorescence is only a proxy for gene disruption in *E. coli*. Results obtained on editing efficiency in this bacterial system do not necessarily transfer directly to mammalian neurons, the eventual context for NR2B (*GRIN2B*) targeting in ischemic stroke.

Together, these results indicate that no stop codon edit was detected at codon 5 of GFP among the samples sequenced. They also indicate that the present design, pooled-culture sequencing of a small sample without a matched uninduced control, would not have reliably detected editing occurring at the low efficiencies reported previously for substitution-type edits. Distinguishing between these possibilities will require one of two approaches: isolating and sequencing single colonies at a scale comparable to Tong et al.'s (2021) screen, or applying amplicon-based deep sequencing to detect low-frequency edited variants. Either approach would also need to be paired with generating the missing no-TET control needed for a proper induced versus uninduced comparison. Testing a deletion-type edit, which showed the highest efficiency in the Tong et al. (2021) toolkit, could additionally serve as a positive system validation control. This would confirm that the cloning, culturing, and sequencing workflow used here is capable of detecting a successful edit before returning to the more technically demanding substitution required for stop codon insertion into NR2B. Establishing these methodological foundations in this bacterial reporter system is a necessary prerequisite for extending prime editing toward NR2B (*GRIN2B*) silencing in mammalian models of ischemic stroke.

5. Conclusion

This study implemented and evaluated a Sanger sequencing and colony fluorescence workflow for assessing CRISPR prime editing of a stop codon at codon 5 of GFP in *E. coli*. The sensitivity of that workflow proved to be the factor limiting what the data can show. Across three sequenced samples, no evidence of the intended GAA to TAG edit was detected, and the accompanying fluorescence measurements were subject to confounds that limit their ability to indicate editing status on their own. Because this edit type is reported elsewhere as markedly less efficient than insertion or deletion type edits, and because only pooled-culture sequencing of three samples was performed, this null result is best read as inconclusive rather than as evidence against editing having occurred. Establishing a more sensitive verification workflow, through single-colony sequencing and a matched no-TET control, is a necessary next step before extending this approach toward NR2B (*GRIN2B*) silencing in ischemic stroke.

6. References

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., Chen, P. J., Wilson, C., Newby, G. A., Raguram, A., & Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature*, 576(7785), 149–157. https://doi.org/10.1038/s41586-019-1711-4

Chacón, J. M., Möbius, W., & Harcombe, W. R. (2018). The spatial and metabolic basis of colony size variation. *The ISME Journal*, 12(3), 669–680. https://doi.org/10.1038/s41396-017-0038-0

Dammavalam, V., Lin, S., Nessa, S., Daksla, N., Stefanowski, K., Costa, A., & Bergese, S. (2024). Neuroprotection during thrombectomy for acute ischemic stroke: A review of future therapies. *International Journal of Molecular Sciences*,

25(2), 891. https://doi.org/10.3390/ijms25020891

El Itawi, H., Zoghlami, A., Charbonnier, M., González-Hernández, Y., Ghali, F., Balti, R., & Perré, P. (2026). Ability of a microfluidic droplets system to assess the dynamics of green fluorescence protein (GFP) production by *Escherichia coli* for modelling and bioprocess purposes. *New Biotechnology*, 92, 12–24. https://doi.org/10.1016/j.nbt.2026.01.006

Feigin, V. L., Brainin, M., Norrving, B., Martins, S. O., Pandian, J., Lindsay, P., Grupper, M. F., & Rautalin, I. (2025). World Stroke Organization: Global stroke fact sheet 2025. *International Journal of Stroke*, 20(2), 132–144. https://doi.org/10.1177/17474930241308142

Tong, Y., Jørgensen, T. S., Whitford, C. M., Weber, T., & Lee, S. Y. (2021). A versatile genetic engineering toolkit for *E. coli* based on CRISPR-prime editing. *Nature Communications*, 12(1), 5206. https://doi.org/10.1038/s41467-021-25541-3

Wang, I.-M., & Zhang, B. (2013). *Profiling brain cortex from tMCAO rat stroke model* (GSE38037) [Data set]. NCBI Gene Expression Omnibus. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE38037

Wu, T., Chen, S.-R., Pan, H.-L., & Luo, Y. (2023). The α2δ-1-NMDA receptor complex and its potential as a therapeutic target for ischemic stroke. *Frontiers in Neurology*, 14, 1148697. https://doi.org/10.3389/fneur.2023.1148697

Yu, S. P., Jiang, M. Q., Shim, S. S., Pourkhodadad, S., & Wei, L. (2023). Extrasynaptic NMDA receptors in acute and chronic excitotoxicity: Implications for preventive treatments of ischemic stroke and late-onset Alzheimer's disease. *Molecular Neurodegeneration*, 18(1), 43. https://doi.org/10.1186/s13024-023-00636-1

Code and Data Availability

The measurements behind Figures 2 and 3 and Tables 1 and S1 are deposited at Zenodo: https://doi.org/10.5281/zenodo.23530892. The raw Sanger traces, the GEO2R output table, and the figure scripts will be added to the same record as a later version. The gene expression dataset analyzed in Section 2.8 is publicly available from NCBI GEO under accession GSE38037.

Acknowledgments

I would like to thank my **parents**, who have always supported me in making this research project happen. I would also like to thank **Dr. Xavier Portillo** from Harvard Medical School and **Vasave Bhavesh** for their guidance and mentorship on this project, for training me in lab activities, data collection, methodologies, data analysis, and for reviewing and editing the paper. This research was conducted through the Indigo Innovator Global Lab at Life Science Incubator Singapore.

AI Use Statement

Throughout the process, Claude, an AI assistant made by Anthropic, was used as part of this manuscript preparation. Claude helped interpret the GEO2R output for the public expression dataset and helped me plan and evaluate reviewers' responses. It also helped me check for formatting errors, typos, and grammar mistakes.

Claude did not perform any experiments. All laboratory work, data collection, and measurement are mine, with Dr. Xavier Portillo and Vasave Bhavesh. Every decision about what the data mean, what to claim, and what to leave out is mine, and I read and approved every sentence before using it.

Author Biography

Azrey Auranugraha Supriadi is a student at Husnul Khotimah Islamic Boarding School. He plans to pursue a Bachelor's degree in neuroscience or informatics, with a specific interest in neurological diseases, genetics, and computation. He was awarded a bronze medal at the most prestigious national research competition in Indonesia for his neuroscience research on working memory, and has been running a free mentoring and manuscript review program for students preparing science fair projects, which has reached over 100 mentees across Indonesia. He was also awarded a full scholarship for the Indigo Innovator Global Lab at Life Science Incubator Singapore, where he evaluated the CRISPR prime editing workflow in *E. coli* systems, producing this research manuscript. Recently, he was awarded a Garuda Scholarship, a full-ride scholarship to pursue higher education in the top world universities, scholarship by the Ministry of Higher Education, Science, and Technology Indonesia.

Mentor Contribution Statement

Dr. Xavier Portillo, faculty mentor for this project, provided extensive feedback on the organization, language, and structure of the manuscript. He taught the author how to construct figures and facilitated the author's access to paywalled journal articles through his institutional access. **Vasave Bhavesh**, teaching assistant to Dr. Xavier Portillo, provided additional support throughout the project. Together, Dr. Xavier Portillo and Vasave Bhavesh worked closely with the author on constructing the plasmid, troubleshooting experimental issues, and explaining the CRISPR prime editing system step by step as part of the laboratory and mentoring activities. They also assisted extensively with collecting lab data, including OD and fluorescence intensity measurements, and tutored the author on using Fiji and Benchling to analyze the resulting datasets.

Appendix

Supplementary Table S1: Per-colony fluorescence measurements for Figure 3. Background-corrected RawIntDen, region-of-interest area in pixels, and background-corrected mean intensity for each of the six colonies measured on each plate, with plate means and the background region on each plate. Region area was recovered as RawIntDen divided by the mean gray value. Background-corrected mean intensity is the colony mean gray value minus the background mean gray value.

Plate	Colony	RawIntDen	ROI area (px)	Mean Intensity
Experiment Set A	1	25,992	340	68.8

	2	40,270	484	62.7
	3	24,782	306	78.2
	4	25,239	336	68.0
	5	19,747	320	56.8
	6	21,846	342	56.0
	Mean	**26,313**	**355**	**65.1**
	Background ROI	14,714	289	-
Sample C_f	1	10,691	154	47.5
	2	4,660	132	17.4
	3	5,650	120	32.0
	4	12,542	144	66.9
	5	3,128	96	25.4
	6	7,093	100	62.1
	Mean	**7,294**	**124**	**41.9**
	Background ROI	3,747	81	-

Supplementary Sequence S1: Complete pegRNA cassette as synthesized. The 180-nucleotide gBlock ordered for this study, 5' to 3', with the components marked.

Upstream homology arm, 35 nt
ttgacagctagctcagtcctaggtataatactagt

Protospacer, 20 nt
GAGAAATACTAGATGAGCAA

Guide RNA scaffold, 76 nt
GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC

Reverse-transcription template, 16 nt
AAGTTCCTATCCTTTG

Primer-binding site, 13 nt
CTCATCTAGTATT

Downstream homology arm, 20 nt
ttttttgaagcttgggccc