2021

Sharpening new scissors: Designing a CRISPR-Cas9 construct for inducible and interchangeable HSATII expression in Tig-1 fibroblasts

Andrew Y. Cheng , '21

Sharpening new scissors:

Designing a CRISPR-Cas9 construct for inducible and interchangeable HSATII expression in Tig-1 fibroblasts

by

Andi Y. Cheng

Advised by Dawn Carone

ABSTRACT

As research delves ever deeper into the causes and progression of cancers, complex relationships between noncoding RNAs (ncRNAs) and tumorigenesis have been revealed. One such RNA is human satellite II (HSATII). While HSATII exists across multiple human chromosomes and comprises large portions of pericentromeric heterochromatin, it is transcriptionally silent in healthy cells. However, HSATII expression is aberrantly upregulated in many tumors, suggesting an etiological or exacerbatory role for the RNA transcripts. In cancerous cells, many aspects of cellular function are dysregulated, complicating the assessment of the specific effects of HSATII RNA. To investigate the role of HSATII RNA expression in cancer progression, a constitutive expression construct was previously designed for insertion into healthy Tig-1 fibroblast cells, generating cell lines that stably express HSATII over multiple generations. This project refines stable line generation and expression behavior by redesigning the expression construct to incorporate three new functionalities: 1) leveraging CRISPR-Cas9 editing enables targeted, sequence-specific genomic integration, 2) Golden Gate cloning facilitates the insertion of different/multiple HSATII sequence variants, and 3) a Tet-on promoter system produces inducible and controlled temporal expression of HSATII. An intermediary plasmid containing RFP as a marker for successful cloning was designed to expedite and simplify Golden Gate cloning. In the future, stable lines generated with this novel construct have the potential to more accurately parallel endogenous HSATII expression in cancers and offer high sequential, spatial, and temporal resolution for interrogating the role of HSATII ncRNA in tumorigenesis.

INTRODUCTION

Sometimes considered a disease of "living too long," cancer—the unfettered growth of a body's own cells—presents a major obstacle to health and longevity in the modern world (Brown, 2015; White *et al.*, 2014). In the United States, cancers are the second leading cause of death, and incidence and mortality rates are expected to rise over time due to aging populations, resulting in economic burdens in the hundreds of billions (Siegel *et al.*, 2020, 2021; J. Xu, 2020; Yabroff *et al.*, 2011). At the cellular level, cancer cells manifest phenotypic differences compared with their healthy counterparts (Carew & Huang, 2002; Wallace, 2012). Further, DNA replication and repair are impaired in colon, ovary, endometrium, prostate, and breast cancers (Abkevich *et al.*, 2012; Boyer *et al.*,

1995; Jiricny & Nyström-Lahti, 2000). Cancer not only impacts DNA replication, however, as genomic integrity itself may be compromised. Cancer cells frequently display abnormal genomes characterized by aneuploidy, chromosomal rearrangements, and rapidly accumulating mutations (Hartwell, 1992; Yoon *et al.*, 2002). This genomic instability often precedes and may precipitate tumorigenesis, potentially via a positive feedback loop involving a defective mitotic spindle checkpoint and other cell-cycle complications (Bartek *et al.*, 1999; Hartwell, 1992; Yoon *et al.*, 2002).

An etiological factor of cancer under novel consideration is the role of noncoding RNAs (ncRNAs). First observed in the mid-1990s, the number of classified ncRNAs has exploded in past decades, and it is estimated that ncRNAs constitute 60% of total transcriptional output in human cells (Anastasiadou *et al.*, 2018). ncRNAs therefore likely provide key functionalities and have been implicated in regulatory roles in most cellular functions. Of note, they have been identified as oncogenes and tumor suppressors in every major cancer type, and expression levels of some ncRNAs have been observed to correlate with cancer drug sensitivity and response (Anastasiadou *et al.*, 2018; Corrà *et al.*, 2018; Nicolas, 2017). One tumor-associated ncRNA of interest here is human satellite II (HSATII).

HSATII is a satellite DNA, sequences which occur in long arrays of near-identical tandem repeats, form a significant component of constitutive heterochromatin, and are often enriched in pericentromeric regions of chromosomes (Altemose *et al.*, 2014; Bersani *et al.*, 2015; Younger & Rinn, 2015). Although satellite repeats comprise 15% of the human genome, their roles are still relatively unknown. While classically considered evolutionary relics and transcriptionally inert, they have been shown to undergo regulated

transcription with implications for chromosome organization and segregation, kinetochore formation, and heterochromatin status (Hall *et al.*, 2017; Nogalski *et al.*, 2019). HSATII has been identified to reside on 11 human chromosomes, comprising the major portion of pericentromeric heterochromatin on chromosomes 2, 10, and 16, as well as a 13Mb chromosome band at 1q12 (Tagarro *et al.*, 1994). While there are generally 23-26 bp consensus sequences for a single HSATII repeat, HSATII displays considerable sequence heterogeneity. HSATII can be divided into 3 overall subfamilies based on sequence similarity, within each of which there is potentially further variance (Altemose *et al.*, 2014). Thus, while HSATII is a ubiquitous satellite sequence in the human genome, there is high potential for individual variation at both the sequence and repeat levels.

Numerous studies have associated HSATII expression and dysregulation with cancer physiology. Aberrant expression of HSATII RNA has been observed in pancreatic ductal adenocarcinomas in both mice and humans. This relationship was observed at the levels of both *bona fide* tumors and precancerous intraductal papillary mucinous neoplasm lesions. Further, disease severity correlated with level of HSATII expression (Kishikawa *et al.*, 2016.; Pantano *et al.*, 2015; Ting *et al.*, 2011). The induction of HSATII RNA transcription is not limited to pancreatic cancers, as overexpression is also observed in human lung, kidney, ovarian, and prostate carcinomas (Ting *et al.*, 2011). Upon expression in cancer cells, HSATII RNA has been observed to aggregate in the nucleus *in cis* to its transcription sites into punctate foci. These aberrant focal accumulations have been observed in fibrosarcoma, breast adenocarcinoma, osteosarcoma, prostate adenocarcinoma, hepatocellular carcinoma, astrocytoma,

glioblastoma, gastrointestinal adenocarcinomas, and more cancer cell lines and tissues (Hall *et al.*, 2017). These results indicate that the association between HSATII RNA upregulation and tumorigenesis is widely relevant across tissue types with a broadly conserved, focal accumulation profile.

The consistent coincidence of HSATII RNA with cancer posits a causative or exacerbatory role of HSATII transcripts on cancer progression. While the exact relationship has yet to be clearly defined, previous studies have revealed some potential mechanistic linkages.

One model focuses on protein sequestration and relies on the previously described focal accumulations of HSATII RNA. In many cancers, master regulatory proteins—key chromatin and epigenetic modifiers—aggregate into two types of cancer-specific nuclear bodies. PRC1, a major polycomb group complex and histone ubiquitylase, accumulates into cancer-associated polycomb (CAP) bodies at demethylated 1q12 HSATII DNA (Hall *et al.*, 2017; Vidal & Starowicz, 2017). Meanwhile, HSATII RNA foci attract methyl CpG binding protein 2 (MeCP2) —an intrinsically disordered nucleic acid binding protein with both transcriptional activator and repressor roles—into cancer-associated satellite transcript (CAST) bodies (Ausió *et al.*, 2014; Hall *et al.*, 2017). The abnormal sequestration of these epigenetic regulators into CAP and CAST bodies thus disrupts chromatin state and transcriptional activity. By serving as molecular "sponges," HSATII sequences may thus contribute to the genomic instability and aberrant physiology of cancer cells (Hall *et al.*, 2017).

A second model considers HSATII transcripts as immunomodulators. Some key HSATII transcripts have been identified as immunostimulatory self-agonists, leading to

immune activation. The innate immune response in tumors may thus rely on the upregulation of HSATII expression for initiation. However, long-term activation and inflammation may detrimentally impact cell physiology, and an unsuccessful immune response with concomitant accumulation of copious HSATII RNA could thus further cancer progression (Tanne *et al.*, 2015).

While these proposals offer tantalizing possibilities for the role of HSATII in cancer, further work is necessary to interrogate and validate them. In cancers, a whole host of other defects and abnormalities are at play in addition to aberrant upregulation of HSATII. Thus, it remains unclear for which aspects of cancer phenotype HSATII expression is essential. One approach to address this issue is to express HSATII RNA in a normal, healthy cell line. As normal cells lack cancer abnormalities, any phenotypic changes upon HSATII introduction can be inferred to result from the presence of the RNA. Previous studies have applied this methodology in the study of α-sat, another major class of human satellite sequences, overexpressing the transcript via methods such as transfection and microinjection. These assays have revealed that introducing α-sat to normal cells leads to cell division defects, chromosomal instability, and aneuploidy (Chan *et al.*, 2017; Ichida *et al.*, 2018; Zhu *et al.*, 2018).

Prior work in the Carone lab has utilized this approach in the development of stable lines—normal cell lines that stably express HSATII RNA. Stable lines were generated from the Tig-1 primary (non-transformed) human fibroblast cell line using lipid mediated transfection. In this technique, plasmids containing the HSATII expression construct are encapsulated in a lipid reagent that delivers the plasmid into the nucleus. The construct is then randomly integrated into the genome (Landers *et al.*, 2021). As

HSATII sequences native to human chromosomes 7 and 10 undergo preferential expression in cancer lines, a sequence from Chr7 was selected for incorporation in the expression construct (Hall *et al.*, 2017; Landers *et al.*, 2021). In that study, HSATII construct transcription was under control of a CMV promoter, resulting in constitutive expression upon genomic integration (Landers *et al.*, 2021). Introduction of this constitutively expressing HSATII construct into Tig-1 cells was successful, as following transfection and neomycin selection, RNA fluorescence *in situ* hybridization (FISH) revealed focal accumulations of HSATII RNA in Tig-1 nuclei. These results mirror the behavior of HSATII transcripts in cancer cells. Furthermore, MeCP2 was observed to accumulate on HSATII RNA foci in a CAST body-like manner, and long-term expression of HSATII resulted in chromosomal instability and cell division defects such as lagging chromosomes, blebbing, and formation of micronuclei (Landers *et al.*, 2021). Overexpression of HSATII in normal cells thus appears to result in a cancer-like phenotype with reproducibility of these specific genetic and cellular abnormalities.

However, while the development of stable lines has been powerful and allowed for the isolation of some of the effects of HSATII RNA, questions that this expression construct design cannot answer remain. Overall, the construct has three major weaknesses: 1) random genomic integration, 2) a singularly expressed HSATII sequence variant, and 3) constitutive, high level expression.

These issues are inherent to the construct design, and thus, while it has enabled general interrogation into the presence vs. absence of HSATII RNA in an integration-site independent manner, it cannot approach some more nuanced questions. As HSATII RNA is observed to accumulate *in cis* and is endogenously found in pericentromeric regions,

the effects of expression may differ depending on the integration locus and surrounding chromatin state (Hall *et al.*, 2017; Landers *et al.*, 2021). Despite being a tandem repeat, HSATII has notable sequence diversity. These differences result in a myriad of expressed RNAs with variance at both the sequence and repeat structural levels. Changing the HSATII sequence has been observed to modify its secondary structure and biophysical properties, thus potentially impacting its cellular effects (Rubien, 2020). Finally, given the potential of low-level HSATII expression to actually benefit the immune response, the observed cellular and genetic defects of HSATII RNA expression may be dose-dependent or occur only at a certain threshold (Tanne *et al.*, 2015). This project hopes to shed light on these questions by developing a novel construct containing a Dox-inducible and Golden Gate interchangeable HSATII sequence with CRISPR-Cas9 mediated genomic insertion.

The design of this construct incorporates three key elements that each address one of the identified shortcomings with the previous stable line construct.

Firstly, this construct utilizes CRISPR-Cas9 biotechnology for genomic integration. There are two components of a CRISPR-Cas system: the nucleic acid component, CRISPR (clustered regularly interspaced short palindromic repeats), and the protein component, Cas (CRISPR associated protein). The presence of both components is requisite for activity, as transcription of the CRISPR sequence results in guide RNAs (gRNA) that complexes with a Cas endonuclease. The gRNA then targets the complex to its complementary sequence, resulting in sequence-specific cleavage at that site (Doudna & Charpentier, 2014; Jinek *et al.*, 2012). Cas9 is the most commonly used Cas variant, though others exist, and was thus selected for use in this construct (Gasiunas *et al.*, 2012;

Jinek *et al.*, 2012). Utilization of CRISPR-Cas9 activity thus prevents random genomic integration and instead allows for specific targeting to genomic loci.

Secondly, the HSATII plasmid is constructed using Golden Gate cloning. Golden Gate cloning differs from regular cloning in its use of type IIS restriction enzymes. These enzymes cleave external to their recognition sites, resulting in the excision of the recognition site from the final sequence (Engler *et al.*, 2008; Engler & Marillonnet, 2014; Weber *et al.*, 2011). By designing an intermediate plasmid containing type IIS restriction sites, the same overhangs can be added to multiple HSATII sequences. Any one of these HSATII variants can then be inserted into the plasmid via digestion and ligation, allowing for the ability to test multiple HSATII sequences that are integrated into the same genomic site.

Finally, the HSATII sequence in the construct is under a doxycycline (Dox)-inducible promoter. This promoter contains a Tetracycline (Tet)-responsive element (TRE) which binds a reverse-Tet-controlled transactivator. In the absence of Tet, the transactivator is inactivated and cannot bind the TRE, and as a result, the downstream sequence is not expressed. Addition of Tet results in a conformational change that activates the transactivator, causing it to bind the TRE and recruit RNA polymerase to the promoter. Dox is a derivative of Tet that is more frequently used due to its higher strength of activation (Krueger *et al.*, 2004). Addition of Dox should therefore induce expression of HSATII in this construct in a dose-dependent manner, allowing for more controlled transcription and temporal studies.

By combining all of these features, this novel construct will enable the generation of new stable lines that: 1) exhibit targeted, sequence-specific genomic integration, 2)

express different/multiple HSATII sequence variants, and 3) undergo inducible and controlled expression of HSATII.

MATERIALS AND METHODS

Cell Culture and Transfection

Tig-1 primary human fibroblast cells were thawed rapidly at 37°C from liquid nitrogen storage and then maintained in culture. MEM media was made with 15% fetal bovine serum (FBS) and supplemented with 1% L-Glutamine and 1% Penicillin/Streptomycin. Growth was maintained in a 5% CO_2-controlled 37°C water-jacketed incubator. Cells were fed with fresh media bidaily and split with trypsin as they reached maximum confluency.

Transfections for the introduction of CRISPR-Cas9 editing constructs were performed in T25 flasks on 50-70% confluent cells. Transfections were lipid-mediated, using FuGENE HD reagent (Promega:E2311) and following the previously established protocol for stable line generation (Landers *et al.*, 2021). The CRISPR-Cas9 editing system is encoded in two plasmids, one containing the CRISPR-Cas9 sequence and AAVS1 gRNA (AddGene#85802) and a second containing the donor sequence and HAs (AddGene#85798). 8 μg total DNA is required for transfection, so this co-transfection used 4 μg of each plasmid. 4 control conditions were used for the transfection: a G418 control that was untransfected, a FuGENE control that was exposed to FuGENE but no editing plasmids, a CRISPR control that was transfected with only 4 μg of the CRISPR plasmid, and a wild-type control that was untransfected and allowed to grow normally. Cells were exposed to transfection media for 24 hours before initiation of antibiotic

selection. All conditions except the WT control were exposed to MEM media with 15% FBS and supplemented with 1% L-Glutamine and 1% neomycin-derived geneticin (G418). Normal cell culture ensued with cells continuously grown on G418 media. Selection was defined as complete following the death of all cells in the G418 control, though transfected cells were maintained on G418 media to continue selection pressure and ensure construct expression.

An analogous transfection was concurrently performed in HeLa cells. Cell culture conditions differed only in the media composition, as HeLa cells grow optimally in 10% FBS. The transfection was performed with the same CRISPR-Cas9 plasmid and pT077 (AddGene#137879), a plasmid containing AAVS1 HAs and an independent GFP protein under Dox-inducible promoter.

Live Cell Imaging and Analysis

Following completion of selection, Dox was added at 1, 2, and 4 μg/mL in G418 media to induce H2B:GFP construct expression. G418 and Dox were not observed to interact in a cytotoxic manner even at the highest Dox concentration (Fig. S2A). Cells were exposed to G418/Dox media for 48 hours. At 24 and 48 hours, media was aspirated and replaced with sterile 1x phosphate buffered saline (PBS). The T25 flasks were then imaged on a ZEISS Axio Observer Z1 epifluorescent microscope, with a Zeiss Axiocam 702 mono camera using a 40x objective lens. Images were captured in brightfield and with excitation at 488 nm. Cells were initially scored for GFP expression (presence vs. absence) and the percent of cells expressing GFP in each condition was calculated. As signal intensity correlates with amount of expression, GFP intensity in the green channel

within each cell was quantified using FIJI (ImageJ). Cellular fluorescence signal was quantified by subtracting the cellular background (area × minimum signal intensity) from the mean integrated density. This value was then divided by the area to yield the average cellular signal intensity per micron^2.

Fluorescence intensities were statistically analyzed using GraphPad Prism (v9.1.0). As Brown-Forsythe and Bartlett's tests revealed standard deviations differed within each time point, Brown-Forsythe and Welch's ANOVA tests were run with Dunnett's T3 multiple comparisons test ($n_{condition}<50$) or Games-Howell's multiple comparisons test ($n_{condition}>50$) *post hoc* analyses to determine the effect of Dox concentration. To compare data with respect to concentrations and time points, two-way ANOVAs were run with Tukey's multiple comparisons test *post hoc* analyses. Unpaired t-tests were used for isolated comparisons regarding the effects of G418.

RFP gBlock PCR

A 1089 bp RFP gBlock was designed to contain an RFP cassette for bacterial expression with flanking BsmBI and ClaI/KpnI recognition sites (Fig. 4A, Fig. S5). The RFP cassette sequence was obtained from pAV113 (AddGene#63180) and ordered for synthesis from IDT as a dsDNA gBlock (Agmon *et al.*, 2015). Primers complementary to the 5' and 3' ends of the gBlock were ordered from IDT for amplification of the RFP gBlock (Table S1). Serial dilutions of template were tested for optimal and specific amplification of the gBlock (Fig. 4B). Following amplification via polymerase chain reaction (PCR), products were purified using DNA Clean & Concentrator®-5 (Zymo:D4014) and eluted into pre-warmed nuclease free H_2O for storage.

Verification of RFP Expression

To optimize for bright RFP expression in *Escherichia coli*, pAV113 was transformed into chemically competent bacteria. To assess for strain-dependence, both DH5α and Top10 variants were tested. Potential lineage-dependent effects were tested by transforming pAV113 from two single colonies: a bright red colony and the original stab from AddGene, which was more muted in color (Fig. 5). Plasmids were isolated from these sources using ZymoPURE™ Plasmid Miniprep (Zymo:D4210) and transformed into freshly thawed cells. Transformed bacteria were then grown on two types of LB agar supplemented with 50 μg/mL Ampicillin (Amp): one made with preformulated Difco™ LB Broth, Miller (Luria-Bertani) and the other made with peptone and Difco™ yeast extract. *E. coli* were grown overnight at 37°C and examined the following day.

RFP Plasmid Construction

The RFP cassette was inserted into the H2B:GFP-containing plasmid using restriction-ligation cloning. First, the plasmid and gBlock were separately incubated with ClaI (NEB:R0197) and KpnI (NEB:R3142) restriction enzymes overnight to ensure complete digestion. Restriction digest products were then purified using the DNA Clean & Concentrator-5 kit before a 10-minute ligation using T4 DNA ligase (NEB:M0202). T4 ligase was then heat-inactivated at 65°C for 10 minutes and stored on ice until transformation.

The ligation product was transformed into DH5α chemically competent *E. coli* cells (Thermo Fisher:18265017) and plated on LB agar supplemented with 50 μg/mL Amp. Red colonies were selected for growth in liquid culture (LB+Amp).

HSATII sequence PCR

The chr7 HSATII sequence used to generate stable lines was previously inserted into a pTargeT vector (Promega #1410) for constitutive expression (Landers *et al.*, 2021 | Fig. S6). To extract this sequence and incorporate BsmBI sites, novel primers were designed by modifying the primers used for inserting HSATII into the pTargeT plasmid. 6 bp buffer regions (for appropriate restriction enzyme interaction), BsmBI recognition sites, and corresponding sticky end overhangs were added to the existing Sat2XF and Chr7conF primers (Fig. 6A, Table S1). PCR was then performed using serial dilutions of pTargeT template to amplify the HSATII sequence. The full-length HSATII sequence was then purified using GENECLEAN® Turbo gel extraction to discard truncated sequences (MPbio:1102200).

RESULTS

To streamline the expression vector construction process, a heavy emphasis was placed on thoughtful optimization in an extensive design phase. Based on the available components and the desired functionalities for the system, a four-step workflow was devised (Fig.1). While all steps are required to construct the final plasmid containing HSATII, a key intermediate plasmid containing RFP is formed in step two. The outcomes are reported below, with some steps further broken down into subprocesses.

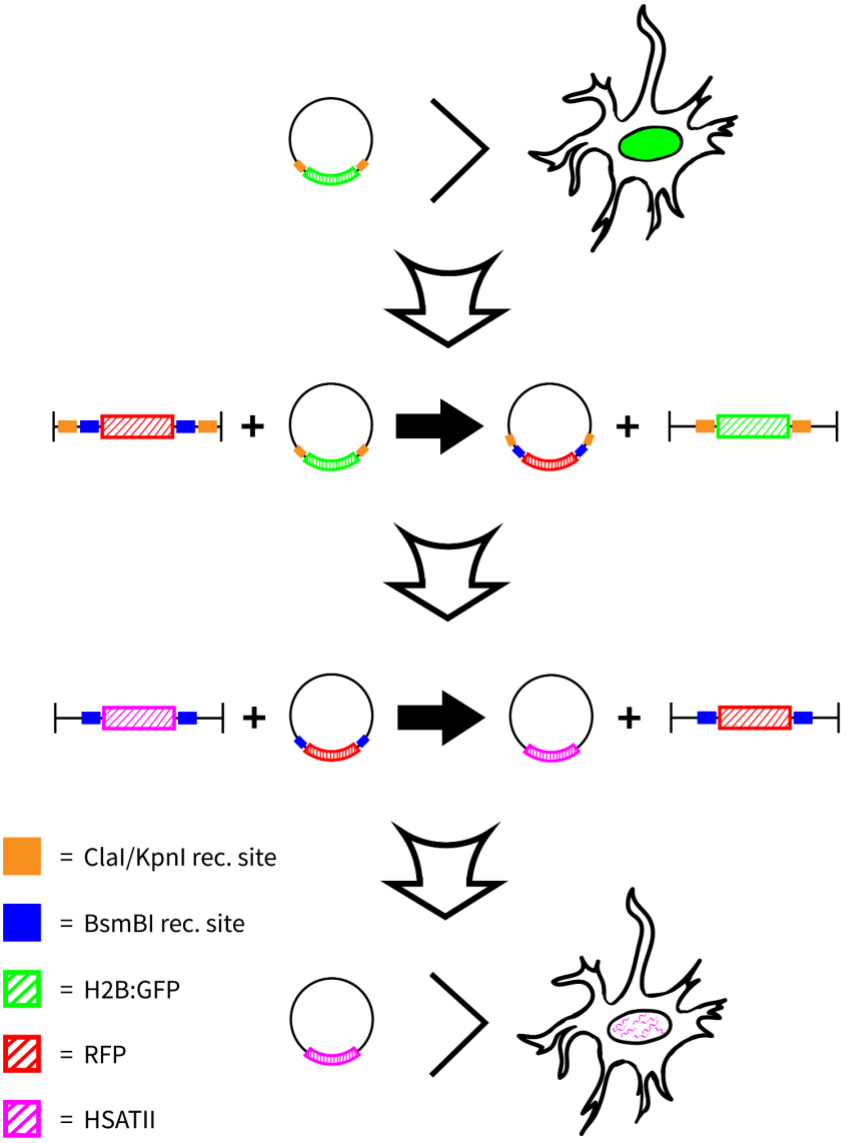


Figure 1. Overview schematic for the cloning procedure. Construction and verification was roughly broken into four steps.
1) Transfect the initial H2B:GFP plasmid into Tig-1 cells to test for CRISPR editing and expression levels.
2) Replace H2B:GFP with RFP using ClaI/KpnI and add BsmBI sites.
3) Replace RFP with HSATII using Golden Gate cloning.
4) Transfect the novel HSATII construct with CRISPR-Cas9 into cells to assess HSATII expression.

Lipid-mediated transfection of the H2B:GFP CRISPR-Cas9 construct results in inducible and dose-dependent expression

The novel construct was synthesized by modifying the pre-existing plasmid pAAVS1-Neo-CAG-M2rtTA-H2B:GFP (AddGene #85798), as it displayed some desired features. Namely, the plasmid contained a Dox-inducible promoter, a Neomycin resistance cassette, and homology arms (HAs) for the AAVS1 (adeno-associated virus site 1) genomic locus (Fig. 2A). This plasmid has been observed to function properly as a donor vector for the insertion of H2B:GFP, a histone 2B and green fluorescent protein fusion, into the genomes of glioblastoma cells in a CRISPR-Cas9 mediated manner (Tejero *et al.*, 2019). To perform this editing, two plasmids were required: the H2B:GFP donor plasmid and a CRISPR-Cas9 plasmid encoding the Cas9 protein and a single-guide RNA (sgRNA) targeting the endonuclease to the AAVS1 locus.

Although Tejero *et al.*'s study demonstrated the efficacy of this CRISPR-Cas9 system, additional validation was necessary to ensure CRISPR-Cas9-mediated insertion in Tig-1 fibroblasts due to several key differences. One such difference involves the targeted cell line. While Tejero *et al.* observed successful editing in cancerous glioblastoma cells, this project hopes to stably insert into Tig-1 cells, a primary cell line that is considerably harder to edit (Han *et al.*, 2015; Li *et al.*, 2017; Søndergaard *et al.*, 2020). Another difference lies in the transfection method. Tejero *et al.* co-transfected the two plasmids using the Neon electroporation system, whereas this project intends to use lipid-mediated transfection due to its previously observed success in the Carone lab for stable line generation (Landers *et al.*, 2021; Tejero *et al.*, 2019). To confirm CRISPR-Cas9 editing, a lipid-mediated co-transfection of the H2B:GFP donor and CRISPR-Cas9 targeting plasmids was thus performed for Tig-1 fibroblasts (Fig. 2).

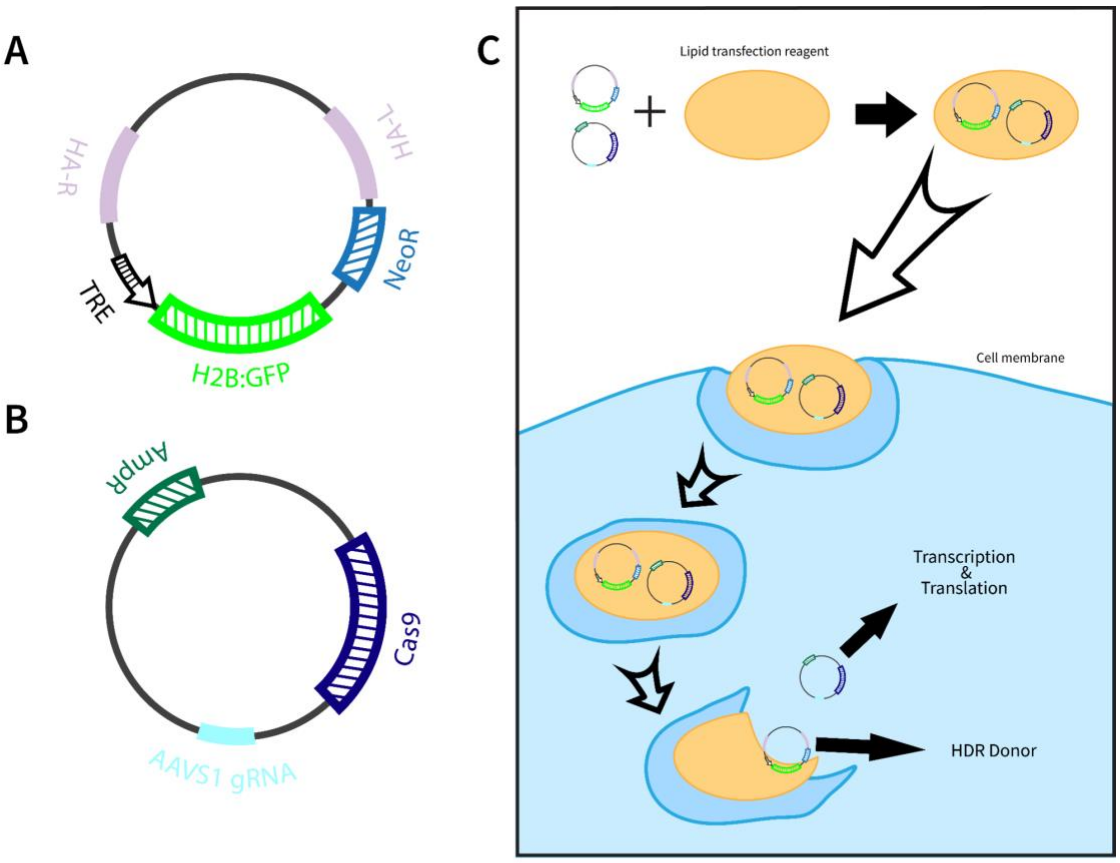


Figure 2. Co-transfection methodology for CRISPR-Cas9 editing.
A) The structure of the H2B:GFP plasmid with HAs, neomycin resistance cassette, Dox-inducible promoter, and H2B:GFP coding sequence highlighted. B) The structure of the CRISPR-Cas9 plasmid with ampicillin resistance, AAVS1 gRNA, and Cas9 coding sequence highlighted. C) The mechanism of lipid mediated transfection. Construct plasmids are engulfed by the lipid reagent (FuGENE) and uptaken into the cell by endocytosis. Upon release, Cas9 and gRNAs are synthesized and create DSBs. DSBs are then repaired via HDR using the H2B:GFP donor.

Following transfection, selection was performed with G418. G418 is cytotoxic to normal human cells, so continuous exposure kills Tig-1 cells that have not been successfully edited with the donor sequence containing a neomycin resistance cassette (Fig 2A). Cells were then treated with 1 µg/mL, 2 µg/mL, or 4 µg/mL of Dox to test for promoter activity. Imaging was performed with live-cell fluorescence microscopy 24h and 48h post-treatment initiation to assess GFP production. Cells were then scored for presence vs. absence of GFP to determine the proportion of GFP-expressing cells. There

was very little difference in the proportion of GFP+ cells within each Dox concentration between 24h and 48h. In both time points, the proportion of GFP+ cells increased from 1.0 to 2.0 μg/mL of Dox but not between 2.0 and 4.0 μg/mL (Fig. 3B). Fluorescence signal, a proxy for expression, was statistically analyzed. While the mean signal intensity did increase from 24h to 48h in all concentrations, the difference was not found to be significant (Fig. 3C). However, within each time point, Brown-Forsythe and Welch's one-way ANOVAs revealed a significant effect of Dox concentration (24h: $F^*(3, 25.8)=13.66$, $p<0.0001$; $W(3, 13.85)=36.61$, $p<0.0001$; 48h: $F^*(3, 8.46)=32.26$, $p<0.0001$; $W(3, 4.77)=270.2$, $p<0.0001$). Dunnett's T3 *post hoc* analyses revealed that at 24h, the increases in fluorescence intensity from 1.0 to 2.0 μg/mL and from 1.0 to 4.0 μg/mL of Dox were significant ($p=0.0285$; $p=0.0008$), though the increase from 2.0 to 4.0 μg/mL was not significant ($p=0.1955$). The signal intensities at both 2.0 and 4.0 μg/mL Dox were significantly higher than the no Dox treatment ($p<0.0001$, $p<0.0001$). The difference in intensities between 1.0 and the none condition approached significance ($p=0.07$). At 48h, the signal at 4.0 μg/mL was significantly higher than at 1.0 μg/mL ($p=0.0036$), and the increases from 1.0 to 2.0 and from 2.0 to 4.0 μg/mL Dox both approached significance ($p=0.069$, $p=0.059$). All treatment conditions displayed significantly higher signal than the no Dox treatment ($p=0.0033$, $p=0.012$, $p=0.0006$).

Transfection, selection, induction, imaging, and analysis were similarly performed with HeLa cells. Unlike Tig-1 fibroblasts, HeLa cells were transfected with pT077, which contains an unconjugated GFP—rather than H2B:GFP—under a Dox-inducible promoter. Additionally, induction of expression was performed with only 1.0 and 2.0 μg/mL Dox. The proportion of GFP+ cells increased with an increase in Dox concentration at both

24h and 48h. Furthermore, at both 1.0 and 2.0 μg/mL Dox, the proportion of GFP+ cells doubled between 24h and 48h (Fig. S1A, B). Signal quantification and analysis yielded similar results. A two-way ANOVA considering Dox concentration and time as the independent factors determined that there was a significant effect of time ($F(1, 1826)=2278$, $p<0.0001$).Tukey's *post hoc* test revealed that the increase in signal from 24h to 48h at both 1.0 and 2.0 μg/mL Dox were significant ($p<0.0001$; $p<0.0001$ | Fig. S1C). Further, Brown-Forsythe and Welch's ANOVAs demonstrated that at both 24h and 48h, there is a significant effect of Dox concentration (24h: $F^*(2, 366.4)=570.8$, $p<0.0001$; $W(2, 248.4)=706.4$, $p<0.0001$; 48h: $F^*(2, 896.8)=669.6$, $p<0.0001$; $W(2, 514.3)=12370$, $p<0.0001$). Games-Howell's *post hoc* tests revealed that at 24h, increasing Dox from none to 1.0, from 1.0 to 2.0, and from none to 2.0 all resulted in a significant increase in GFP expression ($p<0.0001$; $p<0.0001$; $p<0.0001$ | Fig. S1D). Similarly, at 48h, all increases in Dox concentration resulted in significant increases in signal ($p<0.0001$; $p<0.0001$; $p<0.0001$ | Fig. S1D).

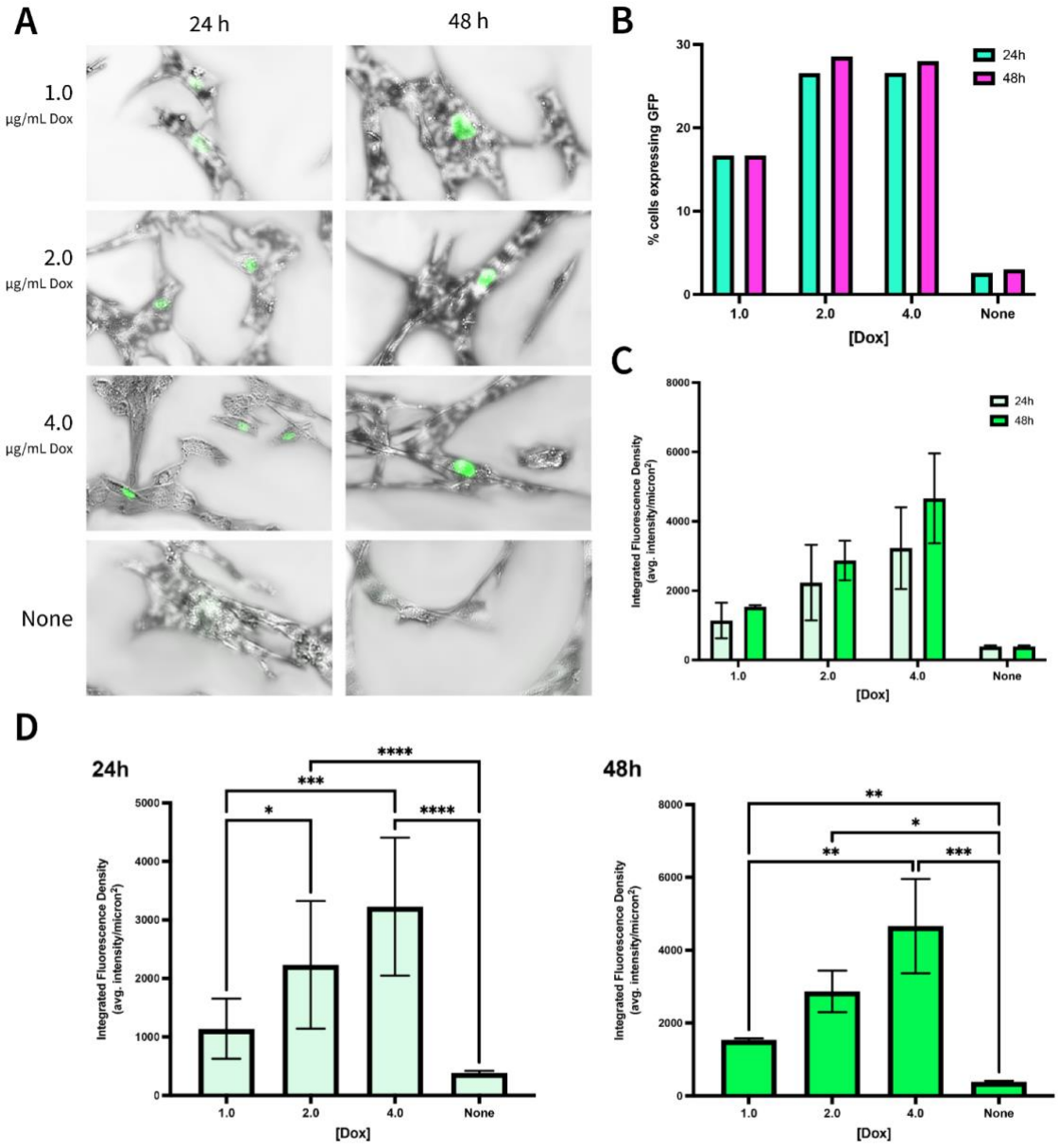


Figure 3. Verification of H2B:GFP expression in Tig-1 fibroblasts.
Following transfection and G418 selection, expression was induced with exposure to varying Dox concentrations for up to 48 h. A) Representative images of cells exposed to 1.0, 2.0, and 4.0 µg/mL of Dox at 24 and 48 h. B) The percentage of cells expressing H2B:GFP for all concentrations and timepoints. The proportion was not observed to increase over time. Increasing Dox concentration from 1.0 to 2.0 µg/mL, but not from 2.0 to 4.0 µg/mL, increased the proportion of expressing cells. C) Average signal intensities per micron² in expressing cells. Induction of expression did not change over time, as intensities within each concentration were not significantly different. D) A significant effect of Dox concentration on expression was observed within each timepoint.
(*) p<0.05, (**) p<0.01, (***) p<0.001, (****) p<0.0001

Construction of RFP gBlock

Prior to exchanging H2B:GFP for RFP, the RFP gene block (gBlock) was amplified (Fig. 1). The gBlock contains an RFP construct under *lac* promoter for expression in *E. coli* and was initially ordered from AddGene (Fig. S5). The RFP sequence used has been observed to be expressed in *E. coli*, turning the colonies red to the naked eye (Agmon *et al.*, 2015). Primers were designed to align with the ends of the gBlock to produce full-length copies by PCR. Primer sequences spanned 8 bp buffer regions that enabled enzymatic activity, restriction enzyme recognition sites, and the outer edges of the RFP (Fig. 4A). Multiple gBlock template concentrations were tested to determine an appropriate amount for gBlock amplification. At high template concentrations, a great deal of nonspecific product formed as evidenced by multiple bands in the gel. In this reaction, the ideal dilution was found to be 1:10,000,000 as at this dilution, only a single band was observed at 1.1 kb, the expected length of the RFP gBlock (Fig. 4B).

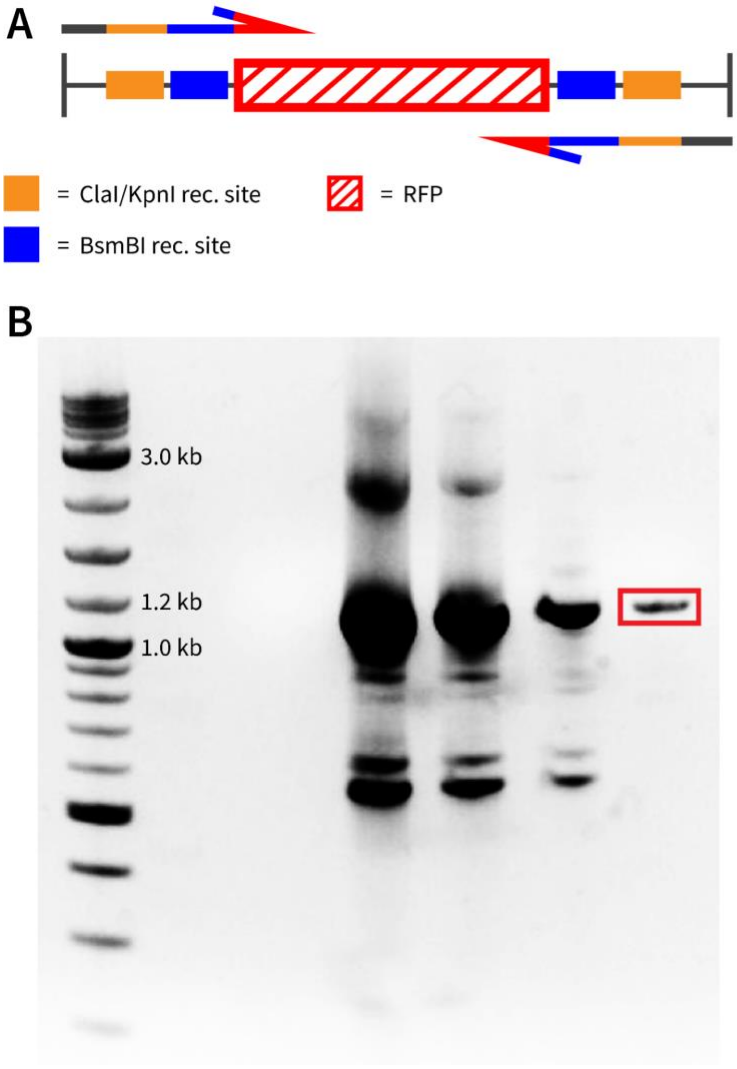


Figure 4. RFP gBlock amplification. A) A schematic representation of the RFP gBlock with key enzymatic recognition sites highlighted. The gBlock was PCR amplified using primers that overlapped at least a portion of all key features. B) Multiple concentrations of gBlock template were tested to determine the optimal dilution factor for strong and specific amplification. Highest performance was observed at a 1:10,000,000 dilution (initial concentration = 10 ng/µL). From left to right, observed bands are: 2-Log ladder, 1:10000, 1:100000, 1:1000000, 1:10000000.

Bacterial RFP expression is strain and media-dependent

To optimize expression of the RFP cassette, pAV113, the original RFP plasmid, was isolated from single colonies and transformed into DH5α and Top10 chemically competent cells. While the origin colonies displayed differing red intensities, the plasmid source (an original stab colony vs. a bright red colony) did not change RFP expression in re-transformed bacteria (Fig. 5). Choice of chemically competent bacterial strain did, however, change RFP expression. Regardless of media, colonies were brighter in DH5α cells than in Top10 cells (Fig. 5). LB agar composition also impacted RFP expression.

Growth on preformulated Difco™ LB agar resulted in lower RFP production than growth on LB agar made with peptone and Difco™ yeast extract (Fig. 5). Overall, optimal RFP expression was observed in DH5α cells grown on LB agar containing peptone and Difco™ yeast extract. While DH5α colonies were still slightly red, Top10 cells grown on Difco™ LB agar did not express RFP and were completely colorless (Fig. 5).

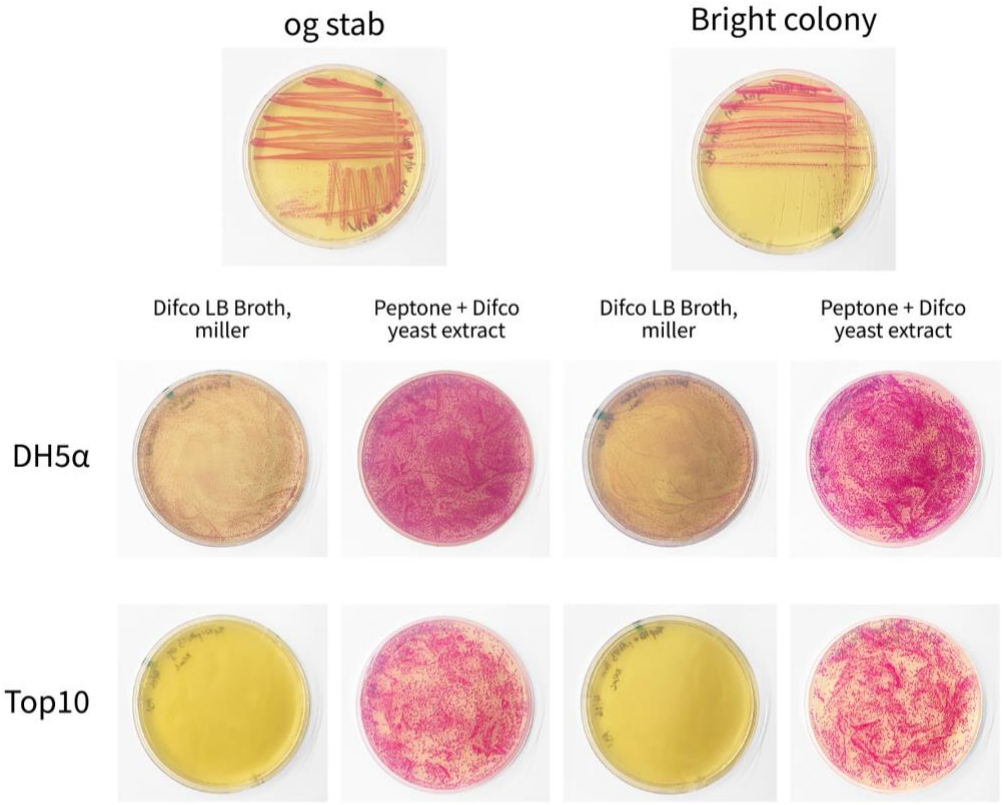


Figure 5. Assay for variation of RFP expression in *E. coli*.
The dependence of RFP expression on media, bacterial strain, and plasmid origin were tested. RFP was most highly expressed by DH5α cells grown on peptone LB agar. Top10 cells expressed lower amounts of RFP across the board. RFP expression was not impacted by plasmid source.

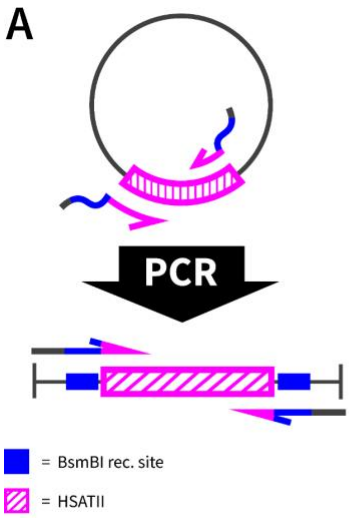
Insertion of RFP into the H2B:GFP plasmid

Following preparation of sufficient RFP gBlock, the sequence was added to the plasmid via restriction enzyme cloning. The gBlock was designed to have two flanking regular restriction enzyme sites, one for ClaI and one for KpnI (Fig. 4A, Fig. S5). Following restriction digest with ClaI and KpnI and ligation with T4 DNA ligase, the resultant plasmids were transformed into DH5α chemically competent *E. coli* cells. Cells were plated on LB Agar made with peptone and Difco™ yeast extract to induce RFP expression.

Synthesis of HSATII gBlock

The expressed RFP sequence in the plasmid was then exchanged for an HSATII sequence. The selected HSATII sequence for this first-round test of CRISPR-mediated insertion was the same Chr7 HSATII sequence expressed in stable lines (Landers *et al.*, 2021). The Chr7 HSATII sequence was selected as it has been observed to be successfully expressed and to impact cell phenotype, giving a point of comparison for this new targeted and inducible construct (Landers *et al.*, 2021). The HSATII sequence was located on a pTargeT vector that was used for stable line generation, and as such, was amplified off of the plasmid (Fig. 6A). Primers based on the original primers that generated this sequence were designed to amplify the HSATII repeat and to add BsmBI (a IIS restriction enzyme) recognition sites (Fig. 6A). PCR amplification resulted in successful generation of the full-length product. However, due to the repetitive nature of tandem repeat sequences, multiple shorter products were also formed, leading to a

"satellite ladder" in gel electrophoresis (Fig. 6B). Gel extraction was performed to isolate the correct, full-length 382 bp product for downstream incorporation (Fig. S6).



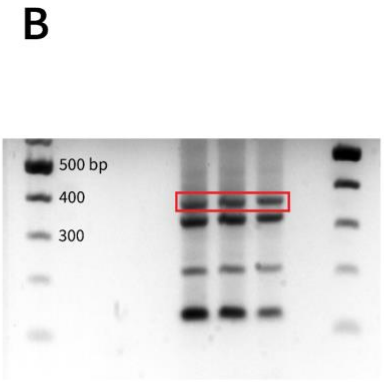


Figure 6. Chr7 HSATII sequence amplification and addition of BsmBI recognition sites. A) Primers were designed with inner portions complementary to the chr7 HSATII sequence and outer portions containing BsmBI recognition sites. The chr7 HSATII sequence was extrapolated from the pTargeT vector into a linearized sequence with flanking BsmBI sites. B) PCR resulted in a "satellite ladder" containing multiple truncated sequences. The full-length sequence (corresponding band boxed in red) was isolated using gel extraction.

Insertion of HSATII into the RFP Plasmid (in progress)

With the production of the RFP plasmid and the HSATII gBlock, the final plasmid containing HSATII can be synthesized via Golden Gate cloning (Fig. 1). Both the plasmid and gBlock were designed to include BsmBI recognition sites with corresponding sticky end overhangs. Incubation with BsmBI thus leads to cleavage external to the recognition sites, leaving distinct sticky ends at each side of the gBlock/plasmid (Fig. 7A, 7B). As overhangs are distinct, directional integration of the HSATII sequence occurs to specifically orient the HSATII sequence to the Dox-inducible promoter (Fig 7C). Following ligation, the final plasmid will be transformed into DH5α cells to screen for successful editing. As editing removes the RFP construct, non-edited colonies should appear red while successfully edited colonies should appear white.

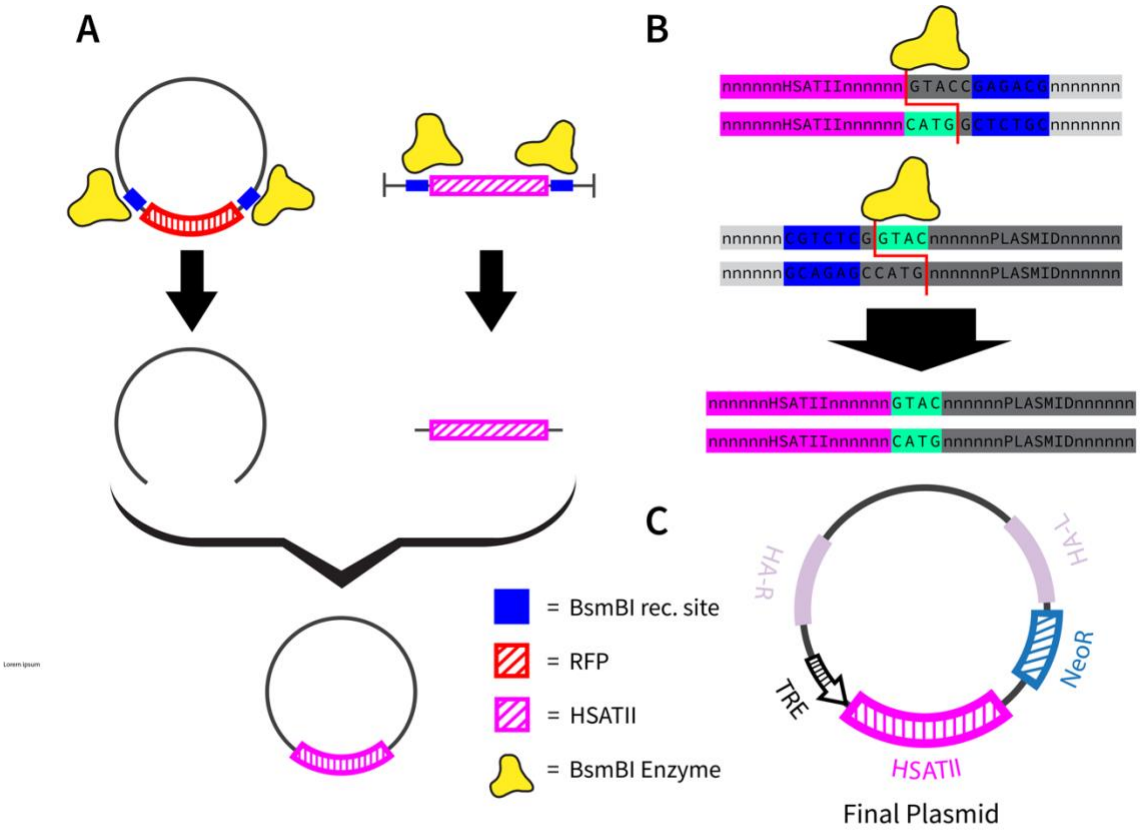


Figure 7. Construction of the final HSATII plasmid using Golden Gate cloning.
A) A single-pot reaction containing BsmBI and T4 ligase is used to synthesize the HSATII plasmid. B) Type IIS restriction enzymes cleave external to their recognition sites. BsmBI cleavage results in a sticky end overhang. With appropriate recognition site direction, cleavage and ligation will result in the excision of the recognition site, preventing the reverse reaction. C) The structure of the final HSATII plasmid with key HAs, neomycin resistance cassette, Dox-inducible promoter, and HSATII sequence highlighted.

DISCUSSION

Building on previous successes with the introduction of HSATII expression in Tig-1 fibroblasts, this project develops a new method for stable line generation. This new construct extends the malleability of HSATII expression in stable line cells in three key ways: enabling targeted genomic integration, exhibiting modulable and inducible HSATII expression, and facilitating the expression of different HSATII sequence variants.

CRISPR-Cas9: accurate and site-specific cleavage

To achieve targeted genomic insertion, this construct relies on CRISPR-Cas9 targeted integration. As molecular biological questions have become increasingly precise, the advent of CRISPR-Cas biotechnology has revolutionized genomic manipulations (Fellmann *et al.*, 2017). First discovered in 1987 in *E. coli*, the CRISPR-Cas system is prevalent in microbial genomes, functioning as a microbial mirror of adaptive immunity to combat viral infections (Horvath & Barrangou, 2010; Ishino *et al.*, 1987; Jinek *et al.*, 2012; Mojica *et al.*, 2005). The CRISPR-Cas immunological response is composed of three distinct stages: an adaptive phase, an expression phase, and an interference phase. Viral challenge in the adaptive phase results in acquisition of foreign genetic material that is conjugated with Cas proteins in the expression phase. Upon reexposure to the virus, the Cas complex specifically degrades the foreign material in the interference phase (Horvath & Barrangou, 2010; Jinek *et al.*, 2012; Marraffini & Sontheimer, 2010; Mojica *et al.*, 2005).

While CRISPR-Cas systems have been known of for decades, their use as a biotechnology tool is relatively recent. Nobel prize winners Doudna and Charpentier

pioneered a path forward by simplifying the CRISPR-Cas system for genome editing. Endogenously, CRISPR-Cas systems have many RNA components that must coordinate for appropriate endonuclease activity (Jinek *et al.*, 2012). However, these RNAs can be linked in a single, chimeric sequence: a single guide RNA (sgRNA) (Cong *et al.*, 2013; Doudna & Charpentier, 2014; Jinek *et al.*, 2012; Mali *et al.*, 2013). As sgRNAs are easily customizable to target genomic regions, so long as the proper protospacer adjacent motif (PAM) is present, a CRISPR-Cas construct can be easily modified to introduce double-strand breaks (DSBs) at specific target loci (Jinek *et al.*, 2012).

For this initial HSATII CRISPR-Cas construct, sgRNAs specific to the human AAVS1 locus were used due to the permissible characteristics of the site (Tejero *et al.*, 2019). While no *bona fide* safe harbor locus, a genomic site that supports predictable transgene expression while minimizing disruption of the host genome, has been identified in humans, the AAVS1 locus is one of the three top contenders (Dubois *et al.*, 2018; Papapetrou & Schambach, 2016; Pellenz *et al.*, 2019). As the sgRNAs and HAs were not modified from Tejero *et al.*, targeting of the construct to the AAVS1 locus should not be impacted. In their work. Tejero *et al.* confirmed proper integration of the H2B:GFP transgene into the AAVS1 locus using genomic PCR (Tejero *et al.*, 2019). While other insertion sites throughout the genome will be of interest in downstream questions, the benefits of stable, non-interrupted HSATII expression for a first-pass construct underlies our decision.

Genomic editing using HDR

In this construct, Cas9 was selected as the Cas endonuclease due to its ubiquity in the published literature. Further, Cas9 has been successfully used to edit primary human and porcine fibroblasts using homology directed repair (HDR) to insert novel genetic material (Li *et al.*, 2017, 2018; Lin *et al.*, 2014; Xie *et al.*, 2017).

HDR is similarly essential for our insertion of HSATII. Once a DSB arises in the genome, cells can resolve it in one of two ways: HDR or non-homologous end joining (NHEJ) (He *et al.*, 2016; Lieber, 2010). While NHEJ occurs more frequently and can repair DSBs, it generally results in the formation of insertion-deletion (indel) mutations rather than the insertion of a novel genetic sequence (He *et al.*, 2016; Lieber, 2010; Liu *et al.*, 2019; Mali *et al.*, 2013; Zhang *et al.*, 2017). Thus, despite its low rate of occurrence, HDR remains the optimal method for genomic insertions. As its name suggests, HDR depends on homology, specifically, homologous sequences. Endogenously in cells, HDR primarily occurs during the S or G2 phases of the cell cycle, as these are the times when a homologous template is present to guide DSB repair (Lieber, 2010; Liu *et al.*, 2019). In biotechnology, this repair mechanism is leveraged with the addition of homology arms (HAs) to either end of the desired insert sequence. HAs are homologous to the genomic regions upstream and downstream of the induced DSB and the internal edge of each arm should be within 10 bp of the DSB site (Liu *et al.*, 2019). By adding these regions of homology, HDR can repair the DSB using the supplied sequence as a template, resulting in the incorporation of the insert into the repaired genomic sequence. HA length can vary depending on the length of the insert, but generally is between 50-100% insert size, though the proportion may be smaller for very large insertions (Liu *et al.*, 2019; Zhang *et*

al., 2017). In this construct, the HAs are ~800 bp while the insert is 6.2 kb, aligning with the expectations of proportionally shorter HAs for large insertions.

Lipid-mediated transfection enables accurate and high-efficiency CRISPR-Cas9 editing

Following co-transfection of the pAAVS1-H2B:GFP donor and the CRISPR-Cas9 + gRNA plasmids, the H2B:GFP sequence was likely integrated into Tig-1 genomes (Fig. 2B). Dox-induced H2B:GFP expression was observable from P25 to senescence at P31, indicating that the sequence was genomically incorporated. Tejero *et al.* confirmed on-target genomic integration into the AAVS1 locus using genomic PCR to amplify the region of insertion (Tejero *et al.*, 2019). Given that their work was also performed in human cells, the genomic PCR was not initially performed in this study. However, future work will verify proper insertion using the same primer sets as Tejero *et al.* with subsequent PCR product sequencing to verify the appropriate junction of the AAVS1 locus and the H2B:GFP construct.

H2B:GFP expression in Tig-1 fibroblasts mimicked expression in the glioblastoma model used by Tejero *et al.* In their study, Tejero *et al* devised the H2B:GFP fusion protein as a marker for cell division. H2B:GFP is a fusion of the histone 2B peptide and a functional GFP. As histone proteins are only incorporated into a cell's genetic material when it is actively dividing and new nucleosomes are being added to replicated DNA, GFP fused to histone 2B accumulates only in actively dividing cells (Tejero *et al.*, 2019). In Tig-1 cells, H2B:GFP expression was similarly observed in either actively dividing or recently divided cells. Many cells with GFP signal were observed to contain dividing nuclei, indicating that the cells had just finished telophase but not

cytokinesis (Fig. S3). GFP-expressing cells tended to accumulate in clusters, perhaps indicating regions of active division. As the rate of cell division likely does not change, a roughly consistent number of cells should be dividing at any point. Further, the amount of chromatin and nucleosomes synthesized should be the same for all cells. The observations that both the proportions of H2B:GFP expressing cells and the signal strengths did not vary over time align with these expectations (Fig. 3B, 3C). Since accumulation of GFP signal was limited to dividing cells, the values for proportion of GFP+ cells are likely an underestimate for the true proportion of successfully edited cells that could have expressed the transgene construct (Fig. 3B). This underestimate is especially likely given that most cells did not divide. The low rate of division may have been due to the low confluency of cells, as primary fibroblasts rely on interactions with neighboring cells for optimal growth, compounded with G418 stress.

Although these results indicated that H2B:GFP was genomically integrated and functioning properly, they did not lend themselves well to estimate transfection efficiency. The simultaneous transfections performed with pT077 provided a more accurate representation. As with the H2B:GFP plasmid, the pT077 plasmid contains a Dox-inducible promoter and HAs for the AAVS1 locus. The only difference is that the expressed protein is an independent GFP (not fused to H2B). Thus, unlike H2B:GFP, GFP could accumulate throughout the cell and throughout the cell cycle. In line with these expectations, GFP was not nuclear-localized, and signal significantly increased four to five-fold from 24h to 48h (Fig. S1A, C). As any expressed GFP could accumulate, the proportion of GFP+ cells following Dox induction closely related to the transfection and editing efficiency. The proportion of GFP+ cells increased with time and with increasing

Dox concentration. The highest proportion of over 60% expressing was observed at 48h with 2.0 μg/mL Dox (Fig. S1B). Therefore, the efficiency of functional construct insertion with lipid-mediated transfection of CRISPR-Cas9 is at least 60%, a high rate of success comparable to other transfection methods (Xu *et al.*, 2018).

The Dox-inducible promoter displays dose-dependent, rheostat-like behavior

Expression of the H2B:GFP sequence was induced with Dox treatment, as the sequence is under a Dox-inducible promoter (Fig. 2A). Dox is a Tet derivative with a substituent hydroxyl group shifted from one 6C ring to a benzene ring, and it is commonly used in place of Tet due to its more potent effector ability (Krueger *et al.*, 2004). While microbial genetic regulation commonly features inducible sequences (as in operon control), these mechanisms do not normally function in eukaryotic cells. A Tet-controlled transactivator (tTA) was developed by fusing an *E. coli* Tet repressor with a *Herpes simplex* virion protein 16 activating domain to enable inducible expression in human cells. This initial system is known as a "Tet-off" design, as expression of the sequence from a minimal promoter sequence containing a Tet-response element (TRE) is normally active, with Dox treatment leading to unbinding of the tTa and expression termination (M. Gossen & Bujard, 1992). Subsequent mutagenesis assays led to development of a reverse Tet repressor, which when incorporated into the fusion protein resulted in a reverse Tet-controlled transactivator (rtTA). This new system is known as a "Tet-on" design. rtTA reverses the phenotype of the tTA, as the rtTA is not normally bound to the TRE. Dox treatment leads to rtTA binding and sequence expression (Gossen *et al.*, 1995). As we are interested in the specific expression of HSATII, a Tet-on model

was selected for this construct to reduce the duration of Dox exposure due to concerns about toxicity and other side effects (Gossen *et al.*, 1995; Krueger *et al.*, 2004).

In addition to simple binary-switch induction of HSATII expression, the ability to modulate expression level to vary the amount of transcript present can answer more detailed questions. As Dox-inducible promoters have been observed to display dose-dependent expression and rheostat-like behavior, Dox-dependent behavior of the H2B:GFP construct was assessed *in vitro* (Loew *et al.*, 2010; Rossi *et al.*, 2000; Urlinger *et al.*, 2000). Increasing Dox concentration was observed to positively correlate with the proportion of H2B:GFP expressing cells up to a maximum (Fig. 3B). At both 24h and 48h, increasing the concentration of Dox increased the signal intensity (Fig. 3D). While this observed increase in H2B:GFP may seem counterintuitive given that a constant amount of H2B should be incorporated into new nucleosomes, it can be explained by ratios. In edited cells, there were two versions of H2B: the endogenous variant and the introduced H2B:GFP. While the endogenous H2B would be expressed at the same levels during DNA replication, the novel H2B:GFP was under a Dox-inducible promoter. If there were a dose-dependent effect, then an increase in Dox would have led to an increase in H2B:GFP production. As the ratio of H2B:GFP to normal H2B increased, it was more likely that H2B:GFP was incorporated into nucleosomes, resulting in stronger signal. The positive relationship between Dox concentration and signal intensity thus indicated that expression of H2B:GFP was dose-dependent.

The dose-dependent nature of the Dox-inducible promoter was further supported by pT077 transfections. At both 24h and 48h, a significant effect of concentration on signal intensity was observed such that an increase in concentration significantly

increased the amount of GFP expression (Fig. S1D). Dose-dependent expression was thus observed for both H2B:GFP and unconjugated GFP.

These results suggest that when H2B:GFP is eventually exchanged for HSATII, a dose-dependent effect of Dox on HSATII expression should be observed, providing preliminary reassurement for modulable expression of HSATII.

G418 and leaky expression

A perennial issue with Tet-controlled promoter systems is the possibility for "leaky" expression, or expression that occurs either when the inducer is absent (Tet-on) or the repressor is present (Tet-off). This issue is especially prevalent in Tet-on systems (Meyer-Ficca *et al.*, 2004, p.; Pham *et al.*, 2008; Sun *et al.*, 2007). Optimally, this background expression should be reduced as much as possible to prevent false positives and to reduce noise. To test for any unintended interactions or effects of G418, some H2B:GFP transfected Tig-1 cells were exposed to 4.0 μg/mL Dox + G418 and none + G418. At even the highest levels of Dox, G418 was not observed to negatively interact with Dox and reduce the induction of expression (Fig. S2A). Cells also grew similarly in 4.0 μg/mL Dox + G418 and 4.0 μg/mL Dox conditions and appeared normal. However, G418 was observed to influence leaky expression. Exposure to G418 significantly reduced background H2B:GFP expression (Fig. S2B). This reduction in basal expression level may have resulted from G418 selection stress or potential chemical interactions with the rtTA or other transcriptional regulators that impaired RNA polymerase binding. Given the lack of negative interactions and the positive reduction of background, cells should remain exposed to G418 even during Dox induction.

Golden Gate cloning provides an easy and efficient method for HSATII sequence exchange

Golden Gate cloning methods were leveraged to enable the potential expression of multiple HSATII sequence variants. First designed in 1996, Golden Gate cloning differs from normal restriction-ligation cloning in its reliance on the class IIS restriction enzyme subset (Engler *et al.*, 2008; Lee *et al.*, 1996; Padgett & Sorge, 1996). Unlike other restriction enzymes, IIS endonucleases interact with two discrete sites on double-stranded DNA: a 4-7 bp recognition site and a cleavage site 1-20 bp away (Szybalski *et al.*, 1991 | Fig. 7). With proper orientation of the recognition sequence, the use of IIS restriction enzymes in cloning enables the excision of the recognition sites from the final product (Fig. 7B). Therefore, the Golden Gate cloning workflow involves only a single reaction containing the IIS enzyme and a T4 ligase as the product cannot be reverted (Engler *et al.*, 2008; Engler & Marillonnet, 2014, p. 201; Weber *et al.*, 2011). The ease of Golden Gate cloning lends itself well to the goal of testing multiple HSATII sequence variants: so long as the HSATII sequence has complementary sticky ends with the plasmid, it can be slotted into place with little technique. Further, excision of recognition sites results in seamless integration, reducing the flanking non-HSATII sequences that could impact transcript function.

An RFP intermediate serves as a marker for successful cloning

To make the H2B:GFP plasmid Golden Gate compatible, it first has to be modified to include appropriate IIS recognition sites. An intermediate plasmid was thus designed that contains an RFP cassette flanked with BsmBI recognition sites (Fig. 1, Fig. 4A). The RFP cassette adds an additional functionality to aid with cloning. Initially, as

H2B:GFP is exchanged for the RFP gBlock, successfully edited plasmids contain RFP. Unlike GFP which must be fluorescently excited, RFP is visible to the naked eye (Agmon *et al.*, 2015). Following transformation, colonies that appear red would thus contain the intermediate plasmid. This RFP marker is useful in two respects. While the H2B:GFP plasmid is successfully cut, the restriction digest is not highly efficient. Therefore most colonies simply contain unedited plasmid (Fig. S4). Further, the RFP gBlock is inserted via restriction digestion with ClaI/KpnI, but the RFP coding sequence has internal KpnI sites (Fig. S5). Since only the full-length sequence produces functional RFP, the color marker allows for selection of full-length inserts. In future Golden Gate-mediated editing using BsmBI, the RFP cassette will be exchanged for different HSATII sequence variants. In this case, white colonies will have successful edits whereas red colonies will still contain the unedited intermediate. RFP thus provides a screening system for successful Golden Gate editing, further facilitating the testing of HSATII sequence variants.

A note on RFP expression

As RFP expression from the cassette is not under a constitutive promoter but rather the *lac* operon promoter, environmental changes can impact RFP expression (Fig. S5). pAV113, the source of the RFP sequence, was thus used to determine optimal conditions for RFP expression. Preliminary single colony streaks differed in RFP production and suggested that expression may be impacted by lineage (the expression level of the original colony) and by bacterial strain. To determine if parent colony brightness reflects a change in the plasmid itself, plasmids were isolated and miniprepped

from colonies and then transformed into freshly thawed chemically competent cells. Two strains of chemically competent *E. coli*, DH5α and Top10, were used to assess for strain-dependent effects. Two formulations of LB agar were used for growth to find conditions for optimal RFP expression. Overall, no effect of lineage was observed, as bacteria of the same strain growing on the same LB expressed comparable levels of RFP (Fig. 5). However, differences were observed with respect to strain and LB formulation. Replacement of tryptone in preformulated Difco™ LB with peptone resulted in much higher RFP expression, with colonies appearing magenta in color (Fig. 5). DH5α cells expressed RFP more readily than Top10 cells. Downstream cloning should thus use DH5α chemically competent cells and peptone LB for high levels of RFP marker expression.

Wrapping up construction

Unfortunately, due to the constraints of COVID-19, there was insufficient time to complete the workflow and exchange RFP for HSATII (Fig. 1). Thus, the immediate next step is to test insertion of the initial HSATII construct—already prepped and purified—via Golden Gate cloning (Fig. 6). Following the production of the final plasmid, a co-transfection of the HSATII donor plasmid and the CRISPR-Cas9 plasmid will be performed for Tig-1 fibroblasts. Similarly to initial tests with H2B:GFP, cells will be exposed to varying levels of Dox to test for dose-dependence. HSATII will be detected using RNA FISH, and signal will be quantified using Fiji. Once initial HSATII expression from the construct is verified, any number of sequence variants can be expressed via the RFP intermediary.

Further, while the AAVS1 locus was targeted initially due to its safe harbor-like properties, downstream investigation will likely consider the effects of integration into different genomic loci. To that end, edits must be made for both the sgRNA sequence in the CRISPR-Cas9 plasmid and the HA sequences in the RFP intermediate plasmid. Building off of our success with Golden Gate cloning and its ease of use, these further insertions could also be performed via Golden Gate edits. Class-IIS endonuclease recognition sites must first be incorporated flanking the HAs and the sgRNA sequence, likely via normal restriction-ligation cloning. While any IIS enzyme can be used for the sgRNA, BsmBI cannot be used for the HAs given that they are on the RFP plasmid. Choosing a different enzyme prevents inappropriate excision of the RFP cassette. With these modifications, single pot cloning can thus be performed to change the targeted genomic locus for HSATII insertion.

A perspective on HSATII, cancer, and the relevance of this construct

Altogether, the features of this construct present an easy-to-use system for dose-dependent expression of specific HSATII sequences from determinate genomic loci in human cells. This construct thus grants a higher level of control than our previous stable line expression vector and can be used to answer detailed questions regarding the site, time, amount, and sequence of expression. With our growing understanding of the complexity of HSATII behavior, interrogating these detailed issues is essential.

In stable lines, the accumulation of HSATII has been observed to correlate with chromosomal instability and cell cycle defects—key hallmarks of cancer (Bartek *et al.*, 1999; Hartwell, 1992; Yoon *et al.*, 2002). The defects took the form of errors in mitosis

such as nuclear blebbing, lagging chromosomes, chromatin bridges, and the formation of micronuclei (Landers *et al.*, 2021). These results raise the possibility that the effects of HSATII RNAs may differ depending on cell cycle stage, with effects possibly maximized during S or M phase and minimized during the G phases. HSATII RNA expression in U2OS osteosarcoma cells has been observed to display cell cycle dependence, with high levels of expression during S phase (Akkipeddi, 2020). As expression of HSATII is inducible, this new construct enables the assessment of cell cycle dependent HSATII RNA effects. Briefly, a population of cells would first be cell cycle synchronized using chemical inhibitors such as Nocodazole, thymidine, and Brefeldin A (Liu *et al.*, 2019; Zhang *et al.*, 2017). Following release, cells will cycle together, and Dox can be added at the appropriate time points to induce HSATII expression at specific cell-cycle stages.

Since previous stable lines relied on random genomic integration of the expression construct, the observed cancer-like traits were integration site-independent (Landers *et al.*, 2021). However, that independence may be due to the consistent, high-level of HSATII exposure from the constitutive CMV promoter. In normal cancer progression, HSATII expression may gradually increase over time, with low levels being immunostimulatory until an inflection point is reached following which the high levels of HSATII detrimentally impact processes cell-wide (Tanne *et al.*, 2015). Essentially, the previous stable line model skips over the entire process by which a normal cell becomes cancerous to the endpoint with high levels of HSATII (Landers *et al.*, 2021). Stable lines generated with this novel construct can model the natural increase of HSATII in cancer, as HSATII expression can be dialed up or down over time depending on the concentration of Dox. This model may thus be more biologically representative of actual

tumorigenesis. Coupled with other assays such as RNAseq, immunostaining, or RNA pulldown, the system could reveal gradual effects of increasing levels of HSATII on gene regulation, protein activity, and cell cycle defects.

Further, given the high level of endogenous HSATII repeat heterogeneity in cells, different sequences of HSATII may play different roles or interact to promote tumorigenesis (Altemose *et al.*, 2014; Bersani *et al.*, 2015; Rubien, 2020). As the expressed HSATII sequence is easily mutable using Golden Gate cloning, different or even multiple sequences could be expressed in the gradual accumulation model to determine if certain RNA sequences are essential for cancer-like phenotype development.

In stable lines, the transcriptional regulator protein, MeCP2, was observed to colocalize with HSATII RNA foci, phenocopying the behavior of MeCP2 in cancerous cell lines (Hall *et al.*, 2017; Landers *et al.*, 2021). These observations support the molecular sponge model for HSATII contribution to tumorigenesis. While not yet validated in cancer, inappropriate protein sequestration by a repeat RNA has been implicated in other diseases, such as myotonic dystrophy, a form of muscular dystrophy (Swinnen *et al.*, 2020). In this heritable disorder, a toxic repeat expansion of $(CTG)_n$ occurs in the 3' untranslated region of DM1 protein kinase. Upon transcription, the CUG repeats sequester muscleblind-like RNA-binding proteins into nuclear foci (Miller *et al.*, 2000). Sequestration of these proteins results in missplicing of certain mRNAs in a tissue-dependent manner, results that have been confirmed in patients to correlate with clinical disease presentation (Freyermuth *et al.*, 2016; Fugier *et al.*, 2011; Savkur *et al.*, 2001). Stable lines generated with this novel construct can further verify the role of HSATII RNA protein sequestration in tumorigenesis. Given that HSATII RNA

sequesters MeCP2 which has an RNA-binding domain, the transcript sequence likely impacts the interaction. Certain HSATII sequences may thus more robustly recruit MeCP2, a sequence-dependent effect that is easily testable with this construct. Additionally, sequestration of MeCP2 by HSATII RNA may require the aggregation of HSATII into nuclear foci, a phase separation facilitated by RNA secondary structure (Rubien, 2020). As certain bulk effects are necessary for RNA interactions and for this separation to occur, foci may not be observed at low levels of HSATII expression. By using varying concentrations of Dox, a minimum concentration for focal HSATII RNA accumulation and subsequent RNA-protein interactions could be determined, relating back to the gradual increase model for HSATII expression and accumulation in cancer cells.

Overall, this construct can thus approach and model the HSATII biology of cancer cells down to very specific nuances with a high degree of resolution. By digging into these minutiae, this approach may shed light on the nature of HSATII and its complex relationship with cell health and tumorigenesis.

ACKNOWLEDGMENTS

As I watch the sun rise on this misty May morning, I cannot help but to think of all the people who have helped me reach this moment. First and foremost, I would like to thank Professor Dawn Carone for taking me under her wing almost four years ago. At the time, I most certainly did not know anything about research, but with her patience, kindness, and guidance, I have grown to love the lab as a second home. Indeed, there were many late nights where it may as well have been my home. As the rock grounding the lab, I would like to thank Christina Rabeler for all of her help and advice with transfections and just generally knowing where anything and everything could be found. It has truly been a blessing to meet everyone in the Carone lab over the years. In particular, I would like to thank Jack Rubien ’20 for helping me find my footing when I first stepped foot into Martin and Sajal Akkipeddi ’20 for his humor, friendship, and excessive guitar playing late at night in the microscopy room. Thank you to Twan Sia ’21 for being far too smart and a thoroughly distracting influence on Martin third. You are a great friend. For teaching a plebian like myself about the intricacies of statistics, I would like to thank Professor Sara Hiebert Burch. For everything she has done for me, for all of the conversations, support, advice, and occasional learning, I would like to thank Professor Amy Vollmer. For the relationships I have made in biology at Swarthmore, I would do it again. I would also like to thank Sam Yan ’21 for being my friend over all my years at Swarthmore and listening to biology even if it was not your own interest. Additionally, thank you to my senior spring suitemates, Jihye Yoon ’21 and Sally Peng ’21. I appreciated your company and support in this last stretch. Finally, thank you to my family and especially my sister. Even if we don’t talk, you are always with me.

REFERENCES

Abkevich, V., Timms, K. M., Hennessy, B. T., Potter, J., Carey, M. S., Meyer, L. A., Smith-McCune, K., Broaddus, R., Lu, K. H., Chen, J., Tran, T. V., Williams, D., Iliev, D., Jammulapati, S., FitzGerald, L. M., Krivak, T., DeLoia, J. A., Gutin, A., Mills, G. B., & Lanchbury, J. S. (2012). Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. *British Journal of Cancer*, *107*(10), 1776–1782. https://doi.org/10.1038/bjc.2012.451

Agmon, N., Mitchell, L. A., Cai, Y., Ikushima, S., Chuang, J., Zheng, A., Choi, W.-J., Martin, J. A., Caravelli, K., Stracquadanio, G., & Boeke, J. D. (2015). Yeast Golden Gate (yGG) for the Efficient Assembly of S. cerevisiae Transcription Units. *ACS Synthetic Biology*, *4*(7), 853–859. https://doi.org/10.1021/sb500372z

Akkipeddi, S. (2020). Transcription patterns of HSATII RNA in U2OS cells. Thesis submitted to Swarthmore College Department of Biology.

Altemose, N., Miga, K. H., Maggioni, M., & Willard, H. F. (2014). Genomic Characterization of Large Heterochromatic Gaps in the Human Genome Assembly. *PLOS Computational Biology*, *10*(5), e1003628. https://doi.org/10.1371/journal.pcbi.1003628

Anastasiadou, E., Jacob, L. S., & Slack, F. J. (2018). Non-coding RNA networks in cancer. *Nature Reviews Cancer*, *18*(1), 5–18. https://doi.org/10.1038/nrc.2017.99

Ausió, J., Paz, A. M. de, & Esteller, M. (2014). MeCP2: The long trip from a chromatin protein to neurological disorders. *Trends in Molecular Medicine*, *20*(9), 487–498. https://doi.org/10.1016/j.molmed.2014.03.004

Bartek, J., Lukas, J., & Bartkova, J. (1999). Perspective: Defects in cell cycle control and cancer. *The Journal of Pathology*, *187*(1), 95–99. https://doi.org/10.1002/(SICI)1096-9896(199901)187:1<95::AID-PATH249>3.0.CO;2-#

Bersani, F., Lee, E., Kharchenko, P. V., Xu, A. W., Liu, M., Xega, K., MacKenzie, O. C., Brannigan, B. W., Wittner, B. S., Jung, H., Ramaswamy, S., Park, P. J., Maheswaran, S., Ting, D. T., & Haber, D. A. (2015). Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer. *Proceedings of the National Academy of Sciences*, *112*(49), 15148–15153. https://doi.org/10.1073/pnas.1518008112

Boyer, J. C., Umar, A., Risinger, J. I., Lipford, J. R., Kane, M., Yin, S., Barrett, J. C., Kolodner, R. D., & Kunkel, T. A. (1995). Microsatellite Instability, Mismatch Repair Deficiency, and Genetic Defects in Human Cancer Cell Lines. *Cancer Research*, *55*(24), 6063–6070.

Brown, G. C. (2015). Living too long. *EMBO Reports*, *16*(2), 137–141. https://doi.org/10.15252/embr.201439518

Carew, J. S., & Huang, P. (2002). Mitochondrial defects in cancer. *Molecular Cancer*, *1*(1), 9. https://doi.org/10.1186/1476-4598-1-9

Chan, D. Y. L., Moralli, D., Khoja, S., & Monaco, Z. L. (2017). Noncoding Centromeric RNA Expression Impairs Chromosome Stability in Human and Murine Stem Cells. *Disease Markers*, *2017*, e7506976. https://doi.org/10.1155/2017/7506976

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, *339*(6121), 819–823. https://doi.org/10.1126/science.1231143

Corrà, F., Agnoletto, C., Minotti, L., Baldassari, F., & Volinia, S. (2018). The Network of Non-coding RNAs in Cancer Drug Resistance. *Frontiers in Oncology*, *8*. https://doi.org/10.3389/fonc.2018.00327

Doudna, J. A., & Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science (New York, N.Y.)*, *346*(6213), 1258096. https://doi.org/10.1126/science.1258096

Dubois, V. P., Zotova, D., Parkins, K. M., Swick, C., Hamilton, A. M., Kelly, J. J., & Ronald, J. A. (2018). Safe Harbor Targeted CRISPR-Cas9 Tools for Molecular-Genetic Imaging of Cells in Living Subjects. *The CRISPR Journal*, *1*(6), 440–449. https://doi.org/10.1089/crispr.2018.0030

Engler, C., Kandzia, R., & Marillonnet, S. (2008). A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *PLOS ONE*, *3*(11), e3647. https://doi.org/10.1371/journal.pone.0003647

Engler, C., & Marillonnet, S. (2014). Golden Gate Cloning. In S. Valla & R. Lale (Eds.), *DNA Cloning and Assembly Methods* (pp. 119–131). Humana Press. https://doi.org/10.1007/978-1-62703-764-8_9

Fellmann, C., Gowen, B. G., Lin, P.-C., Doudna, J. A., & Corn, J. E. (2017). Cornerstones of CRISPR-Cas in drug discovery and therapy. *Nature Reviews. Drug Discovery*, *16*(2), 89–100. https://doi.org/10.1038/nrd.2016.238

Freyermuth, F., Rau, F., Kokunai, Y., Linke, T., Sellier, C., Nakamori, M., Kino, Y., Arandel, L., Jollet, A., Thibault, C., Philipps, M., Vicaire, S., Jost, B., Udd, B., Day, J. W., Duboc, D., Wahbi, K., Matsumura, T., Fujimura, H., … Charlet-Berguerand, N. (2016). Splicing misregulation of SCN5A contributes to cardiac-conduction delay and heart arrhythmia in myotonic dystrophy. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms11067

Fugier, C., Klein, A. F., Hammer, C., Vassilopoulos, S., Ivarsson, Y., Toussaint, A., Tosch, V., Vignaud, A., Ferry, A., Messaddeq, N., Kokunai, Y., Tsuburaya, R., de la Grange, P., Dembele, D., Francois, V., Precigout, G., Boulade-Ladame, C., Hummel, M.-C., Lopez de Munain, A., … Charlet-Berguerand, N. (2011). Misregulated alternative splicing of BIN1 is associated with T tubule alterations and muscle weakness in myotonic dystrophy. *Nature Medicine*, *17*(6), 720–725. https://doi.org/10.1038/nm.2374

Gasiunas, G., Barrangou, R., Horvath, P., & Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences*, *109*(39), E2579–E2586. https://doi.org/10.1073/pnas.1208507109

Gossen, M., & Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences*, *89*(12), 5547–5551. https://doi.org/10.1073/pnas.89.12.5547

Gossen, M., Freundlieb, S., Bender, G., Müller, G., Hillen, W., & Bujard, H. (1995). Transcriptional Activation by Tetracyclines in Mammalian Cells. *Science*, *268*(5218), 1766–1769.

Hall, L. L., Byron, M., Carone, D. M., Whitfield, T. W., Pouliot, G. P., Fischer, A., Jones, P., & Lawrence, J. B. (2017). Demethylated HSATII DNA and HSATII RNA Foci Sequester PRC1 and MeCP2 into Cancer-Specific Nuclear Bodies. *Cell Reports*, *18*(12), 2943–2956. https://doi.org/10.1016/j.celrep.2017.02.072

Han, X., Liu, Z., Jo, M. chan, Zhang, K., Li, Y., Zeng, Z., Li, N., Zu, Y., & Qin, L. (2015). CRISPR-Cas9 delivery to hard-to-transfect cells via membrane deformation. *Science Advances*, *1*(7), e1500454. https://doi.org/10.1126/sciadv.1500454

Hartwell, L. (1992). Defects in a cell cycle checkpoint may be responsible for the genomic instability of cancer cells. *Cell*, *71*(4), 543–546. https://doi.org/10.1016/0092-8674(92)90586-2

He, X., Tan, C., Wang, F., Wang, Y., Zhou, R., Cui, D., You, W., Zhao, H., Ren, J., & Feng, B. (2016). Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Research*, *44*(9), e85–e85. https://doi.org/10.1093/nar/gkw064

Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science*, *327*(5962), 167–170. https://doi.org/10.1126/science.1179555

Ichida, K., Suzuki, K., Fukui, T., Takayama, Y., Kakizawa, N., Watanabe, F., Ishikawa, H., Muto, Y., Kato, T., Saito, M., Futsuhara, K., Miyakura, Y., Noda, H., Ohmori, T., Konishi, F., & Rikiyama, T. (2018). Overexpression of satellite alpha

transcripts leads to chromosomal instability via segregation errors at specific chromosomes. *International Journal of Oncology*, *52*(5), 1685–1693. https://doi.org/10.3892/ijo.2018.4321

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., & Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology*, *169*(12), 5429–5433. https://doi.org/10.1128/jb.169.12.5429-5433.1987

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)*, *337*(6096), 816–821. https://doi.org/10.1126/science.1225829

Jiricny, J., & Nyström-Lahti, M. (2000). Mismatch repair defects in cancer. *Current Opinion in Genetics & Development*, *10*(2), 157–161. https://doi.org/10.1016/S0959-437X(00)00066-6

Kishikawa, T., Otsuka, M., Yoshikawa, T., Ohno, M., Yamamoto, K., Yamamoto, N., Kotani, A., & Koike, K. (n.d.). Quantitation of circulating satellite RNAs in pancreatic cancer patients. *JCI Insight*, *1*(8). https://doi.org/10.1172/jci.insight.86646

Krueger, C., Pfleiderer, K., Hillen, W., & Berens, C. (2004). Tetracycline derivatives: Alternative effectors for Tet transregulators. *BioTechniques*, *37*(4), 546–550. https://doi.org/10.2144/04374BM04

Landers, C. C., Rabeler, C. A., Ferrari, E. K., D'Alessandro, L. R., Kang, D. D., Malisa, J., Bashir, S. M., & Carone, D. M. (2021). Ectopic expression of pericentric HSATII RNA results in nuclear RNA accumulation, MeCP2 recruitment, and cell division defects. *Chromosoma*, *130*(1), 75–90. https://doi.org/10.1007/s00412-021-00753-0

Lee, J. H., Skowron, P. M., Rutkowska, S. M., Hong, S. S., & Kim, S. C. (1996). Sequential amplification of cloned DNA as tandem multimers using class-IIS restriction enzymes. *Genetic Analysis: Biomolecular Engineering*, *13*(6), 139–145. https://doi.org/10.1016/S1050-3862(96)00164-7

Li, G., Liu, D., Zhang, X., Quan, R., Zhong, C., Mo, J., Huang, Y., Wang, H., Ruan, X., Xu, Z., Zheng, E., Gu, T., Hong, L., Li, Z., Wu, Z., & Yang, H. (2018). Suppressing Ku70/Ku80 expression elevates homology-directed repair efficiency in primary fibroblasts. *The International Journal of Biochemistry & Cell Biology*, *99*, 154–160. https://doi.org/10.1016/j.biocel.2018.04.011

Li, G., Zhang, X., Zhong, C., Mo, J., Quan, R., Yang, J., Liu, D., Li, Z., Yang, H., & Wu, Z. (2017). Small molecules enhance CRISPR/Cas9-mediated homology-directed genome editing in primary cells. *Scientific Reports*, *7*(1), 8943. https://doi.org/10.1038/s41598-017-09306-x

Lieber, M. R. (2010). The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annual Review of Biochemistry*, *79*(1), 181–211. https://doi.org/10.1146/annurev.biochem.052308.093131

Lin, S., Staahl, B. T., Alla, R. K., & Doudna, J. A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *ELife*, *3*, e04766. https://doi.org/10.7554/eLife.04766

Liu, M., Rehman, S., Tang, X., Gu, K., Fan, Q., Chen, D., & Ma, W. (2019). Methodologies for Improving HDR Efficiency. *Frontiers in Genetics*, *9*, 691. https://doi.org/10.3389/fgene.2018.00691

Loew, R., Heinz, N., Hampf, M., Bujard, H., & Gossen, M. (2010). Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnology*, *10*, 81. https://doi.org/10.1186/1472-6750-10-81

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., & Church, G. M. (2013). RNA-Guided Human Genome Engineering via Cas9. *Science*, *339*(6121), 823–826. https://doi.org/10.1126/science.1232033

Marraffini, L. A., & Sontheimer, E. J. (2010). CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nature Reviews. Genetics*, *11*(3), 181–190. https://doi.org/10.1038/nrg2749

Meyer-Ficca, M. L., Meyer, R. G., Kaiser, H., Brack, A. R., Kandolf, R., & Küpper, J.-H. (2004). Comparative analysis of inducible expression systems in transient transfection studies. *Analytical Biochemistry*, *334*(1), 9–19. https://doi.org/10.1016/j.ab.2004.07.011

Miller, J. W., Urbinati, C. R., Teng-umnuay, P., Stenberg, M. G., Byrne, B. J., Thornton, C. A., & Swanson, M. S. (2000). Recruitment of human muscleblind proteins to (CUG)n expansions associated with myotonic dystrophy. *The EMBO Journal*, *19*(17), 4439–4448. https://doi.org/10.1093/emboj/19.17.4439

Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of Molecular Evolution*, *60*(2), 174–182. https://doi.org/10.1007/s00239-004-0046-3

Nicolas, F. E. (2017). Role of ncRNAs in Development, Diagnosis and Treatment of Human Cancer. *Recent Patents on Anti-Cancer Drug Discovery*, *12*(2), 128–135. https://doi.org/10.2174/1574892812666170105113415

Nogalski, M. T., Solovyov, A., Kulkarni, A. S., Desai, N., Oberstein, A., Levine, A. J., Ting, D. T., Shenk, T., & Greenbaum, B. D. (2019). A tumor-specific endogenous repetitive element is induced by herpesviruses. *Nature Communications*, *10*(1), 90. https://doi.org/10.1038/s41467-018-07944-x

Padgett, K. A., & Sorge, J. A. (1996). Creating seamless junctions independent of restriction sites in PCR cloning. *Gene*, *168*(1), 31–35. https://doi.org/10.1016/0378-1119(95)00731-8

Pantano, F., Perrone, G., Vincenzi, B., Iuliani, M., Fioramonti, M., Zoccoli, A., Ribelli, G., Borzomati, D., Nappo, G., Pellegrini, C., Amato, M. M., Righi, D., Di Matteo, F. M., Trodella, L., Santini, D., Coppola, R., Onetti-Muda, A., & Tonini, G. (2015). Long non-coding RNA HSAT II as a new biomarker for the identification of high risk intraductal papillary mucinous neoplasms (IPMNs). *Journal of Clinical Oncology*, *33*(15_suppl), e15246–e15246. https://doi.org/10.1200/jco.2015.33.15_suppl.e15246

Papapetrou, E. P., & Schambach, A. (2016). Gene Insertion Into Genomic Safe Harbors for Human Gene Therapy. *Molecular Therapy*, *24*(4), 678–684. https://doi.org/10.1038/mt.2016.38

Pellenz, S., Phelps, M., Tang, W., Hovde, B. T., Sinit, R. B., Fu, W., Li, H., Chen, E., & Monnat, R. J. (2019). New Human Chromosomal Sites with "Safe Harbor" Potential for Targeted Transgene Insertion. *Human Gene Therapy*, *30*(7), 814–828. https://doi.org/10.1089/hum.2018.169

Pham, D. H., Moretti, P. A. B., Goodall, G. J., & Pitson, S. M. (2008). Attenuation of leakiness in doxycycline-inducible expression via incorporation of 3′ AU-rich mRNA destabilizing elements. *BioTechniques*, *45*(2), 155–162. https://doi.org/10.2144/000112896

Rossi, F. M. V., Kringstein, A. M., Spicher, A., Guicherit, O. M., & Blau, H. M. (2000). Transcriptional Control: Rheostat Converted to On/Off Switch. *Molecular Cell*, *6*(3), 723–728. https://doi.org/10.1016/S1097-2765(00)00070-8

Rubien, J. (2020). RNA phase separation in cancer: Investigating HSATII RNA folding and function. Honors thesis submitted to Swarthmore College.

Savkur, R. S., Philips, A. V., & Cooper, T. A. (2001). Aberrant regulation of insulin receptor alternative splicing is associated with insulin resistance in myotonic dystrophy. *Nature Genetics*, *29*(1), 40–47. https://doi.org/10.1038/ng704

Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2021). Cancer Statistics, 2021. *CA: A Cancer Journal for Clinicians*, *71*(1), 7–33. https://doi.org/10.3322/caac.21654

Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, *70*(1), 7–30. https://doi.org/10.3322/caac.21590

Søndergaard, J. N., Geng, K., Sommerauer, C., Atanasoai, I., Yin, X., & Kutter, C. (2020). Successful delivery of large-size CRISPR/Cas9 vectors in hard-to-transfect human cells using small plasmids. *Communications Biology*, *3*(1), 1–6. https://doi.org/10.1038/s42003-020-1045-7

Sun, Y., Chen, X., & Xiao, D. (2007). Tetracycline-inducible Expression Systems: New Strategies and Practices in the Transgenic Mouse Modeling. *Acta Biochimica et Biophysica Sinica*, *39*(4), 235–246. https://doi.org/10.1111/j.1745-7270.2007.00258.x

Swinnen, B., Robberecht, W., & Van Den Bosch, L. (2020). RNA toxicity in non-coding repeat expansion disorders. *The EMBO Journal*, *39*(1). https://doi.org/10.15252/embj.2018101112

Szybalski, W., Kim, S. C., Hasan, N., & Podhajska, A. J. (1991). Class-IIS restriction enzymes—A review. *Gene*, *100*, 13–26. https://doi.org/10.1016/0378-1119(91)90345-C

Tagarro, I., Fernández-Peralta, A. M., & González-Aguilera, J. J. (1994). Chromosomal localization of human satellites 2 and 3 by a FISH method using oligonucleotides as probes. *Human Genetics*, *93*(4), 383–388. https://doi.org/10.1007/BF00201662

Tanne, A., Muniz, L. R., Puzio-Kuter, A., Leonova, K. I., Gudkov, A. V., Ting, D. T., Monasson, R., Cocco, S., Levine, A. J., Bhardwaj, N., & Greenbaum, B. D. (2015). Distinguishing the immunostimulatory properties of noncoding RNAs expressed in cancer cells. *Proceedings of the National Academy of Sciences*, *112*(49), 15154–15159. https://doi.org/10.1073/pnas.1517584112

Tejero, R., Huang, Y., Katsyv, I., Kluge, M., Lin, J.-Y., Tome-Garcia, J., Daviaud, N., Wang, Y., Zhang, B., Tsankova, N. M., Friedel, C. C., Zou, H., & Friedel, R. H. (2019). Gene signatures of quiescent glioblastoma cells reveal mesenchymal shift and interactions with niche microenvironment. *EBioMedicine*, *42*, 252–269. https://doi.org/10.1016/j.ebiom.2019.03.064

Ting, D. T., Lipson, D., Paul, S., Brannigan, B. W., Akhavanfard, S., Coffman, E. J., Contino, G., Deshpande, V., Iafrate, A. J., Letovsky, S., Rivera, M. N., Bardeesy, N., Maheswaran, S., & Haber, D. A. (2011). Aberrant Overexpression of Satellite Repeats in Pancreatic and Other Epithelial Cancers. *Science*, *331*(6017), 593–596. https://doi.org/10.1126/science.1200801

Urlinger, S., Baron, U., Thellmann, M., Hasan, M. T., Bujard, H., & Hillen, W. (2000). Exploring the sequence space for tetracycline-dependent transcriptional activators: Novel mutations yield expanded range and sensitivity. *Proceedings of*

the National Academy of Sciences, *97*(14), 7963–7968. https://doi.org/10.1073/pnas.130192197

Vidal, M., & Starowicz, K. (2017). Polycomb complexes PRC1 and their function in hematopoiesis. *Experimental Hematology*, *48*, 12–31. https://doi.org/10.1016/j.exphem.2016.12.006

Wallace, D. C. (2012). Mitochondria and cancer. *Nature Reviews Cancer*, *12*(10), 685–698. https://doi.org/10.1038/nrc3365

Weber, E., Gruetzner, R., Werner, S., Engler, C., & Marillonnet, S. (2011). Assembly of Designer TAL Effectors by Golden Gate Cloning. *PLOS ONE*, *6*(5), e19722. https://doi.org/10.1371/journal.pone.0019722

White, M. C., Holman, D. M., Boehm, J. E., Peipins, L. A., Grossman, M., & Henley, S. J. (2014). Age and Cancer Risk. *American Journal of Preventive Medicine*, *46*(3 0 1), S7-15. https://doi.org/10.1016/j.amepre.2013.10.029

Xie, Z., Pang, D., Wang, K., Li, M., Guo, N., Yuan, H., Li, J., Zou, X., Jiao, H., Ouyang, H., Li, Z., & Tang, X. (2017). Optimization of a CRISPR/Cas9-mediated Knock-in Strategy at the Porcine Rosa26 Locus in Porcine Foetal Fibroblasts. *Scientific Reports*, *7*(1), 3036. https://doi.org/10.1038/s41598-017-02785-y

Xu, J. (2020). Mortality in the United States, 2018. *NCHS Data Brief*, *355*, 8.

Xu, X., Gao, D., Wang, P., Chen, J., Ruan, J., Xu, J., & Xia, X. (2018). Efficient homology-directed gene editing by CRISPR/Cas9 in human stem and primary cells using tube electroporation. *Scientific Reports*, *8*(1), 11649. https://doi.org/10.1038/s41598-018-30227-w

Yabroff, K. R., Lund, J., Kepka, D., & Mariotto, A. (2011). Economic Burden of Cancer in the US: Estimates, Projections, and Future Research. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *20*(10), 2006–2014. https://doi.org/10.1158/1055-9965.EPI-11-0650

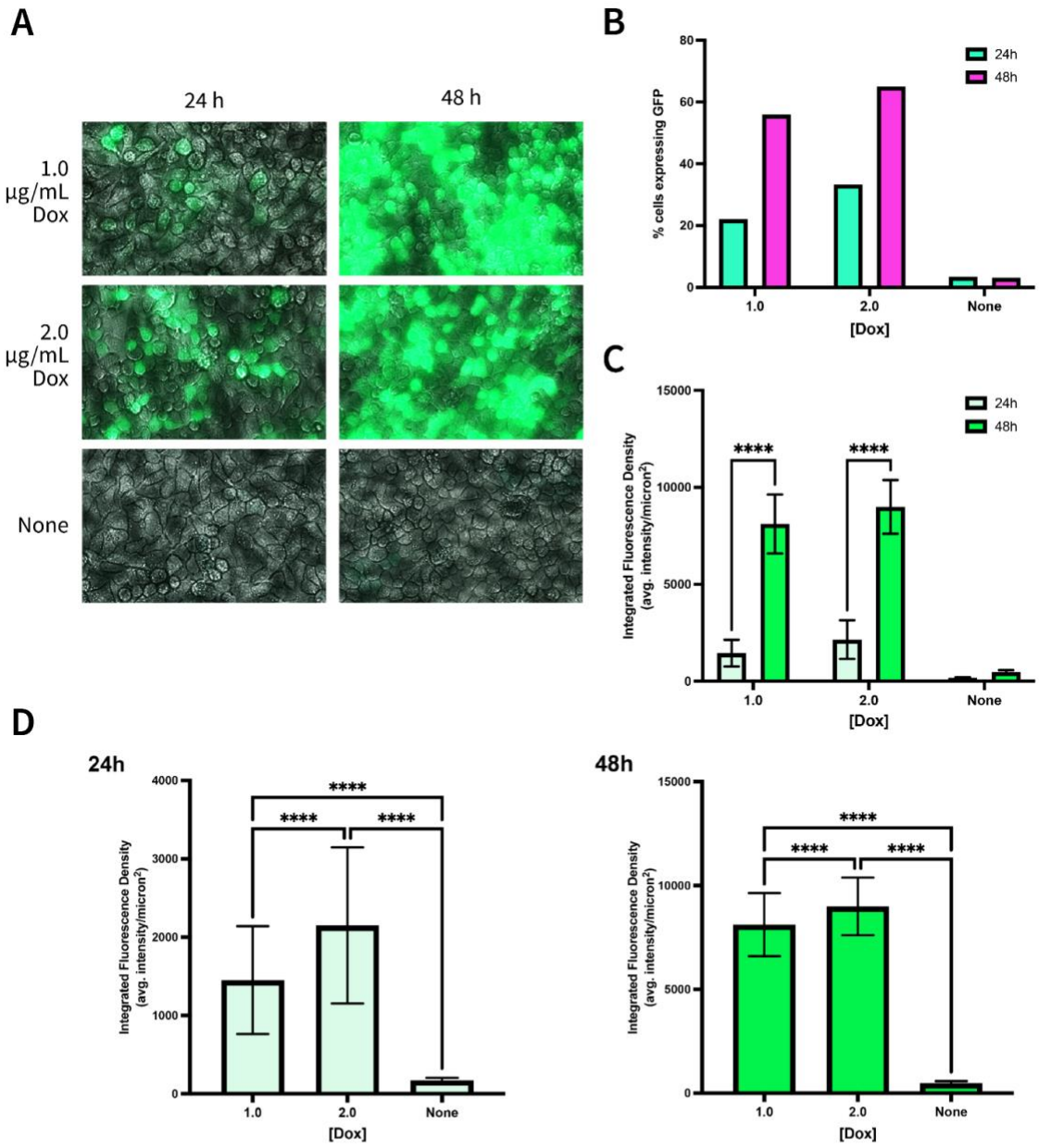
Yoon, D.-S., Wersto, R. P., Zhou, W., Chrest, F. J., Garrett, E. S., Kwon, T. K., & Gabrielson, E. (2002). Variable Levels of Chromosomal Instability and Mitotic Spindle Checkpoint Defects in Breast Cancer. *The American Journal of Pathology*, *161*(2), 391–397. https://doi.org/10.1016/S0002-9440(10)64194-6

Zhang, J.-P., Li, X.-L., Li, G.-H., Chen, W., Arakaki, C., Botimer, G. D., Baylink, D., Zhang, L., Wen, W., Fu, Y.-W., Xu, J., Chun, N., Yuan, W., Cheng, T., & Zhang, X.-B. (2017). Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biology*, *18*(1), 35. https://doi.org/10.1186/s13059-017-1164-8

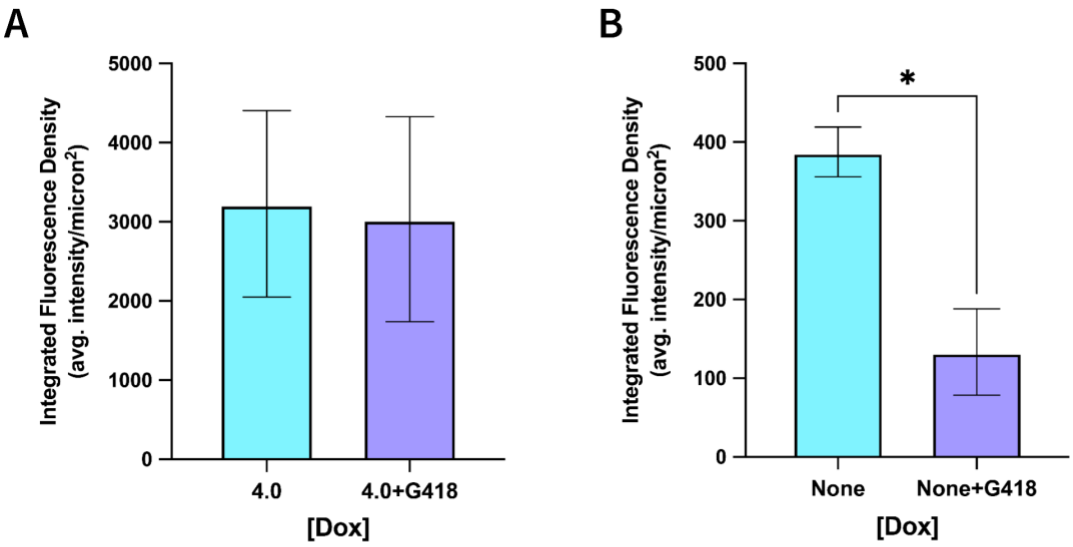
Zhu, Q., Hoong, N., Aslanian, A., Hara, T., Benner, C., Heinz, S., Miga, K. H., Ke, E., Verma, S., Soroczynski, J., Yates, J. R., Hunter, T., & Verma, I. M. (2018). Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. *Molecular Cell*, *70*(5), 842-853.e7. https://doi.org/10.1016/j.molcel.2018.04.023

Supplemental Table 1: Key Reagents

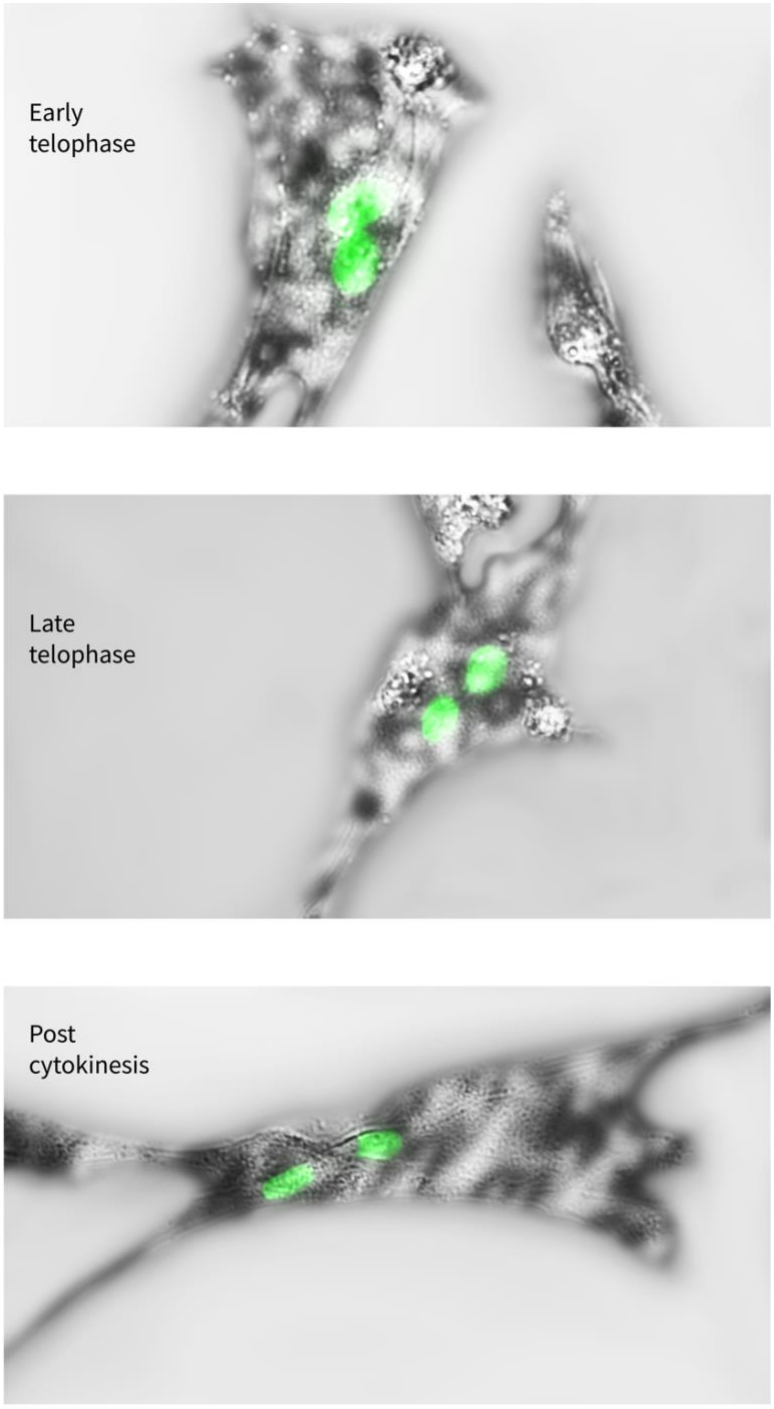
Reagent Type	Designation	Source/Reference	Identifiers	Additional Information
Primer	RFP_gBlock_F	Integrated DNA Technologies		5'-TAGAATTCATCGATGAGACGGC-3'
Primer	RFP_gBlock_R	Integrated DNA Technologies		5'-GTACCGAGACGGCCACCTTAAGC-3'
Primer	Sat2X_F	Integrated DNA Technologies		5'-GGAACCGAATGAATCCTCATTGAATG-3'
Primer	Chr7_conF	Integrated DNA Technologies		5'-ATTCGATTCCATTCGATGATGATTCC-3'
Primer	HSATII_BsmBI_F	Integrated DNA Technologies		5'-GGCTACCGTCTCGGTACGGAACCGAAT-GAATCC-3'
Primer	HSATII_BsmBI_R	Integrated DNA Technologies		5'- GGCTACCGTCTCATCGAATTCGATTCCAT-TCG -3'
Primer	RFP_P_Int_F (Checks for plasmid integration)	Integrated DNA Technologies		5'- GCATTCATTTTATGTTTCAGG -3'
Primer	RFP_P_Int_R (Checks for plasmid integration)	Integrated DNA Technologies		5'-GTCGAGTTTACTCCCTATCAG-3'
Transfected Plasmid	pX330-sgAAVS1	AddGene	Plasmid 85802	Gift from Roland Friedel
Transfected Plasmid	pAAVS1-Neo-CAG-M2rtTA-H2BGFP (donor AB)	AddGene	Plasmid 85798	Gift from Roland Friedel
Transfected Plasmid	pT077	Addgene	Plasmid 137879	Gift from Lindy Barrett
RFP Source Plasmid	pAV113	AddGene	Plasmid 63180	Gift from Jef Boeke
HSATII Source Plasmid	pTargeT_HSATII	Landers *et al.*, 2021		
Cell line (H. sapiens)	HeLa			
Cell line (H. sapiens)	Tig-1	Coriell	Coriell:AG06173	
Chemical compound, drug	Geneticin™ Selective Antibiotic (G418 Sulfate)	Gibco (Thermo Fisher Scientific)	Thermo Fisher:10131027	
Commercial assay or kit	FuGENE® HD Transfection Reagent	Promega	Promega:E2311	
Commercial assay or kit	DNA Clean & Concentrator®-5	Zymo Research	Zymo:D4014	
Commercial assay or kit	ZymoPURE™ Plasmid Miniprep	Zymo Research	Zymo:D4210	
Commercial assay or kit	GENECLEAN® Turbo	MP Biomedicals	MPbio:1102200	
Enzyme	KpnI-HF®	New England BioLabs	NEB:R3142	
Enzyme	ClaI	New England BioLabs	NEB:R0197	
Enzyme	BsmBI-v2	New England BioLabs	NEB:R0739	
Enzyme	T4 DNA Ligase	New England BioLabs	NEB:M0202	
Bacteria	Subcloning Efficiency™ DH5α Competent Cells	Invitrogen (Thermo Fisher Scientific)	Thermo Fisher:18265017	
Bacteria	One Shot™ TOP10 Chemically Competent E. coli	Invitrogen (Thermo Fisher Scientific)	Thermo Fisher:C404010	



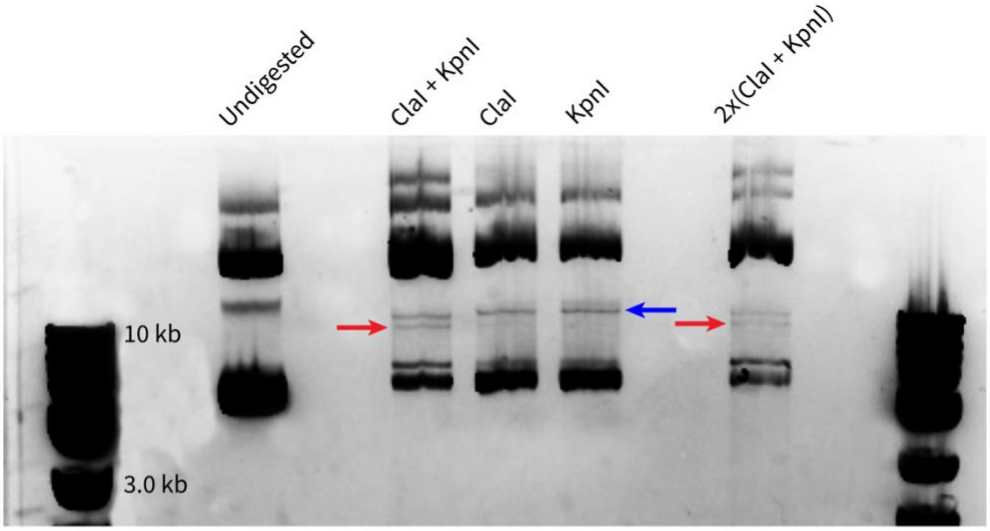
Supplemental Figure 1. Dox-inducible GFP expression in HeLa cells is dose-dependent and accumulative. Following transfection and G418 selection, expression was induced with exposure to varying Dox concentrations for up to 48 h. A) Representative images of cells exposed to 1.0 and 2.0 µg/mL of Dox at 24 and 48 h. B) The percentage of cells expressing GFP for all concentrations and timepoints. The proportion of expressing cells increased over time and with increasing Dox concentration. C) Average signal intensities per micron² in expressing cells. A significant effect of time was noted, as signal intensity significantly increased from 24h to 48h for both tested Dox concentrations. D) A significant effect of Dox concentration on expression was observed within each timepoint.
(****) p<0.0001



Supplemental Figure 2. Assessing G418 for potential side effects and interactions.
As G418 is a cytotoxic chemical, potential interactions with Dox and side effects on construct expression were investigated. A) G418 was not observed to interact with Dox to either increase or decrease construct expression even at the highest Dox concentration. B) G418 alone did not induce inappropriate expression, but rather repressed basal levels of construct expression, thus reducing signal noise.
(*) p<0.05



Supplemental Figure 3. H2B:GFP accumulates in actively dividing cells
Representative images of H2B:GFP expressing Tig-1 cells undergoing different stages of mitosis and cytokinesis. H2B:GFP+ cells frequently were observed to contain two daughter nuclei and were observed to congregate in actively dividing clusters.



Supplemental Figure 4. KpnI and ClaI digests are highly inefficient
To assess for restriction digest efficacy, digestion products were run on an agarose gel. Overall, most bands overlapped between undigested and digested plasmids. Single restriction enzyme digests appeared similar and distinct from the undigested plasmid, with both products displaying a new band at ~11 kb, indicating that linearization occurred (marked in blue). In double digest conditions, a second, unique band was observed (marked in red). Double digestion should result in excision of a 1 kb sequence. As this second band consists of shorter DNA than the single digest band, it is likely that this band represents the double digested product. The band is very thin and faint compared to other bands, indicating that the digest had low efficiency.

5’-TAGAATTCATCGATGAGACGGCGGCCGCTTTTCACACAGGAAACAGCTATGACC
ATGATTACGCCAAGCGCGCAATTAACCCTCACTAAAGGGAACAAAAGCTGGAGCTC
CACCGCGGTGGCGGCCGCTCTAGAACTAGTGGATCCCCAAAATGAGACCCAATAC
GCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAG
GTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAGCTCA
CTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAA
TTGTGAGCGGATAACAATTTCACACATACTAGAGAAAGAGGAGAAATACTAGATGG
CTTCCTCCGAAGACGTTATCAAAGAGTTCATGCGTTTCAAAGTTCGTATGGAAGGTT
CCGTTAACGGTCACGAGTTCGAAATCGAAGGTGAAGGTGAAGGTCGTCCGTACGA
AGGTACCCAGACCGCTAAACTGAAAGTTACCAAAGGTGGTCCGCTGCCGTTCGCTT
GGGACATCCTGTCCCCGCAGTTCCAGTACGGTTCCAAAGCTTACGTTAAACACCCG
GCTGACATCCCGGACTACCTGAAACTGTCCTTCCCGGAAGGTTTCAAATGGGAACG
TGTTATGAACTTCGAAGACGGTGGTGTTGTTACCGTTACCCAGGACTCCTCCCTGC
AAGACGGTGAGTTCATCTACAAAGTTAAACTGCGTGGTACCAACTTCCCGTCCGAC
GGTCCGGTTATGCAGAAAAAAACCATGGGTTGGGAAGCTTCCACCGAACGTATGTA
CCCGGAAGACGGTGCTCTGAAAGGTGAAATCAAAATGCGTCTGAAACTGAAAGACG
GTGGTCACTACGACGCTGAAGTTAAAACCACCTACATGGCTAAAAAACCGGTTCAG
CTGCCGGGTGCTTACAAAACCGACATCAAACTGGACATCACCTCCCACAACGAAGA
CTACACCATCGTTGAACAGTACGAACGTGCTGAAGGTCGTCACTCCACCGGTGCTT
AAGGTGGCCGTCTCGGTACCGGCGCATA-3’

Supplemental Figure 5. RFP gBlock Sequence
Full sequence of the RFP gBlock ordered from AddGene.

5’-ggctaccgtctcggtacGGAACCGAATGAATCCTCATTGAATGGAATCGAATGGAATCAT
CATTGAATGGAATCGAATGGAATCATCATCGAATGGAATCGAATGGAATCATCATCG
AATGGAATCATTATCAAATAAAATGGAATGGAATCATTGAATGGAATCGAATGGAAT
CATCATCGAATGGAATCGAACGGTATCATCAAATGGAATCGAATGGAATCATCATTG
AATAGAATCGAATGGAATCATCGAATGGAACTGAATGGAATCATCATTGAATGGAAT
CGAATGGAATCACCGAATGGAATCGAATGGAATCATCAAATGGAATCGAATGGAAT
CATCATCGAATGGAATCGAATtcgatgagacggtagcc-3’

Supplemental Figure 6. HSATII sequence

The full length sequence of the HSATII product extracted from the pTargeT vector. Chr7 HSATII sequence is in uppercase, while added sequences for BsmBI-mediated Golden Gate cloning are in lowercase.