Aptamer-mediated transactivation of transcription utilizing CRISPR/Cas9 and a photoreceptor

Dissertation

zur Erlangung des Doktorgrades (Dr. rer. nat.)

der Mathematisch-Naturwissenschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

Vorgelegt von

M. Sc. Christian Renzl

aus Köln

Bonn, Juli 2021

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn.

1. Gutachter: Prof. Dr. Günter Mayer

2. Gutachter: Prof. Dr. Michael Pankratz

Tag der Promotion: 03.12.2021

Erscheinungsjahr: 2021

Teile dieser Arbeit sind veröffentlicht in:

Renzl C, Kakoti A, Mayer G. Aptamer-mediated reversible transactivation of gene expression by light. *Angew Chem Int Ed*. 2020;59(50):22414-22418.

DOI: 10.1002/anie.202009240

Beteiligung an erfolgreich eingereichten Manuskripten:

Weber AM, Kaiser J, Ziegler T, et al. A blue light receptor that mediates RNA binding and translational regulation. *Nat Chem Biol*. 2019;15(11):1085-1092.

DOI: 10.1038/s41589-019-0346-y

Content

I Abstract

In synthetic biology, spatio-temporal control of molecular functions is an essential tool for the investigation and manipulation of cellular processes. The usage of light enables a reversible non-invasive trigger, that stimulates photoreceptors, which translate energy into biological signals. Based on the photoreceptor PAL and CRISPR/dCas9, a platform was developed, that mediates light-dependent reversible transactivation of gene expression. While genetic coding space was minimized, the system combines a strong on-switch in the light with a very low background activity in the dark. The new tool adds to the existing building blocks used in synthetic biological circuit design and extends the degree of complexity for these systems.

II Summary

For a better understanding of cellular processes, spatio-temporal control of molecular functions is an important experimental approach. The usage of light enables a reversible non-invasive trigger, that stimulates photoreceptors. PAL is a photoreceptor, that reversibly changes its conformation in the presence of blue light. Aptamers are short structured nucleic acids that bind to targets with high affinity and specificity. The RNA aptamers 04 and 53 bind to the light conformation of PAL, whereas in darkness the binding is much less. CRISPR/Cas9 describes an RNA-guided endonuclease which consists of a complex between the Cas9 protein and a guide RNA (gRNA). In this work, the aptamer 53 and its interaction partner PAL were used as a light-dependent interface, that enables the conditional recruitment of regulatory factors to specific genomic loci. When the aptamer 53 is inserted in the gRNA and PAL is fused to transcriptional activators such as p65 and Heat shock factor 1 (HSF1), the co-expression with dCas9, a cleavage deficient variant of Cas9, enables for light-dependent gene expression. In a prove of principle experiment, an enhanced Blue Fluorescent Protein (eBFP) reporter was co-expressed along with the gRNAs, dCas9 and the PAL fusion protein in HeLa cells, which led to light-dependent transactivation of the reporter gene. The system was then optimized regarding aptamer stem length, plasmid ratios, light-intensity and the order of transactivation domains. Finally, endogenous Achaete-scute homolog 1 (ASCL1) was activated in HeLa cells up to 99-

fold higher in the light, compared to the sample in the dark. In summary, a new tool was generated, which is capable of activating endogenous genes in the light, while in the dark gene expression was maintained at a basal level.

Zeitlich und räumlich aufgelöste Kontrolle über zelluläre Abläufe sind eine wichtige Hilfe um molekulare Mechanismen besser zu verstehen. Hierbei eignet sich Licht als Auslöser für Reaktionen, da es reversibel und nicht-invasiv eingesetzt werden kann um beispielsweise Photorezeptoren zu aktivieren. PAL ist ein Photorezeptor, welcher seine Konformation abhängig vom Licht reversibel verändern kann. Aptamere sind kurze strukturierte Nukleinsäure Sequenzen, die mit hoher Affinität und Selektivität an Zielmolekülen binden können. Das RNA Aptamer 53 bindet die Lichtkonformation von PAL mit hoher Affinität, während die Bindung im Dunkeln stark reduziert ist. CRISPR/Cas9 ist eine auf RNA-Ebene programmierbare Endonuklease, welche aus einem Komplex aus Cas9 und der guide RNA (gRNA) gebildet wird. Innerhalb dieser Arbeit wurden die Aptamere 53 und 04 verwendet, welche die Lichtkonformation von PAL binden. Die Aptamere wurden in die gRNAs integriert um entweder lichtabhängig die Komplexbildung von gRNA und Cas9 zu verhindern oder um mit einer katalytisch inaktiven Form von Cas9 (dCas9) Gene im Licht zu aktivieren. Im zuletzt genannten Fall wurde PAL an die Transkriptionsfaktoren p65 und Heat Shock Factor 1 (HSF1) fusioniert, welche sich lichtabhängig mit Hilfe der gRNA und dCas9 an regulatorische Elemente von Promotoren rekrutieren lassen. Mit diesem System konnte ein enhanced Blue Fluorescent Protein (eBFP) Reportergen lichtabhängig in HeLa Zellen überexprimiert werden. Das System wurde hinsichtlich der Stammlänge der Aptamere, der Menge eingesetzter Plasmide, der Lichtintensität und der Ausrichtung der Transkriptionsfaktoren optimiert. Die optimierte Version konnte im Licht und im Vergleich zu einer Probe im Dunkeln endogenes Achaete-scute homolog 1 protein (ASCL1) in HeLa Zellen 99-fach über exprimieren. Zusammenfassend wurde ein neues Werkzeug etabliert, welches in der Lage ist native Gene im Licht zu aktivieren.

III Introduction

III.1 Optogenetics

Almost all organisms depend on light for their survival in regard of growing (photosynthesis), phototaxis, phototropism, circadian rhythm or a higher form of vision[1] **(Figure 1)**. The ability of photoreceptors to mediate photon energy into biological signals, that alter cell fate is the basis for optogenetics. The separation of biochemical reactions within cells by compartmentation is a key factor, since series of these reactions such as oxidative phosphorylation or photosynthesis are bound to membrane surfaces[2]. In order to gain reversible control of biomolecular interactions on a subcellular level, light can be used as a non-invasive stimulus, to control these processes with high spatial and temporal resolution[3, 4].

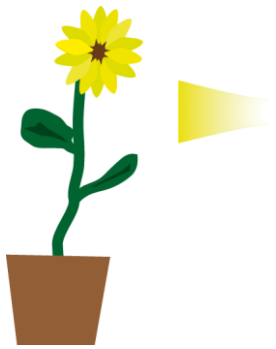


Figure 1: Sunflowers align their leaves towards the position of the sun throughout the day. This phenomenon called phototropism is enabled by photoreceptors.

Light-sensing proteins, i.e. photoreceptor proteins transduce photon energy into cellular responses and can be used to control protein expression and function[5]. The microbial channelrhodopsins were explored by Hegemann and Nagel[6] **(Figure 2)**. They were first applied *in vitro*[7] in 2005 when they were expressed in neurons. Two years later in 2007 they were used *in vivo*[8] to enable control of motor cortex in mice. In 2010, optogenetics was declared to be the 'method of the year' by the journal Nature.

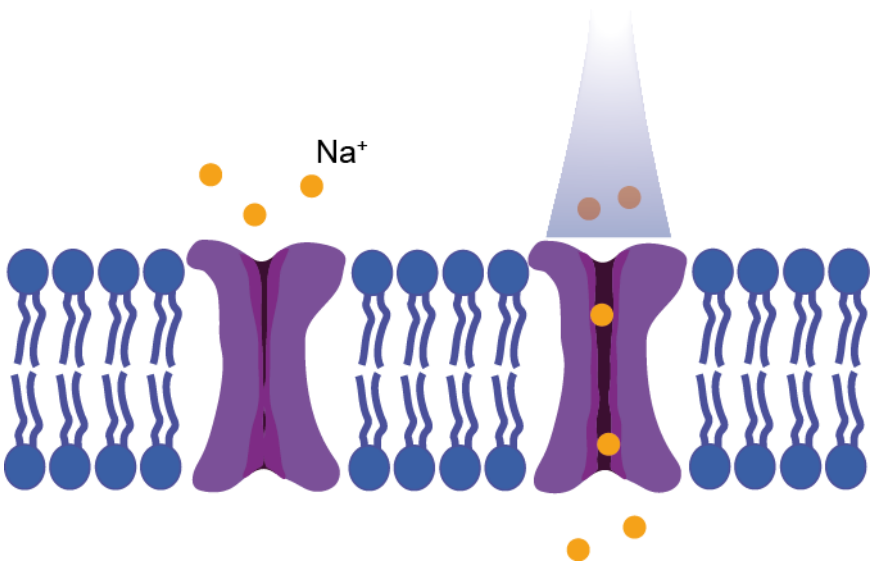


Figure 2: Channelrhodopsins change their transmission characteristics upon irradiation with light. They can be used to activate neurons with a light stimulus.

Advanced genetics and recombinant DNA technology allow the expression of photoreceptors in host cells enabling them to respond to light[9]. These tools provide new possibilities to image, detect and control biological processes with unrivalled precision[10]. Since the first application of channelrhodopsins, optogenetics have been used for genome editing[11, 12], transcription modulation[13-15], mRNA localization[16], protein hydrolysis[17, 18], neuronal inhibition[19], neuronal stimulation[7], protein interaction[20, 21] and other cellular functions such as cell morphology[9]. To date a whole set of photoreceptors has been characterized and used in synthetic biology.

III.2 Photoreceptors in synthetic biology

The generation of fusion proteins of a light sensing unit with and an effector unit is one of the most common methods to upgrade cellular functions with a light switch.[22-24]. In order to sense light, the cells need proteins, that contain a chromophore, which transmits the signal to the protein scaffold[10]. These chromophores are distributed across the visible spectrum of light and some are depicted in **Figure 3**. Flavin mononucleotide (FMN) or flavin adenine dinucleotide (FAD) are chromophores, that are incorporated in photoreceptor proteins like light-oxygen-voltage-sensing (LOV)

domains, blue-light-utilizing flavin adenine dinucleotide (BLUF) domains or cryptochromes[25]. In cyanobacteriochromes, phycocyanobilin is used and in phytochromes biliverdin IXα (BV) is used as the light-sensing molecule[26]. The circular process of a photoreceptor protein changing its conformation to the light conformation and back to the dark conformation is called a photocycle and is reversible for all photoreceptor proteins[10].

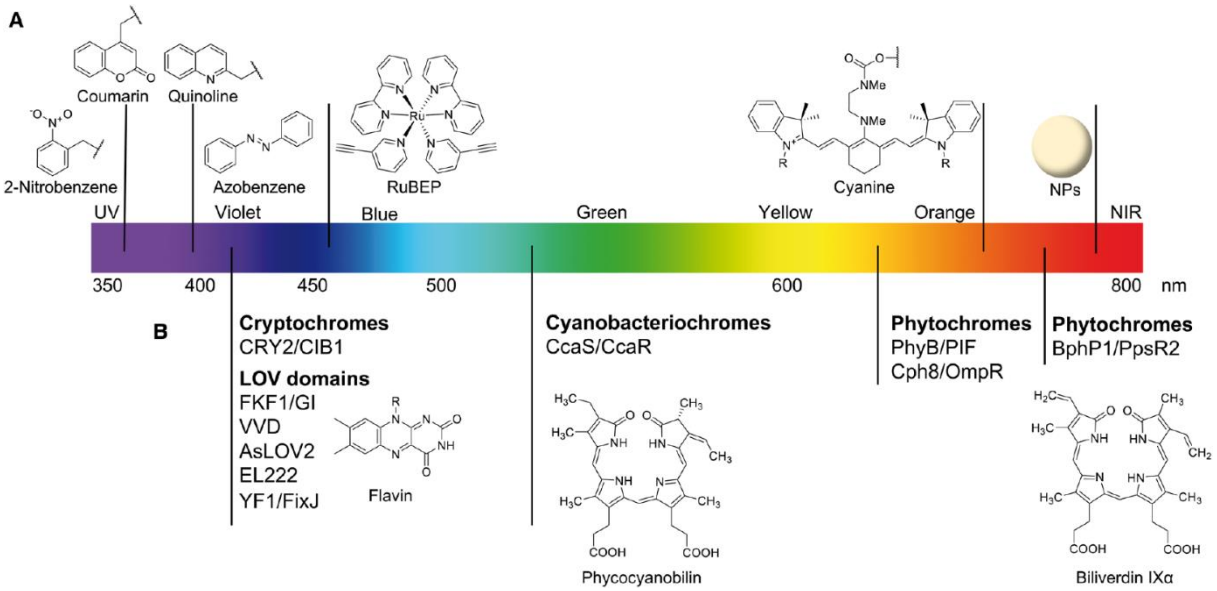


Figure 3: Different chromophores and their absorption characteristics along the spectrum of visible light[27]. The absorption characteristics for chromophores range from ultraviolett up to the near infrared region. They are crucial for sensing light in photoreceptors.

Photoreceptors can be engineered for various applications in synthetic biology. In this work the focus is on light-dependent regulation of transcription using engineered photoreceptors. In all forms of life, transcription is a key step in converting genetic information on DNA level into different kind of RNAs. Transcription generates messenger RNAs that translate into proteins or into regulatory RNAs such as long noncoding RNAs (lncRNA) or micro RNAs (miRNA). Eukaryotic transcription is a tightly regulated process since it involves the formation of a preinitiation complex using transcription factors and cis-acting elements. Briefly, transcription can be described by three steps, which are transcription initiation, elongation and termination. Further, transcription is regulated on an epigenetic level as for example by introducing DNA methylation or histone modification. Therefore, control of transcription is highly

important, since it enables to study protein expression and regulation. Researching biological pathways and gene expression includes the control and analysis of transcription and translation[28]. Gene expression can be controlled using light by implementing chemical or biological sensors[29]. Chemical photocages for example are light-sensitive molecules, that block protein function. Upon irradiation with a specific wavelength, the photocage is dissociated from the target molecule or changes the conformation towards a bioactive domain. Also, small molecule inducers can be photocaged and thus enable light-dependent control of transcription. For example, estradiol and tamoxifen have been photocaged to control reporter gene expression in mammalian cells[30-32]. Gene expression and phenotype control was applied in zebrafish using 2-nitrobenzyl- or thiocoumarin-caged cyclophen[33, 34]. The well-known Tetracycline (Tet) system was photocaged using a Nitrobenzyl-photocage and used to control reporter gene expression in developing mouse embryos and xenopus tadpoles[35]. Light-dependent control of gene expression can also be achieved by photocaging DNA or RNA[36-39].

Besides using small molecules or modified nucleic acids for light-dependent gene regulation, proteins are often the means of choice. Although, proteins like Zinc Finger (ZF) nucleases can be chemically modified using 2-nitrobenzyl[40, 41] or coumarin[42] to conditionally activate recombination, the usage of non-modified light-sensing proteins provides the advantage to completely become expressed by the host cell. Different photoreceptors have been used for the regulation of gene expression. The three most important classes of photoreceptor proteins are LOV-domain proteins, cryptochromes and phytochromes. Phytochromes are sensitive for near infrared (NIR)-light. The incorporated chromophore BV or phycocyanobilin induces a reversible change in protein conformation upon irradiation. The irradiation of phytochrome protein B (PCB) and its interaction partner phytochrome-interacting factor 3 (PIF3) with red light, leads to the formation of a heterodimer. When irradiated with far-red light, the heterodimer splits up. Plants for example use phytochromes to regulate flowering. The inactive form of phytochrome P_r converts into the active form P_{fr}, when irradiated with red light, which is more present during the day. P_{fr} is converted back to P_r, when irradiated with far red light or is automatically converted to P_r in the dark or at night **(Figure 4)**. This system was engineered to regulate gene expression. DNA-binding domains (DBDs) and transactivation domains (ADs) were fused to both modules. Irradiated with light, the modules form dimers, enabling light-dependent control of gene expression in yeast[43]

and mammalian cells[44]. The requirement of the PCB chromophore can be satisfied by adding it to the growth media[43, 44]. However, PCB can also be expressed by the host cells directly[45, 46]. NIR-responsive phytochromes, which use the BV chromophore can rely on endogenous expressed BV of mammalian host cells[47, 48].

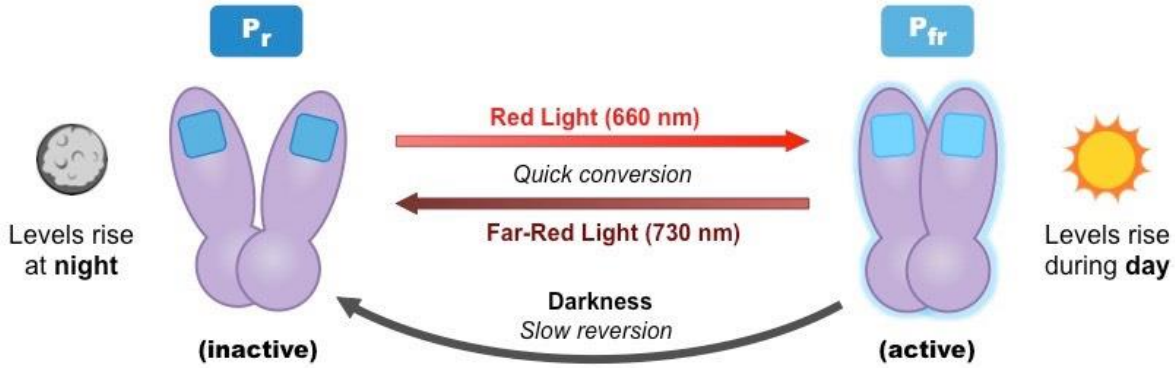


Figure 4: Phytochromes regulate photoperiodism in plants. During the day, inactive phytochrome is converted to its active form by the parts of red light from the sun light. At night, the active form is converted back into the inactive form[a].

Besides the red-light sensitive photoreceptors there are blue-light responsive proteins such as LOV domain proteins and cryptochromes. Both receptors harbor a flavin chromophore to induce conformation changes in the protein backbone upon irradiation with blue light. In 2009, Möglich, Ayers and Moffat used blue light in a synthetic transcription system. They engineered a blue light sensitive kinase YF1 by exchanging an oxygen-sensing domain of the native kinase with the YtvA LOV domain[49]. Often the photoreceptor proteins are fused to DNA binding domains (DBD) and transactivation domains (TAD), to enable light-dependent control of gene expression. The blue light receptor cryptochrome 2 (CRY2) from *Arabidopsis thaliana* and the CRY2 interaction partner CIB1, as well as Vivid (VVD)[50] from *Neurospora* were modified in such a way to modulate transcription[51, 52] and translation[53, 54] in mammalian cells.

In contrast to the heterodimerizing systems described above, most of the single component systems are based on AsLOV2[55] or EL222[56] **(Figure 5)**. AsLOV2 , the

[a] bioninja.com.au/higher-level/topic-9-plant-biology/untitled-3/photoperiodism.html 30.12.20.

LOV2 domain of *Avena sativa* phototropin 1, can release the alpha helix from the LOV domain after a light-induced conformational change, which often is used to mask and modulate an attached domain with cellular function such as nuclear import/export[57] or gene activation[58, 59].

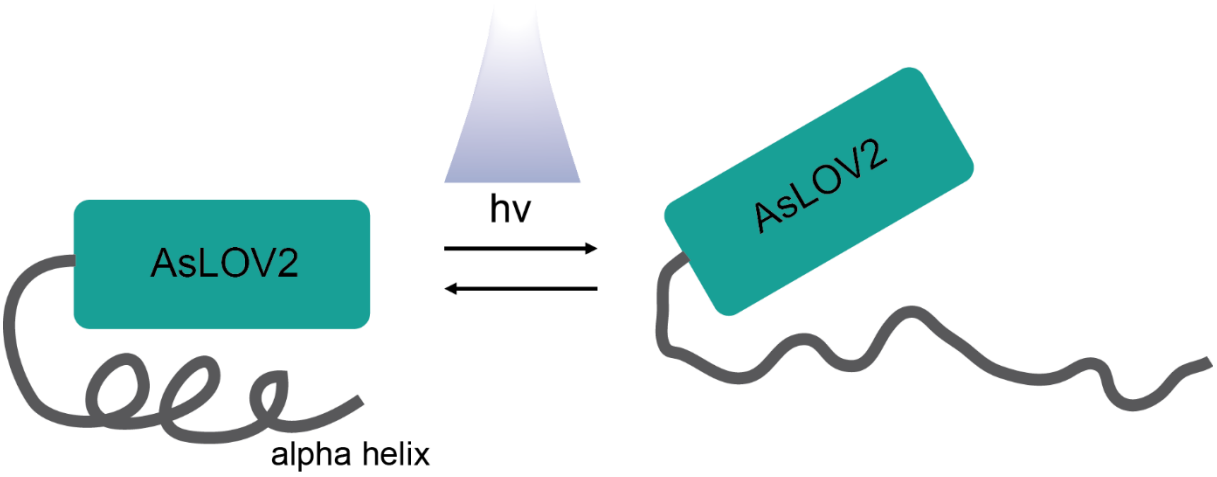


Figure 5: Schematic of AsLOV2 conformational change. Upon irradiation with blue light, a conformational change leads to the release of the alpha helix (grey).

Tunable, light-controlled interacting protein tags (TULIPs)[60] are based on AsLOV2 and bind due to a synthetic interaction between an engineered PDZ domain (ePDZ) and AsLOV2. The alpha helix of AsLOV2 was engineered to contain a PDZ binding motif, which cannot bind to its counterpart ePDZ in the dark. However, blue light induces a conformational change in AsLOV2 and the release of the alpha helix increases the affinity to ePDZ. TULIPs were tested for optogenetic control of gene expression, intracellular localization, and control of GTPase signaling *in vivo* [61, 62]. The *Neurospora crassa* LOV protein VIVID (VVD) LOV domain was engineered to bind to a VVD interaction partner, while both have the opposite electrostatic charge. This interaction was dubbed the "Magnets" system, which consists of two parts. The "positive Magnet" (pMag) variant (I52R/M55R) has a positively charged arginine at residue 52, whereas the "negative Magnet" (nMag) variant (I52D/M55) has a negatively charged aspartic acid at residue 52[63]. The magnets system was used to enable light-dependent control of neuronal differentiation in induced pluripotent stem cells (iPSCs)[64].

Generally, the binding of photoreceptors towards DNA is achieved by fusing DNA binding proteins to them. However, these DNA binding domains rely on conserved

sequences to bind to and do not allow for binding towards any endogenous sequence. Thus, a molecular system is needed, that combines DNA binding by user-defined sequence definition. A system capable of these characteristics is CRISPR/Cas9.

III.3 CRISPR/Cas9

The discovery of CRISPR (clustered, regulary interspaced, short, palindromic repeats)-Cas (CRISPR-associated) began more than 25 year ago, when a sequence of short DNA repeats was found in the *E. coli* genome[65]. Two decades later, these sequences were identified in numerous bacteria, archaea and exogenous mobile elements[66, 67]. So called Cas genes were found to be associated to these sequence arrays, from which a lot showed homology to nucleic acid binding proteins and endonucleases[68-70]. The initial hypothesis by Makarova, that CRISPR/Cas might act similar to an RNA-guided interference against foreign DNA was solidified by an experiment, when Cas proteins induced interference upon bacteriophage infection in bacteria[71].The hypothesis, CRISPR being a system providing adapted immunity in prokaryotes was experimentally confirmed, when bacteria became resistant to bacteriophages after a second infection[72-75]. In the following years, the molecular mechanism of RNA-directed interference by CRISPR-Cas was studied in detail and can be described in three major steps. First, new spacer sequences need to be acquired. Second, the CRISPR array is transcribed and processed into mature small crRNAs and third, an RNA/Cas complex is formed, which binds and cleaves the target DNA[76-78]. CRISPR/Cas divides into three sub-classes called Type I, II and III, which involve unique Cas proteins and differ in structure and molecular function from each other[79, 80]. During the spacer acquisition, Cas1 and Cas2 are both involved in all three Cas sub-classes[81-83] and are necessary for recognizing DNA secondary structures of CRISPR repeat sequences[84]. During crRNA maturation, the sub-classes differ from each other, where sub-classes Type I and Type III involve multimeric protein complexes, sub-class Type II only needs a single protein, Cas9[85]. In this work, the focus is con Cas9, since it is the best studied variant so far. In 2012, Charpentier, Doudna, and colleagues found, that CRISPR/Cas9 can be used as a programmable RNA-guided endonuclease[86]. Only one year later, CRISPR/Cas9 was adapted for genome editing in human cells by Zhang and Church[87, 88]. The most commonly used

Cas9 ortholog is from *Streptococcus pyogenes* (SpyCas9), which is also used in this study. It is a bi-lobed protein with a molecular weight of 160 kDa and contains two nuclease domains, HNH and RuvC **(Figure 6)**[89].

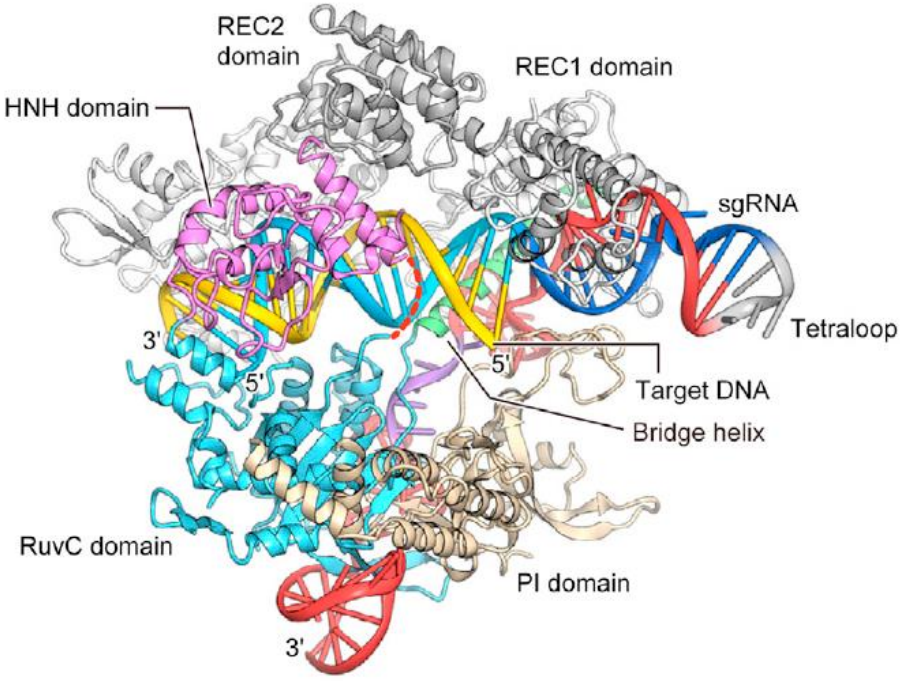


Figure 6: Crystal structure of Cas9. Crystal structure of the Cas9 enzyme in complex with the sgRNA[89]. The enzyme harbors two nuclease domains, the RuvC domain and the HNH domain. The protospacer-adjacent motif (PAM) is bound by the PAM-interacting (PI) domain. The two REC domains promote binding to the DNA. The protein can fold into three conformations, which are unbound, bound to RNA and bound to RNA and DNA.

Upon binding to DNA and in presence of the protospacer-adjacent motif (PAM), which is upstream of the 20 nt guide region, the DNA is cleaved on both strands[90]. It utilizes two non-coding RNAs, the crRNA and the trans-activating crRNA (tracrRNA). They form a crRNA:tracrRNA duplex, which can be bound by Cas9[91]. First, the PI domain of Cas9 recognizes the PAM site, which triggers the R loop formation. Subsequently, the HNH domain is directed towards the complementary strand and induces cleavage on the complementary strand. The RuvC domain cleaves the non-complementary strand[89]. In 2012, Jinek *et al*. modified the crRNA:tracrRNA duplex and fused them to a single guide RNA (sgRNA), by introducing a tetra-loop[86]. Besides the traditional PAM sequence for Cas9, which is NGG, PAM-relaxed mutations of Cas9 were generated[92, 93]. A schematic of the Cas9/sgRNA complex formation and induction of a double-strand

break is shown in **Figure 7**. The double-strand break introduced by Cas9 cleaving the DNA can trigger two repair pathways.

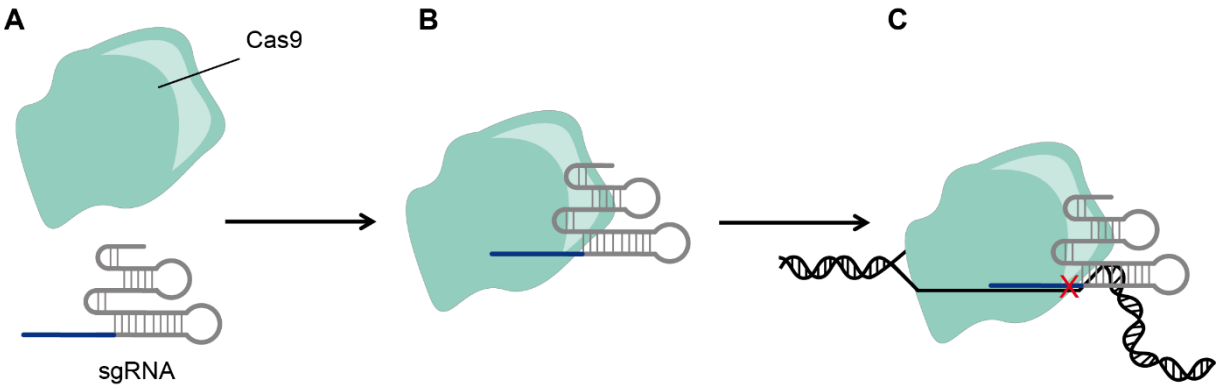


Figure 7: Schematic of Cas9 and sgRNA (A) forming a complex (B), that binds to a target DNA strand and induces a double strand break (C). In order to bind to DNA, Cas9 needs to be in complex with a gRNA or sgRNA. The complex then scans the DNA and upon sequence match, a conformational change induces the cleavage of the bound DNA sequence. After cleavage, the enzyme remains on the DNA.

Nonhomologous end-joining (NHEJ) is highly error-prone and cells introduce mutations in the repaired regions, leading to frame-shifts in coding sequences of proteins. Homology-directed repair (HDR) is another repair mechanism, which involves the sister chromatid as a template to restore the original sequence. This pathway can be used for genome editing by introducing artificial DNA templates with homologous ends[94]. Besides Cas9, further interesting Cas proteins are Cas12 and Cas13. Cas12 is capable of processing RNA guides from multicistronic RNAs, which simplifies gRNA multiplexing. It also generates four base staggered ends upon cleavage[95-97]. Cas13 is capable of cleaving or editing RNA instead of DNA[98-100]. One major restriction of working with Cas9 is its bulky size, which consists of 1368 aa. The size decreases the efficiency to introduce the genetic information into viruses, since the packaging capacity of viruses is limited. The choice of virus is important, since usually Cas9 shall only be expressed transiently. This excludes the usage of viruses that integrate the expression cassette of Cas9 into the genome. The discovery of new and smaller Cas proteins might enable the usage of Cas9 in viruses such as adeno-associated viruses (AAVs). One example is CasX[101], which was found in groundwater bacteria and has a size of 980 aa only. Another interesting Cas variant is CasΦ[102], which was discovered in megabacteriophages and has a size of ~670 aa.

III.4 Control of Cas9 activation

Since for genome editing usually only a transient activation of Cas9 is preferred, several strategies emerged, to conditionally activate Cas9 for the reduction of unspecific genome editing events. One strategy is the chemical masking of sgRNAs using 2-azidomethylnicotinyl (AMN) groups, which can be unmasked by a reduction reaction. Moreover, these sgRNAs are highly resistant to different ribonucleases[103]. A more complex way of modulating sgRNA activity is the insertion of ribozymes into the hairpin structures of the sgRNA. Chen *et al*. introduced the theophylline aptazyme into the tetraloop and stem loop 2 of the sgRNA backbone. Upon addition of the theophylline, the aptazymes are cleaved, hydrolyzing the sgRNA and thereby disabling the function of Cas9[104]. Likewise, based on the theophylline aptamer, Lin *et al*. developed small molecule-activated allosteric aptamer regulating (SMART)-sgRNAs, that enable the control of sgRNA activity by self-hybridization of the sgRNA and refolding upon addition of the ligand[105]. A temperature sensitive variant of Cas9 was generated introducing the LOV domain of *Rhodobacter sphaeroides* (RsLOV) into the protein backbone of Cas9 [106]. In regard of genome editing, specificity enhancement of CRISPR/Cas9 genome editing was achieved in HEK293T cells by adding anti-CRISPR (Acr) proteins. For example, AcrIIA4 was fused to Cas9 and delivered into mammalian cells by Adeno-associated viruses (AAVs)[107]. On this basis several optogenetic strategies for controlling Cas9 activity were developed. For example, the CASANOVA and the CASANOVA-C3 systems use AsLOV2 modified Acr, which light-dependently modulates the accessibility of the Acr towards Cas9[108]. The system was tested in HEK293T cells using a luciferase reporter assay[109]. Photo switchable control of CRISPR/Cas9 was also achieved by engineering a split Cas9 variant, where each part of the split protein was modified using pdDropna domains. Dropna is a green fluorescent photoreceptor protein[110]. The authors investigated the indel frequency in the human GRIN2B locus in HEK293T cells, to demonstrate the conditional arresting of the protein in the dark[111]. In order to circumvent damage to the cells introduced by blue light, a far-red light-activated split-Cas9 (FAST) was developed using the bacterial phytochrome protein BphS[112]. Yu and colleagues enabled far-red light induced genome editing of the PLK1 oncogene in a mouse xenograft tumor model[113] **(Figure 8)**.

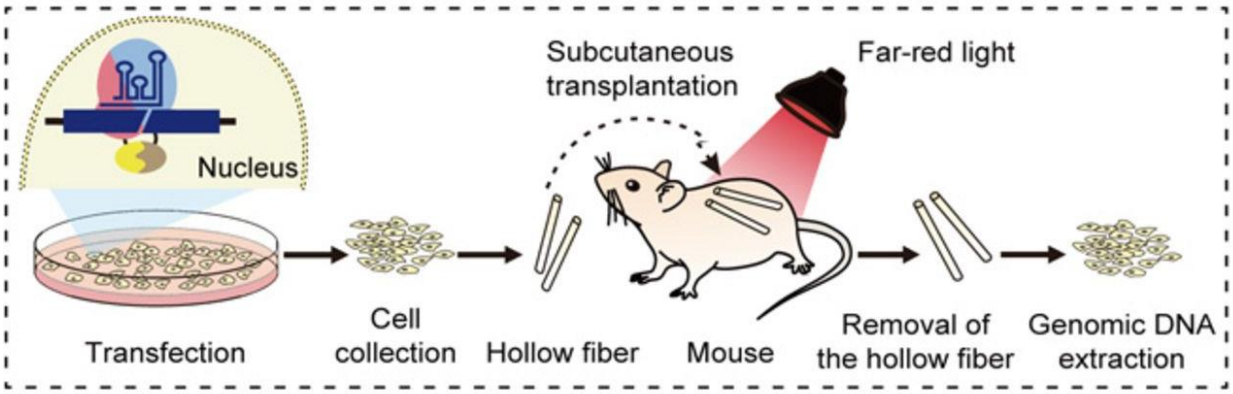


Figure 8: Application of the FAST system in a xenograft mouse model. HEK293 cells expressing the FAST system were subcutaneously implanted in mice using hollow fiber and irradiated with red light for 4 hours per day and a total of 2 days. Gene editing efficiency was measured at the CCR5 locus and detected using a cleavage assay[113].

III.5 dCas9 and gene regulation

Although, Cas9 is an RNA-guided endonuclease, it can also be used as a molecular navigation system, that recruits new functions and modulators towards DNA. A cleavage deficient variant of Cas9 (including mutations D10A and H841A)[86], dubbed dCas9 still binds to DNA and is capable to decrease transcription(also known as CRISPR interference, CRISPRi)[114]. Regulatory proteins, such as transcription factors, repressors or epigenetic modulators can be fused to dCas9 to induce/repress transcription or modulate the chromatin state[115-122] **(Figure 9)**. Before CRISPR/Cas9 emerged, programmable and systematic modulation of gene expression could be achieved using RNA interference (RNAi) or engineered DNA-binding proteins such as zinc finger or transcription-activator-like effector (TALEN) proteins[123-125]. Although RNAi can be programmed to knock down gene expression, it is limited to specific organisms due to the processing machinery needed. Zinc finger and TALEN proteins are costly to design and to develop[126]. The first application using dCas9 for transcriptional activation was published in 2013 by Perez-Pinera *et al.*, when a 4x copy of VP16, a transcriptional activation protein derived from Herpes Simplex Virus, was fused to dCas9[116]. Based on this strategy, further improved variants for CRISPR activation (CRISPRa) were developed.

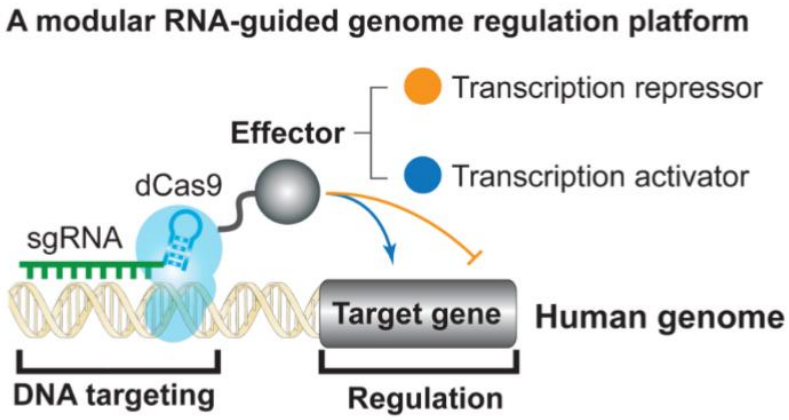


Figure 9: Schematic of targeted genome modulation using dCas9 fused to effector domains. Effectors can be addressed to modulate transcription by increasing or decreasing transcription activity[127].

Tanenbaum *et al.* developed a peptide-based scaffold recruitment system called SunTag, which accumulates 24 repeats of VP64 domains through a GCN4-single-chain variable fragment interaction and greatly increasing transcription activation[128, 129]. The synergistic activation mediators (SAM) system uses rationally designed MS2 aptamer extended sgRNAs, to tether the transcriptional activators p65 and heat shock factor 1 (HSF1) towards the Cas9/sgRNA complex, uncoupling the transactivation domains from a covalent fusion to dCas9[118] **(Figure 10)**. Similarly, the VPR system utilizes VP64, p65 and the Epstein-Barr virus R transactivator (rta) as transcriptional activators[130]. These improved systems were successfully used *in vivo* to upregulate Pdx1 in pancreatic-like beat cells, to restore insulin production in a mouse model. The upregulation of Interleukin 2 (IL-2) maintained renal function upon cisplatin-induced kidney damage[131]. Here, the SAM system was successfully applied for epigenetic gene activation using multiple AAVs[131]. Since dCas9-VP64 is too big for insertion into a single AAV, a mouse model stably expressing Cas9 was used for the study, which expresses wild-type Cas9. However, traditional sgRNAs would induce double-strand breaks (DBS) within the mouse genome. Thus, dead sgRNAs (dgRNAs), which are shortened in length and do not induce a DBS when bound to DNA were used[132, 133]. Neurons could be reprogrammed by upregulation ASCL1, NEUROG2 and NEUROD1[129]. Besides the optimization of transcription activator components, further properties are important for flexible and potent gene expression. Multiplexing of sgRNAs allows for the simultaneous activation of gene networks. However, Cas9 cannot distinguish between different sgRNAs, which is a problem,

when different genes shall be affected with varying transcriptional activators. Tethering the transcriptional activators to interfacing proteins, that bind to RNA aptamers situated inside the sgRNAs is an advantageous option, since the target definition is now directly coupled to the transcriptional activators involved in the upregulation of the gene. In fact, sgRNAs can tolerate modifications of some regions, which enables the insertion of peptide binding aptamers[117]. Using different sgRNAs with varying aptamers such as MS2, PP7 or the com aptamer enables for complex genetic programs and genetic logic gates[134-137]. Multiplexing can also be achieved by expressing sgRNA cognates via RNA polymerase II (RNAP II) using flanking ribozymes[138].

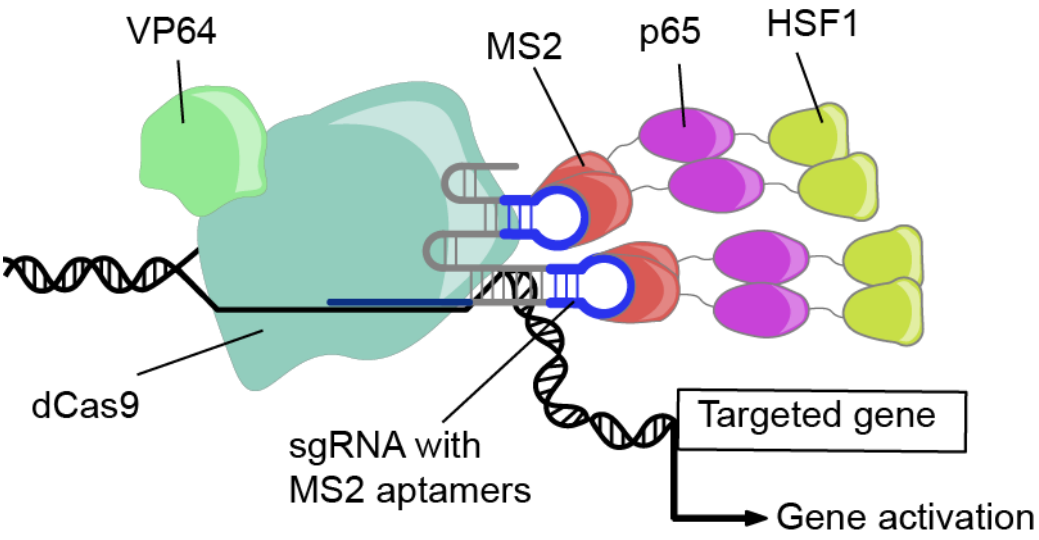


Figure 10: Schematic of the synergistic activation mediators (SAM) system. Effectors are not covalently fused to dCas9 anymore but are tethered by MS2 aptamers towards the sgRNA. The increased local concentration of transactivation domains amplifies the amount of transcription upregulation (image adapted and modified)[118].

III.6 Aptamers

Aptamers are oligonucleotides (ssDNA or RNA), that fold into defined three-dimensional structures, which allow them to bind to ligands with a high specificity and selectivity[139, 140]. Aptamers are found in nature as natural evolved structures for example in gene regulation of the coat proteins of phages MS2 or PP7[141, 142]. They can be generated by systematic evolution of ligands by exponential enrichment (SELEX), in which a library of oligonucleotides is incubated with a target of interest. Sequences,

that bind to the target are recovered and amplified to enter the next round of selection, while sequences that do not bind to the target are discarded. This process was first described in 1990 by Ellington and Szostak and by Gold and Tuerk[143, 144]. Aptamers have the potential to be used as diagnostic tools and therapeutic agents[145, 146]. The first aptamer drug pegaptanib (Macugen) was approved for treating age related macular degeneration[147, 148]. When aptamers are expressed in cells, they are often called intramers[149] **(Figure 11)**. Intramers were selected for nuclear targets[150-153] and cytoplasmic targets[154-156]. It was shown, that intramers are capable of blocking free histones in a murine model[157]. Aptamers can also mimic native binding sites of proteins in order to regulate transcriptional circuits[158]. Fluorogenic aptamers are used for monitoring metabolites *in vivo*[159]. These examples highlight the potential usage of aptamers *in vitro* and *in vivo* and also demonstrate their applicability within cells. Systems like the SAM system use aptamers to tether transcriptional activators towards Cas9/sgRNA complexes. The usage of photoreceptors in this regard enables the strong on-switch activity but under control of user-defined light pulses.

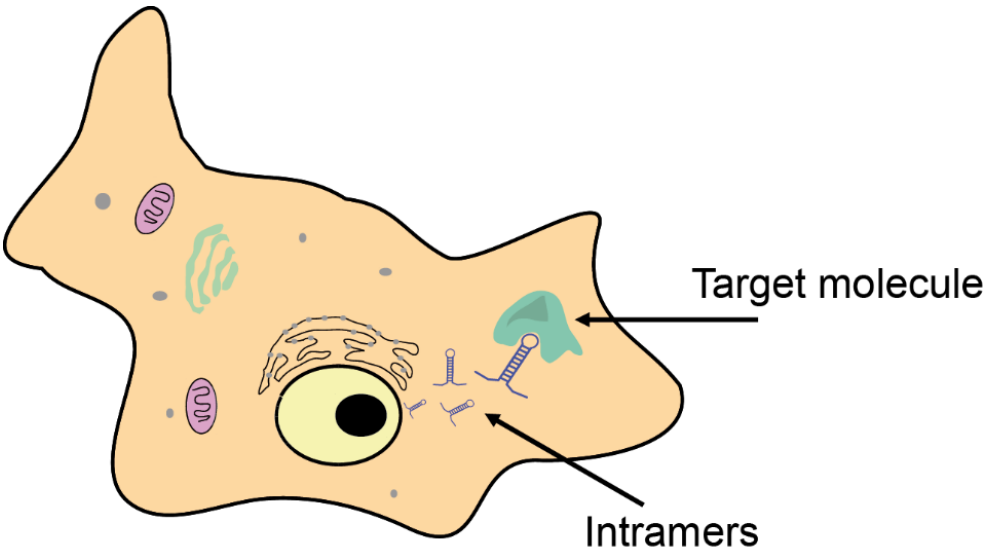


Figure 11: Intramers are nucleic acids that are expressed in host organisms. The intramers (blue) are expressed by the host cell and can bind to intracellular targets (green), in order to modulate their function.

III.7 Optogenetic gene activation with CRISPR/Cas9

Optogenetic control of transcription offers spatio-temporal control and allows for precise and reversible gene activation. The first attempts combining CRISPRa with optogenetics made use of light inducible dimerizers, which were fused to dCas9 or transcriptional activators such as VP64 or p65 and facilitate light-dependent gene expression in mammalian cells[160, 161]. As blue light sensors, the blue light receptor cryptochrome 2 (CRY2) from *Arabidopsis thaliana* and CRY2's interaction partner CIB1 were used. Various genes such as IL1RN, ASCL1, MYOD1 and NANOG were successfully upregulated in human cells, when irradiated with blue light. Finally, these tools are not limited to human cells but were also successfully used in zebra fish[162]. In order to further boost the amount of transactivation, the SAM system was extended with CRY2/CIB1 and named CPTS2.0, which allows for the light-dependent recruitment of multiple transcriptional activators to a single dCas9/sgRNA complex[64]**(Figure 12)**. Split optogenetic dCas9-based transcriptional activation as shown for the Split-CPTS 2.0 system was developed for minimal background activity, since the single split Cas9 fragments do not bind to DNA[64]. Constantly bound dCas9 to promoter regions perturbs gene expression by a road-block mechanism[114]. The split fragments were extended with pMag and nMag, which dimerize upon irradiation with blue light[63]. Split Cas9 can also be used to further reduce background activity by separating one fragment from the nucleus, which is only recruited to the nucleus upon induction[163]. Split-CPTS2.0 was the first split and optogenetic dCas9 based system inducing neuronal differentiation of induced pluripotent stem cells (iPSC)[64]. Light-dependent activation of Cas9 was also achieved by introducing the *Rhodobacter sphaeroides* (Rs)LOV domain at a suitable position within SpyCas9. Irradiation triggers the dissociation of the RsLOV heterodimer, which forms in the dark, and thereby restores Cas9 activity. Reporter expression was increased 6.5-fold in the light, when compared to the expression in the dark[106]. The binding activity of Cas9 towards the DNA was modulated by introducing a dimeric Dropna variant[164] in the Cas9 binding cleft[111], a strategy, that is applicable in Cas9 and dCas9.

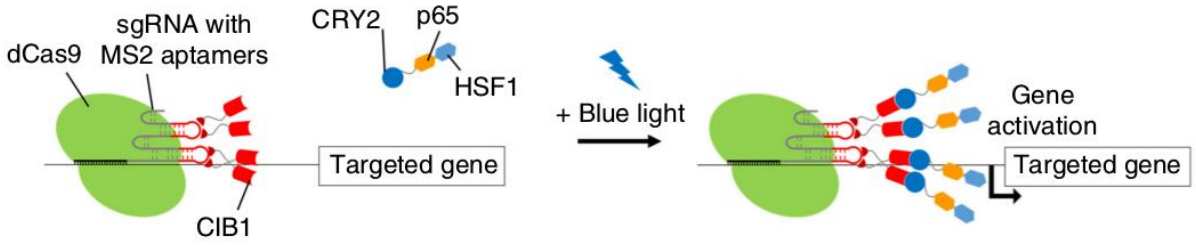


Figure 12: Schematic of the CPTS2.0 system. The SAM system was further developed by introducing CRY2 and CIB2 domains to the tethering mechanism. While MS2-CIB1 remains bound to the MS2 aptamers within the sgRNA, CRY2, which is fused to the transactivation domains is only recruited to the complex, when irradiated with blue light[64].

So far, most of the light-dependent dCas9-based transcription activation systems are triggered by blue light[64, 161], which can induce phototoxic effects on the organisms applied to. Further, shorter wavelengths are less potent in penetrating tissue[44]. Efforts, to use far-red light for photoactivation of Cas9 were made by Shao *et al.* who engineered a far-red light (FRL)-activated CRISPR-dCas9 effector (FACE), which was shown to induce neuronal differentiation *in vivo*[14]. Besides the recruitment of transcriptional activators, Light-activated dynamic looping (LADL) can be applied to drive endogenous enhancer elements towards promoter sequences, which was shown to induce light-dependent expression of Zfp462[165].

The discovery and the disclosure of the molecular mechanisms behind CRISPR/Cas led to a myriad of biotechnological applications, especially for the Cas protein Cas9. However, the published technologies indicate, that CRISPR/Cas9 is a well-suited tool for the combination of aptamers with photoreceptors to modulate protein function or gene expression.

IV Aim

The aim of this work was to modulate Cas9 activity using the photoreceptor PAL. Using PAL and aptamer 04 and 53 offer a direct and conditional interface between RNA and protein biology. One idea was a conditionally sterically blocked Cas9, which can be activated in the dark. It was hypothesized, that the introduction of aptamer 04 at different positions within the sgRNA enables PAL binding to the sgRNA in a light-dependent manner. Upon binding of PAL to the sgRNA, the accessibility of the sgRNA for Cas9 shall be blocked by orientation and electrostatic repulsion. This system would be a great addition to existing systems, that light-dependently regulate Cas9 activity. When co-expressed with PAL, in the light Cas9 and also dCas9 activity would be reduced and the conditional mode of action would contribute to a user-defined genome editing activity which reduces off-target effects caused by non-conditional Cas9 approaches.

Another approach was to generate a variation of a dCas9 based optogenetic transcription modulator, which enables spatio-temporal control of transcription of user defined genes. Since for transcriptional regulation cleavage is not needed, the cleavage deficient dCas9 will be used. In this setup the sgRNA/dCas9 complex is used as a navigation platform, that allows to be programmed to bind to any locus within the genome including promoter regions. PAL shall be fused to the established transactivation domains of the SAM system p65 and HSF1. While in the dark, transcription activity is at base level, in the light, the binding of PAL to the sgRNA/Cas9 complex recruits transactivation domains, that drive transcriptional activity.

V Results

In this work, the focus is on the recently discovered photoreceptor PAL, which belongs to the LOV family photoreceptor proteins[166]. The photoreceptor protein was found in *Nakamurella multipartita* and consists of a Period-ARNT-Singleminded (PAS) domain, an AmiR and NasR transcription antitermination regulators (ANTAR) domain and the LOV domain, resulting in a size of 40 kDa. PAS domains are known as common sensor and interaction modules that facilitate signal transduction by structural and dynamic changes[167]. ANTAR domains are found in dimeric proteins that bind to RNA and modulate transcription termination[168, 169]. It was found, that the LOV domain of PAL – which structurally belongs to the PAS domains[170] - harbors a FMN chromophore and can perform a reversible conformational change upon irradiation with blue light (λ=447 nm), similar to other LOV proteins[171] **(Figure 13)**. The co-factor FMN provides the photosensitivity to blue light, while in darkness FMN exhibits a weak intrinsic green fluorescence[55, 172-174]. Upon irradiation with blue light, a thioether between the C4a atom of FMN and a conserved cysteine residue in the LOV core is formed[173, 175] and as a result, the changing hydrogen bonding pattern of the LOV domain triggers a conformational change of the whole protein[176-178], leading to a modulation of possible effector domains. At the same time, the formation of the flavin-cysteine adduct leads to a loss of intrinsic FMN fluorescence[179, 180]. Incubation in darkness terminates the formation of the thioether, recovers the initial conformation of the protein and restores the hydrogen bonding pattern of the LOV domain[179]. LOV domains were first discovered as the modules that mediate plant phototropism, the orientation of plants towards a source of light[181]. They were also identified in other organisms such as fungal and bacterial systems included in histidine kinases or transcription factors[182].

Since PAL is able to switch its conformation reversibly induced by blue light and the ANTAR domain might enable for RNA binding, the protein was used in a SELEX. Anna Maria Weber selected two aptamers (aptamer 04 and aptamer 53), which were characterized and that specifically bind the light conformation of the photoreceptor PAL with high affinity, while binding in darkness is strongly decreased **(Figure 13, B and C)**. The system is reversible and fully genetically encodable. In a reporter gene assay the aptamer 53 was inserted in the 5' UTR of a Metridia luciferase gene. Upon irradiation with blue light, the expression of a reporter gene was attenuated, while in darkness the expression was ~6 fold higher[166]. This novel aptamer/protein interaction opened the

door for a new field dubbed Optoribogenetics, which enables spatio-temporal control of RNA/Protein assembly. Selection and characterization of the PAL binding aptamers and the development of light-dependent control of translation as a proof of principle experiment using PAL and aptamer 53 are described in the dissertation of Anna Maria Weber.

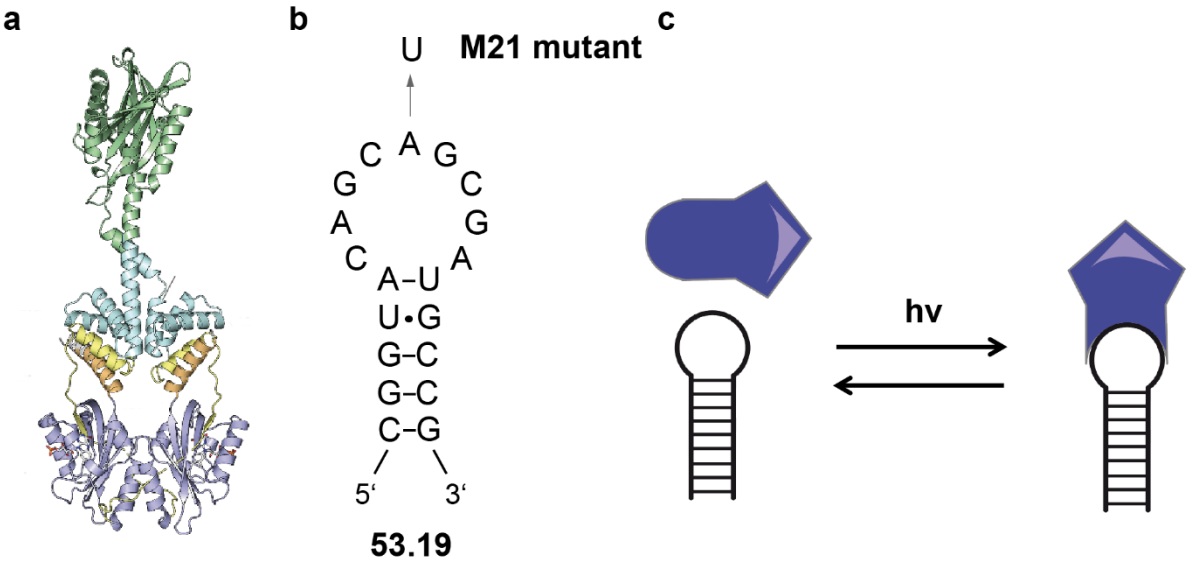


Figure 13: The photoreceptor protein PAL binds specifically to 53.19 aptamer in the light, whereas in the dark no binding was observed. A) Crystal structure of the PAL. B) The hairpin aptamer structure and sequence of 53.19 and the corresponding point mutant M21. C) Schematic of the light-dependent interaction of PAL with aptamer 53.19 (modified)[166].

V.1 Light-dependent Cas9 repulsion

One idea was to construct a light-dependent repulsion system, in which the photoreceptor protein PAL upon irradiation with blue light binds to aptamers, which are introduced in the stem-loops of the sgRNAs. The binding of PAL was believed to repulse bound Cas9 from the sgRNA. It was hypothesized, that incubated in blue light, a Cas9 cleavage reaction using aptamer modified sgRNAs and in the presence of PAL does not lead to cleavage products. Although, it was also elucidated, that wild type Cas9 binds to the sgRNA with an affinity of ~10 pM[183] and PAL binds to aptamer 53 in the light with an affinity of around 20 nM[166] the hypothesis was tested in a cleavage experiment. For the cleavage experiment Cas9 enzyme needs to be expressed, sgRNAs need to be transcribed and a cleavage substrate needs to be generated.

V.1.1 Cas9 production

In order to test modified sgRNAs for their activity in binding to Cas9 a simple way for testing is the Cas9 cleavage assay. Within the Cas9 cleavage assay a complex between Cas9 nuclease and sgRNA is formed. A DNA substrate is added to the Cas9:sgRNA complex, which is cut by the complex at the target site specified by the sgRNA. The Cas9 enzyme was produced by transforming pET-Cas9-6xHis plasmid into *E. coli* DE3 RIL **(VIII.15)**. For the expression the transformed bacteria were cultivated in liquid medium and the expressed protein was purified by Ni-NTA affinity chromatography **(VIII.11)**.

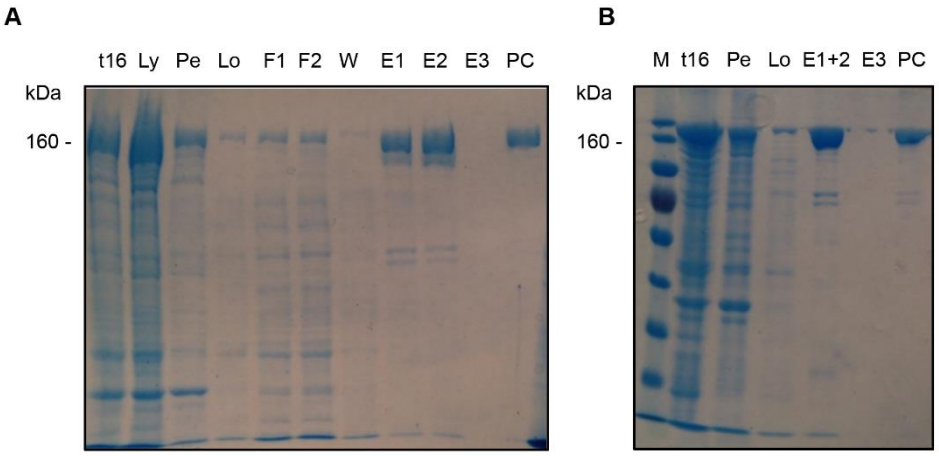


Figure 14: 10% SDS-PAGE of Cas9 large-scale expression. A) SDS-PAGE with samples from the Cas9 expression. T16: culture 16 hours after induction. Ly: Lysate after sonication. Pe: Cell pellet after sonication and centrifugation. Lo: Load. F1: Flowthrough fraction 1. F2: Flowthrough fraction 2. W: Wash fraction. E1: Elution fraction 1. E2: Elution fraction 2. E3: Elution fraction 3. PC: Positive control Cas9. B) M: Thermoscientific PageRuler Prestained Protein Ladder. T16: culture 16 hours after induction. Pe: Cell pellet after sonication and centrifugation. Lo: Load. E1+E2: Pooled elution fraction E1 and E2 from A). PC: Positive control Cas9. The samples were migrating for 60 min at 175 V and were stained using Coomassie Brilliant Blue.

The expression of Cas9 was successful **(Figure 14 A)**. A protein with a molecular weight of ~160 kDa was expressed, which migrates at the MW of the positive control **(Figure 14 A, lanes E1 and E2)**. The protein was visible 16 hours after induction. It was isolated in elution fractions 1 and 2, however in elution fraction 3 no protein was left on the Ni-NTA beads. The elution fractions were pooled and analyzed by western

blot. The gel from **Figure 14 B** was blotted on a membrane to verify Cas9 protein using anti-Cas9 antibody **(VIII.10)**. Blotting was performed using the semi-dry approach. Cas9 was detected within the western blot **(Figure 15, lanes E1+2, E3)** at the size of the positive control **(Figure 15, lane PC)**.

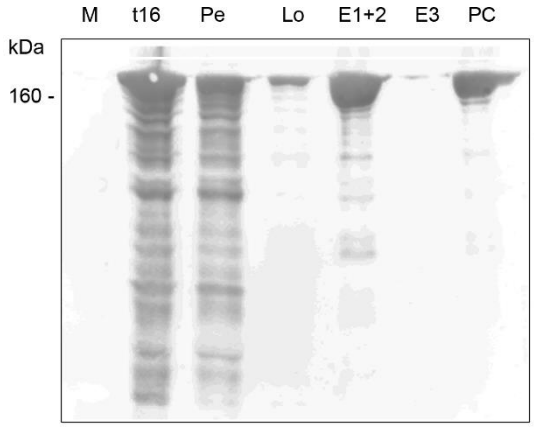


Figure 15: Western blot of Cas9 large-scale expression. M: Thermoscientific PageRuler Prestained Protein Ladder. T16: culture 16 hours after induction. Pe: Cell pellet after sonication and centrifugation. Lo: Load. E1+E2: Pooled elution fraction E1 and E2 from A). PC: Positive control Cas9. The samples were migrating for 60 min at 175 V and were afterwards stained using Coomassie Brilliant Blue.

Finally, the protein concentration was determined using the Bio-Rad protein assay and 1 mg/ml BSA standard **(VIII.9)**.

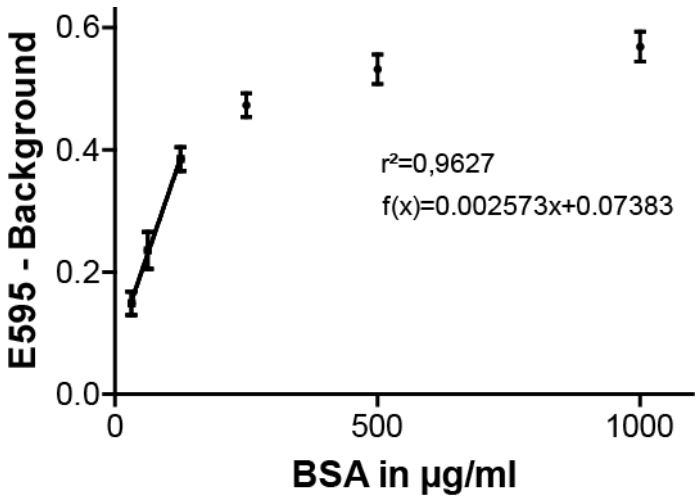


Figure 16: BioRad Proteinassay for the determination of Cas9 concentration.

The BSA standard was diluted from 100 µg/ml to 31.25 µg/ml by serial dilution. The linear range of the assay was measured and ranges up to 125 µg/ml **(Figure 16)**. Before measurement, the Cas9 samples were diluted 1:10. A protein concentration of 400 µg/ml was measured for the pooled elution fractions (1+2) from the large scale Cas9 production. Next, sgRNAs need to be designed and which then shall be used in the cleavage assays.

V.1.2 Design and generation of sgRNAs

In order to create a Cas9 complex, which is capable of cleaving a defined target DNA sequence, a sgRNA needs to be designed accordingly. The sgRNAs shall be used in the Cas9 cleavage assay since it is needed to form the active complex with the Cas9 enzyme. The mCherry protein is a fluorescent protein, which is derived from red fluorescent protein[184]. For a proof of concept, a target site within the coding sequence of the mCherry protein was chosen, which can easily be amplified by PCR for *in vitro* studies and would also allow for experiments within mammalian cells, in which the mCherry signal would be disrupted upon cleavage. A web tool named ChopChop[185] was used, which identifies potential Cas9 cleavage sites within DNA sequences and suggests sites based on their GC content and sequence uniqueness. The target sequence is depicted in yellow **(Figure 17)**, with the PAM motive in depicted blue

>mCherry

ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCG
AGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCA
GTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAAC
TTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACG
GCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCT
GAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATC
ACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAG

Figure 17: CDS of mCherry without stop codon. The target sequence is in yellow, followed by the PAM site in blue.

When designing Cas9 assays, the guide RNAs can either consists of a trRNA and a crRNA or a chimeric sgRNA[86]. In this work, sgRNAs were used, since only one expression cassette is needed and for *in vitro* assays the annealing step can be omitted. The seed sequence was fused to the native form of the sgRNA and was programmed to bind to the target sequence, if complexed with Cas9 **(Figure 18)**.

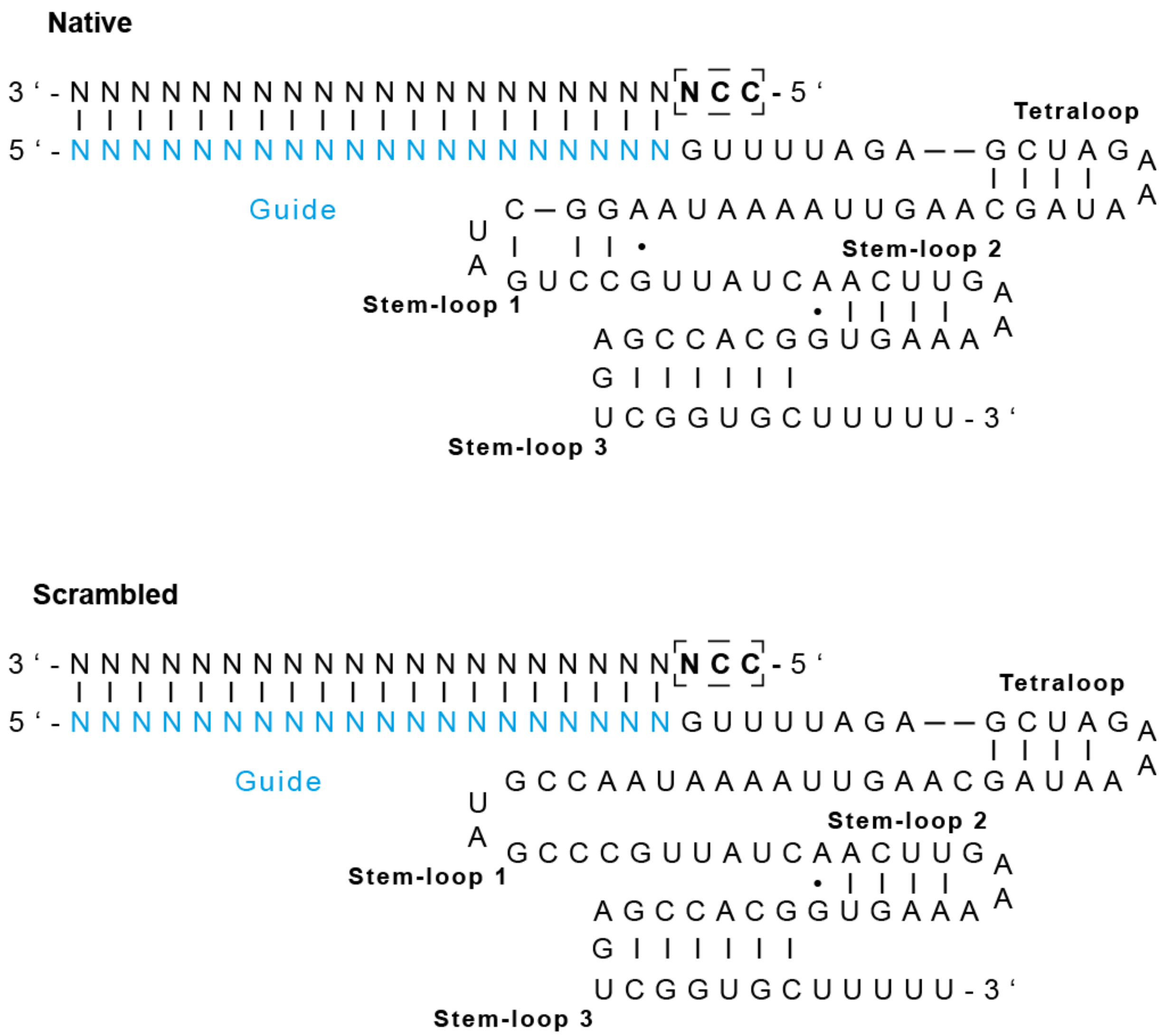


Figure 18: RNA sequences of mCherry CDS targeting native sgRNA (103 nt) and scrambled sgRNA (103 nt) having the stem-loop 1 scrambled.

The newly designed sgRNA were transcribed *in vitro*. The modified sgRNA templates were first amplified using PCR **(VIII.3.1)** as described[186]. The PCR products were visualized on a 4% agarose gel **(Figure 19)**.

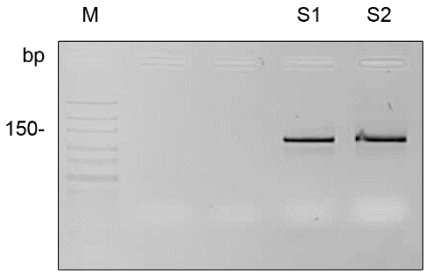


Figure 19: 4% agarose gel with sgRNA templates for *in vitro* transcription. M: Thermoscientific GeneRuler Low Range DNA Ladder. S1: native sgRNA v2 DNA amplicon at 127 bp. S2: scrambled sgRNA DNA amplicon at 127 bp. The samples were migrating for 15 min at 150 V and stained with ethidium bromide.

The PCR templates were then used in *in vitro* transcription reactions **(VIII.12)** and were afterwards purified **(VIII.13)**.

200 ng of the purified RNA was loaded on a 4% agarose gel to analyze the integrity.

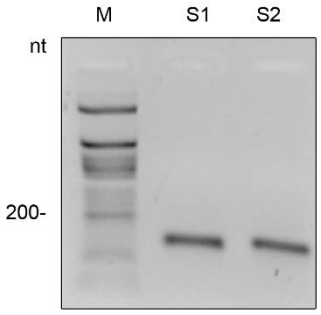


Figure 20: 4% agarose gel with 200 ng of purified sgRNA. M: NEB Low Range ssRNA Ladder (NEB, 0364). S1: native sgRNA, site targeting the mCherry target sequence at 103 nt. S2: scrambled sgRNA, at 103 nt. The samples were migrating for 15 min at 150V and stained with ethidium bromide.

Both sgRNAs were successfully transcribed and are migrating at the expected size **(Figure 20)**. The sgRNAs can be included in further cleavage assays. The concentrations were measured for native sgRNA targeting the mCherry target sequence and for the scrambled sgRNA. In order to test the transcribed sgRNAs in a Cas9 cleavage assay, a cleavable substrate needed to be generated.

V.1.3 Substrate generation

For the detection of Cas9 cleavage activity a DNA substrate is needed, which within the assay is cleaved into two fragments. During the first tests the sgRNAs were programmed to bind to a 20 nt sequence within the mCherry CDS. The substrate, which is used for cleavage contains the mCherry CDS. Two substrates were generated. First the plasmid pmCherry-C1 was digested using several restriction enzymes, rendering circular and supercoiled plasmid DNA to linear fragment. In a later stage, a 1400 bp amplicon containing the mCherry target site was used.

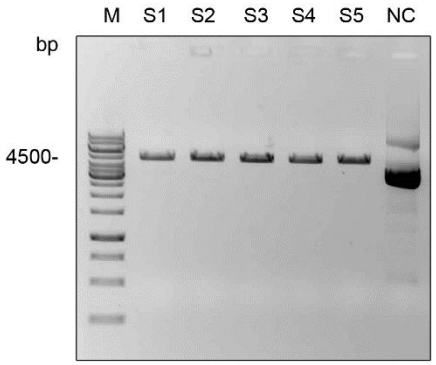


Figure 21: 1% agarose gel with digested and undigested pmCherry-C1 plasmid. M: Thermoscientific GeneRuler 1kb DNA Ladder. S1: *Xhoi* digest. S2: *SalI* digest. S3: *BglII* digest. S4: *BamHI* digest. S5: *NdeI* digest. NC: undigested plasmid. From each sample 5 µl containing 200 ng were loaded onto the gel. The samples were migrating for 30 min at 150 V and stained with ethidium bromide.

2 µg of pmCherry-C1 plasmid (provided by Anna Maria Weber) was digested using *XhoI*, *SalI*, *BglII*, *BamHI* or *NdeI* restriction enzyme. From the reaction 200 ng were analyzed by agarose gel electrophoresis **(Figure 21)**. For further Cas9 cleavage assays the *XhoI* linearized plasmid was used.

Later, and for better visualization, a 1400 bp amplicon was produced, which can be amplified using PCR. Within the cleavage assay, the substrate is cut asymmetrically into two fragments with a fragment size of 400 bp and 800 bp, respectively. As primers the mCherrySubstrate_fwd and the mCherrySubstrate_rev primer were used. As template 50 ng of the pmCherry-C1 plasmid was used. As primers the mCherrySubstrate_fwd and mCherrySubstrate_rev primers were used. The PCR was run using home-made Pfu polymerase in 100 µl volume. T_{ann} was 53 °C and Θ_{elong} was

168s. The PCR was running for 18 cycles **(VIII.3.1)**. After the PCR, 3 µl of the PCR product was loaded on a 1% agarose gel **(Figure 22)**.

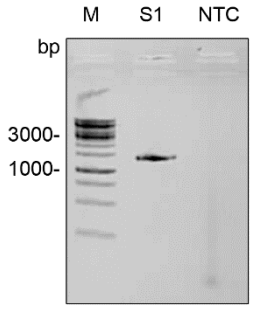


Figure 22: 1% agarose gel with 1400 bp PCR product of the mCherry CDS. M: Thermoscientific GeneRuler 1kb DNA Ladder. S1: mCherry substrate sequence at 1400 bp. NTC: no template control. 3 µl of PCR product were loaded on the gel. The samples were migrating for 20 min at 150 V and stained with ethidium bromide.

The 1400 bp mCherry substrate DNA was successfully amplified. The PCR product was purified **(VIII.6)**. Afterwards, the concentration was measured by spectrophotometric analysis **(VIII.5)**.

sgRNAs, Cas9 enzyme and substrate were then used in a Cas9 cleavage reaction.

V.1.4 Cas9 cleavage assay

To test, whether the expressed Cas9 is active both sgRNAs were tested in Cas9 cleavage assay including Cas9 and NEB Cas9 as positive control **(VIII.18)**. The 1400 bp mCherry amplicon was used as substrate and 2h of incubation time was set for the cleavage reaction. From each reaction 15 µl sample were loaded on an agarose gel **(Figure 23)**.

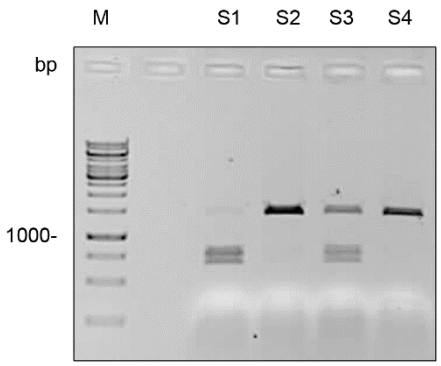


Figure 23: 1% agarose gel with cleavage products of Cas9 comparison. M: Thermoscientific GeneRuler 1kb DNA Ladder. S1: Cas9 and mCherry targeting native sgRNA. S2: Cas9 and scrambled sgRNA. S3: NEB Cas9 and mCherry targeting native sgRNA. S4: NEB Cas9 and scrambled sgRNA. The samples were migrating for 20 min at 130 V and stained with ethidium bromide.

The results from **Figure 23** indicate, that for expressed and NEB Cas9 in combination with mCherry targeting native sgRNA a cleavage occurs **(Figure 23, lanes S1 and S3),** as cleavage products appear at a size of 600 bp and 800 bp, respectively. For samples that contain scrambled sgRNA **(Figure 23, lanes S2 and S4)** no cleavage was detected, indicated by the full-length substrate at 1400 bp. These findings indicate, that the Cas9 is active.

In a next experiment, different buffers were tested for Cas9 activity. The photoreceptor PAL was tested for binding to its aptamers in intracellular Buffer (ICB). Thus, it had to be tested, whether Cas9 can cleave a substrate within ICB. The mCherry targeting native sgRNA was used in a cleavage assay as described above but testing its capability of cleaving the mCherry substrate also in Cas9 reaction buffer and ICB.

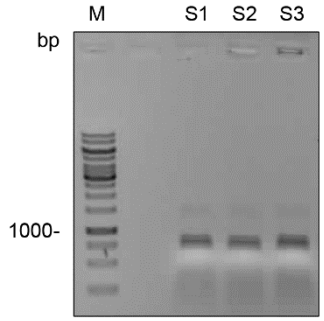


Figure 24: 0.8% agarose gel with cleavage products of Cas9 comparison. M: Thermoscientific GeneRuler 1kb DNA Ladder. S1: NEB Cas9 buffer. S2: Cas9 reaction buffer. S3: ICB. The samples were migrating for 20 min at 130V and stained with ethidium bromide.

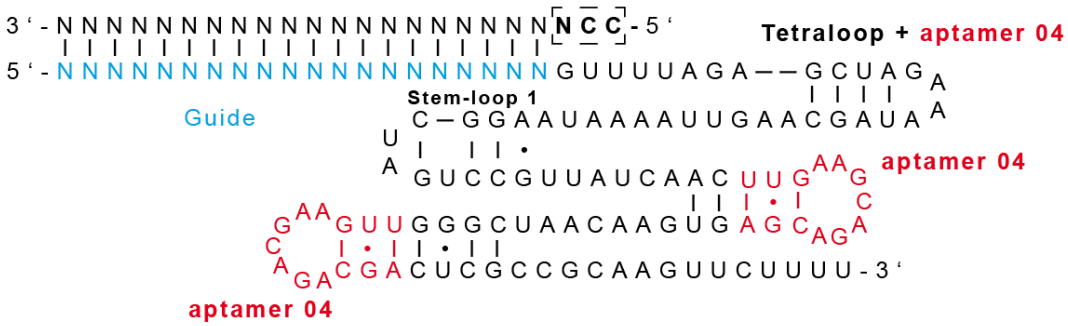
For all three buffers cleavage was observed **(Figure 24)**, indicating, that ICB is suitable for Cas9 cleavage assays, when PAL needs to be included.

In order to modulate Cas9 activity using PAL, the sgRNAs targeting the mCherry CDS were modified with PAL aptamer 04. Afterwards, the sgRNAs were *in vitro* transcribed to test their binding to Cas9 within a Cas9 cleavage assay.

V.1.5 Design and generation of sgRNAs with aptamer 04

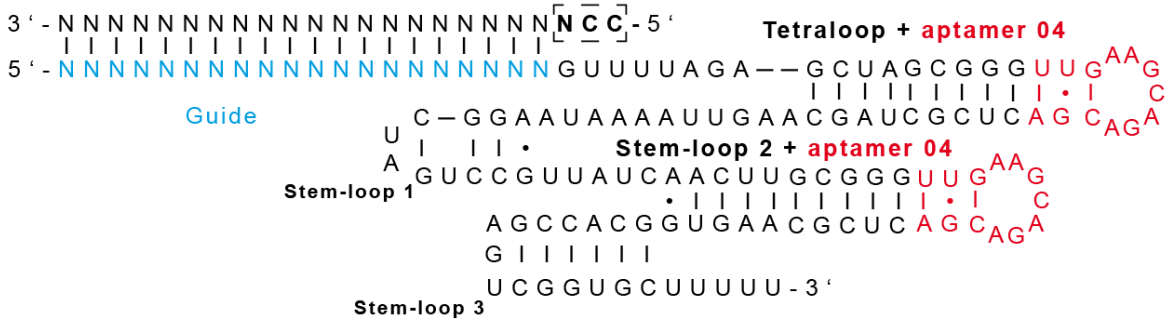
The following sgRNAs were designed together with Anna Maria Weber and contain the 04 PAL aptamer (5'- UUGAAGCAGACGA-3') and a non-binding variant (5'-*AACUUC*CAGACGA-3') **(Figure 25)**. Non-binding variants were designed for sg04.19, dubbed sg04.19dmu and for sg04.23, dubbed sg04.23dmu.

sg04.19di7



sg04.19tri

sg04.23di



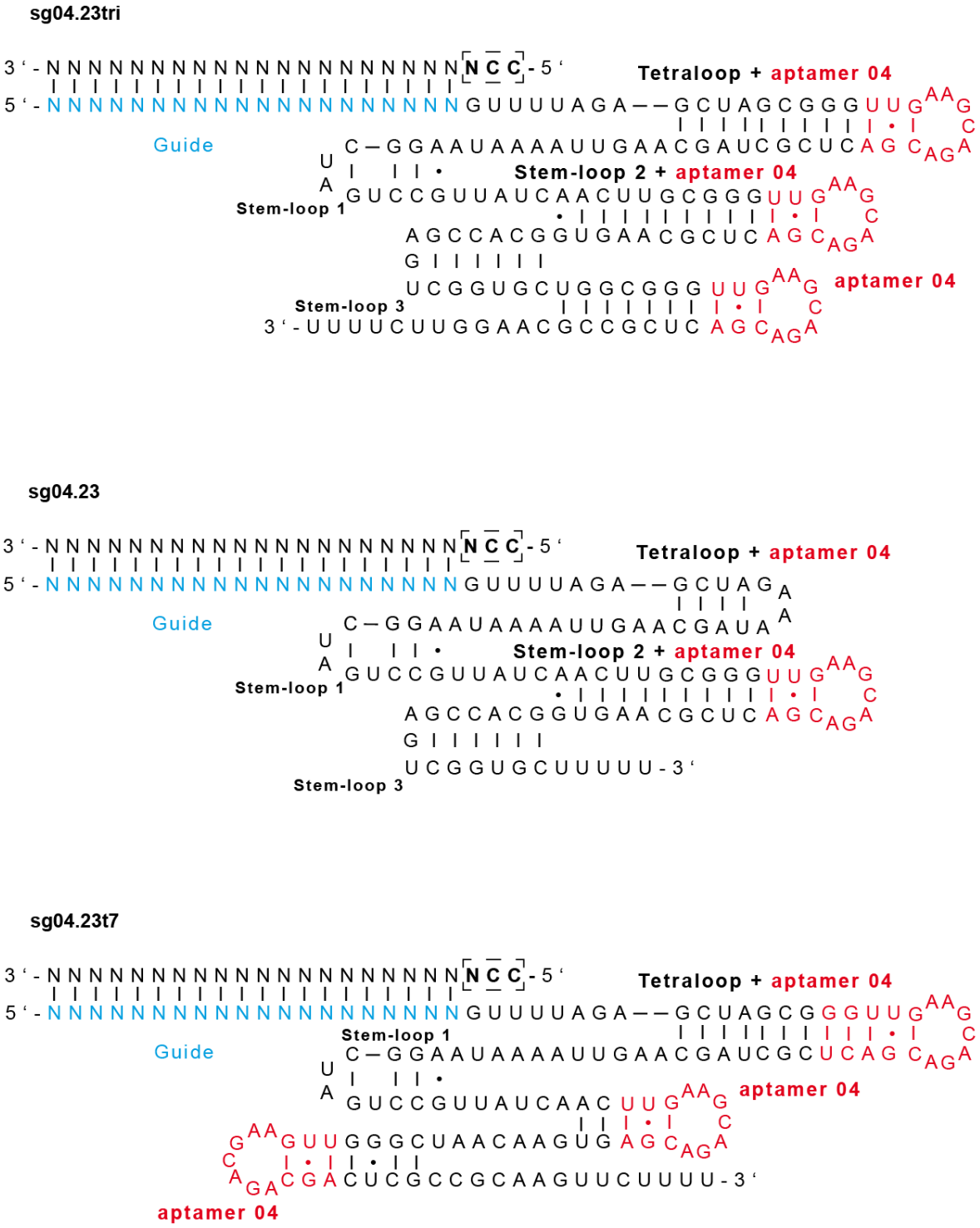


Figure 25: sgRNA variants containing the PAL aptamer 04 at various positions within the sgRNA scaffold. sgRNA variants sg04.19dmu and sg04.23dmu contain aptamer mutants, which should not bind to PAL.

In order to understand the aptamer insertions, the sgRNAs were depicted as schematic **(Figure 26)**.

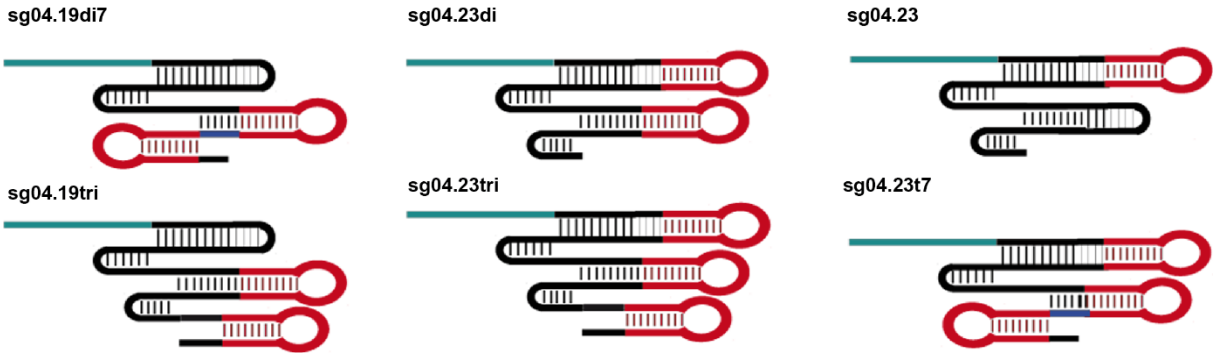


Figure 26: Compact schematic of sgRNAs modified with PAL aptamer 04 variants.

The sgRNA sg04.19di7 was designed using a PAL aptamer 19-mer dimer inserted in stem loop 2 by excluding stem loop 3. For sgRNA sg04.23di the 23-mer aptamer was used and inserted in the tetra-loop and stem loop 2, while keeping stem loop 1 and stem loop 3. In contrast, sgRNA sg04.23 has only a single 23-mer aptamer inserted in the tetra-loop, while keeping stem loops 1, 2 and 3 untouched. sgRNA sg04.19tri was modified by introducing the 19-mer variant into stem loop 2 and extending stem loop 3 with a 3' terminal 19-mer aptamer. Additionally, in sg04.23.tri the 23-mer aptamers were used and compared to sg04.19tri, another aptamer was inserted in the tetra-loop of the sgRNA. Finally, the sgRNA sg04.23t7 is a variant of sgRNA sg04.19di7, however another 23-mer variant was added in the tetra-loop **(Figure 26)**. All variants have in common, that stem loop 1 is unmodified, since it is crucial for binding to Cas9[89]. Once designed, the sgRNAs were transcribed *in vitro*.

The sgRNAs were PCR amplified from oligonucleotides to generate templates for *in vitro* transcription. The PCR was run using Pfu polymerase in 200 µl (2x100 µl) volume for each sgRNA. T_{ann} was 53 °C and Θ_{elong} was 30s. The PCR was running for 18 cycles **(VIII.3.1)**. The reverse primers were changed according to the sequence of the sgRNA **(Table 1)**.

Table 1: PAL aptamer modified sgRNAs, reverse primers and their sizes.

Index	Name	Rev Primer	Size (nt)
1	sg09.19di7	PAL_rev	126
2	sg09.19dmu	PAL_rev	126
3	sg09.19tri	PAL_rev	145
4	sg09.23di	native_rev	137
5	sg09.23tri	PAL_rev	171
6	sg09.23dmu	native_rev	137
7	sg09.23	native_rev	118
8	sg09.23t7	PAL_rev	145

After the PCR, 3 µl of amplicons were loaded on a 4 % agarose gel. The remaining PCR products were purified **(VIII.6)**.

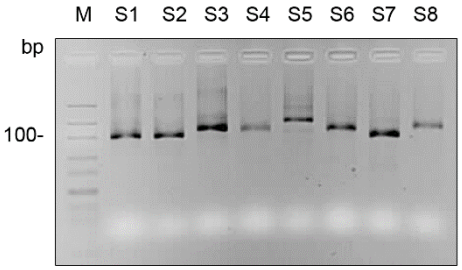


Figure 27: 4% agarose gel with sgRNA templates without PAM site for in vitro transcription. M: Thermoscientific GeneRuler Low Range DNA Ladder. S1: sg09.19di7 DNA template at 153 bp. S2: sg09.19dmu DNA template at 153 bp. S3: sg09.19tri DNA template at 172 bp. S4: sg09.23di DNA template at 164 bp. S5: sg09.23tri DNA template at 198 bp. S6: sg09.23dmu DNA template at 164 bp. S7: sg09.23 DNA template at 145 bp. S8: sg09.23t7 DNA template at 172 bp. The samples were migrating for 20 min at 150 V and stained with ethidium bromide.

All sgRNA DNA templates were successfully amplified and are at their expected size **(Figure 27)**. Some impurities were detected for S1, S2, S3 and S5.

The purified sgRNA DNA templates were used for transcription reactions to generate RNA **(VIII.12)**. 2 µl of purified RNA was loaded on a 4% agarose gel and to analyze the integrity.

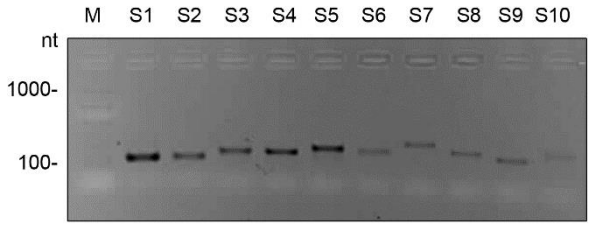


Figure 28: 4% agarose gel with in vitro transcribed sgRNAs modified with PAL aptamers. M: Thermoscientific RiboRuler Low Range ssRNA ladder. S1: native sgRNA targeting mCherry CDS at 103 nt. S2: non-binding sgRNA at 103 nt. S3: sg09.19di7 sgRNA at 126 nt. S4: sg09.19dmu sgRNA at 126 nt. S5: sg09.19tri sgRNA at 145 nt. S6: sg09.23di sgRNA at 137 nt. S7: sg09.23tri sgRNA at 171 nt. S8: sg09.23dmu sgRNA at 137 nt. S9: sg09.23 sgRNA at 118 nt. S10: sg09.23t7 sgRNA at 145 nt. The samples were migrating for 15 min at 150 V and stained with ethidium bromide.

All modified sgRNAs were successfully transcribed and identified regarding their size by agarose gel electrophoresis **(Figure 28)**. The transcribed sgRNAs were then used in a Cas9 cleavage assay to indirectly verify their complex formation with Cas9.

V.1.6 Cas9 cleavage assay with modified sgRNAs

The modified sgRNAs were tested in a Cas9 cleavage assay to indirectly analyze the influence of the Cas9 cleavage capability. Since Cas9 does not cleave DNA without complexed with a sgRNA (except for the presence of Mn^{2+} ions in the buffer, which were not measured)[187], sgRNAs that form weak or even no complex with Cas9 should result in weaker or the absence of cleavage. ICB was used as cleavage buffer, the molar ratio between Cas9 and substrate was 5:1 and between sgRNA and Cas9 2:1.

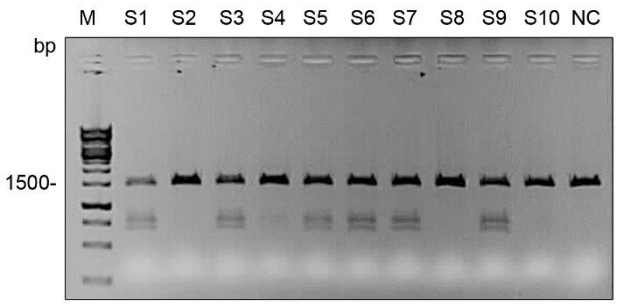


Figure 29: 1% agarose gel with products from the Cas9 cleavage assay using various sgRNAs modified with PAL aptamers. M: Thermoscientific GeneRuler 1kb DNA Ladder S1: Cas9 cleavage reaction with native sgRNA targeting mCherry CDS (positive control). S2: Cas9 cleavage reaction with non-binding sgRNA. S3: Cas9 cleavage reaction with sg04.19di7 sgRNA. S4: Cas9 cleavage reaction with sg04.19dmu sgRNA. S5: Cas9 cleavage reaction with sg04.19tri sgRNA. S6: Cas9 cleavage reaction with sg04.23di sgRNA. S7: Cas9 cleavage reaction with sg04.23tri sgRNA. S8: Cas9 cleavage reaction with sg04.23dmu sgRNA. S9: Cas9 cleavage reaction with sg04.23 sgRNA. S10: Cas9 cleavage reaction with sg04.23t7 sgRNA. NC: substrate only. The samples were migrating for 30 min at 130V and stained with ethidium bromide. The gel picture shown is one representative of two assays, where N=2 is shown in Suppl. Fig. 1.

The cleavage reaction was successful, indicated by the positive control in **Figure 29 lane S1** cutting the 1400 bp substrate into 400 bp and 800 bp fragments. The negative control, non-binding sgRNA in S2 does not induce cleavage. For the aptamer modified sgRNAs only sg09.19di7 (S3), sg09.19tri (S5), sg09.23di (S6), sg09.23tri (S7) and sg09.23 (S8) show cleavage products, qualifying for the light-dependent cleavage assays. Samples S4, S8 and S10 do not induce cleavage however a very small portion of cleavage products are visible for S4.

From a total of eight aptamer modified sgRNAs, five induce cleavage within a Cas9 cleavage assay, indicating that they form a complex with Cas9 and do not block nuclease activity of Cas9 enzyme. These candidates were then tested in a light-dependent Cas9 cleavage assay in order to identify sgRNAs, that enable the light-dependent modulation of Cas9 function in conjunction with PAL.

V.1.7 Light-dependent Cas9 cleavage assay with modified sgRNAs

Before PAL was included in the cleavage assay a few modifications had to be done. Since PAL is not stable at 37 °C *in vitro*, the assay incubation was performed at 25°C. Cas9 was shown to be still active at 25 °C[b]. Again, ICB was used as cleavage buffer, the molar ratio between Cas9 and substrate was increased to 10:1 to obtain more cleavage products and the molar ratio between sgRNA and Cas9 was 2:1. The molar ratio between Cas9 and PAL was 1:10, whereas the molar amount of substrate was set to 54.11 fmol. Each reaction was performed three times, where the two reactions including PAL were in the light or in the dark and one reaction was without PAL as control.

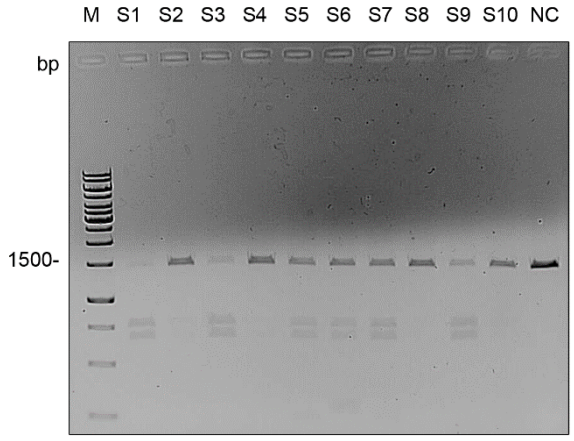


Figure 30: 1% control agarose gel with products from the Cas9 cleavage assay using various sgRNAs modified with PAL aptamers without PAL. M: Thermoscientific GeneRuler 1kb DNA Ladder S1: Cas9 cleavage reaction with native sgRNA targeting mCherry CDS (positive control). S2: Cas9 cleavage reaction with sg04.19di7 sgRNA. S3: Cas9 cleavage reaction with sg04.19dmu sgRNA. S4: Cas9 cleavage reaction with sg04.19tri sgRNA. S5: Cas9 cleavage reaction with sg04.23di sgRNA. S6: Cas9 cleavage reaction with sg04.23tri sgRNA. S7: Cas9 cleavage reaction with sg04.23dmu sgRNA. S8: Cas9 cleavage reaction with non-binding sgRNA. S9: Cas9 cleavage reaction with sg04.23 sgRNA. S10: Cas9 cleavage reaction with sg04.23t7 sgRNA. NC: substrate only. The samples were migrating for 30 min at 130 V and stained with ethidium bromide.

The control gel in **Figure 30** indicates that the increased ratios resulted in a full cleavage in the positive control (S1), whereas the negative control using a non-binding sgRNA (S8) does not induce cleavage, excluding that the concentrations used in the

[b] agilent.com/cs/library/usermanuals/public/5990-7262.pdf, 03.03.2017

assay induce unspecific cleavage. Cleavage was also found for sgRNAs sg04.19dmu, sg04.23di, sg04.23tri, sg04.23dmu and sg04.23.

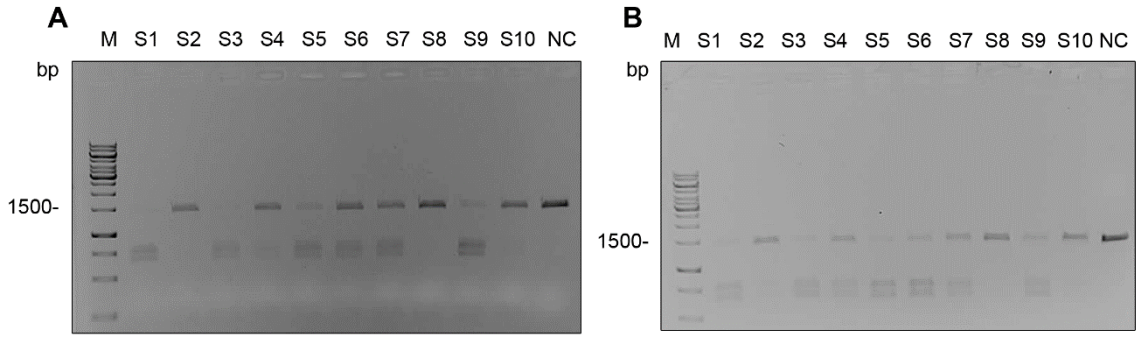


Figure 31: 1% agarose gels with products from the Cas9 cleavage assay incubated in the light/dark using various sgRNAs modified with PAL aptamers PAL. A.) M: Thermoscientific GeneRuler 1kb DNA Ladder S1: Cas9 cleavage reaction with native sgRNA targeting mCherry CDS (positive control) incubated in the light. S2: Cas9 cleavage reaction with sg04.19di7 sgRNA incubated in the light. S3: Cas9 cleavage reaction with sg04.19dmu sgRNA incubated in the light. S4: Cas9 cleavage reaction with sg04.19tri sgRNA incubated in the light. S5: Cas9 cleavage reaction with sg04.23di sgRNA incubated in the light. S6: Cas9 cleavage reaction with sg04.23tri sgRNA incubated in the light. S7: Cas9 cleavage reaction with sg04.23dmu sgRNA incubated in the light. S8: Cas9 cleavage reaction with non-binding sgRNA incubated in the light. S9: Cas9 cleavage reaction with sg04.23 sgRNA incubated in the light. S10: Cas9 cleavage reaction with sg04.23t7 sgRNA incubated in the light. NC: substrate only. B.) M: Thermoscientific GeneRuler 1kb DNA Ladder S1: Cas9 cleavage reaction with native sgRNA targeting mCherry CDS (positive control) incubated in the dark. S2: Cas9 cleavage reaction with sg04.19di7 sgRNA incubated in the dark. S3: Cas9 cleavage reaction with sg04.19dmu sgRNA incubated in the dark. S4: Cas9 cleavage reaction with sg04.19tri sgRNA incubated in the dark. S5: Cas9 cleavage reaction with sg04.23di sgRNA incubated in the dark. S6: Cas9 cleavage reaction with sg04.23tri sgRNA incubated in the dark. S7: Cas9 cleavage reaction with sg04.23dmu sgRNA incubated in the dark. S8: Cas9 cleavage reaction with non-binding sgRNA incubated in the dark. S9: Cas9 cleavage reaction with sg04.23 sgRNA incubated in the dark. S10: Cas9 cleavage reaction with sg04.23t7 sgRNA incubated in the dark. NC: substrate only. The samples were migrating for 30 min at 130V and stained with ethidium bromide.

The cleavage products of the light-dependent assay were analyzed by agarose gel electrophoresis. The cleavage reaction in the light **(Figure 31, A)** was successful for sgRNAs sg04.19dmu, sg04.23di, sg04.23tri, sg04.23dmu, sg04.23 and the positive control, whereas the negative control using a non-binding sgRNA (S8) does not induce cleavage. A slight enhancement in cleavage was detected for sgRNA sg04.19tri compared to the control gel in **Figure 30**. Cleavage was also found for sgRNAs sg04.19dmu, sg04.23di, sg04.23tri, sg04.23dmu and sg04.23. Compared to the cleavage assay in the light **(Figure 31, A)** the cleavage assay in the dark

(Figure 31, B) does not show visual differences, as the same cleavage pattern was detected there. In summary, no light-dependent cleavage was detected in this assay.

The further assessment of the idea was skipped and another approach was chosen, in which PAL and Cas9 can coexist on a sgRNA. PAL would recruit a biological function towards the sgRNA/Cas9 complex, which can be freely programmed to bind to user-defined genes. One idea was the development of a freely programmable and light-dependent transactivation system, using the direct interaction of PAL with the aptamers 04 or 53 in mammalian cells. The system was adapted from the MS2 system but enables control by blue light.

V.2 Light-dependent activation of transcription

It was hypothesized, that the extension of stem loops of the sgRNAs and the insertion of aptamer 53 might allow for PAL binding to the aptamers, while the sgRNA is in a complex with Cas9. This scenario was successfully used in the SAM system. The new approach would allow for light-dependent recruitment of transcriptional activators towards promoters to enable for control of transcription. First it was tested, that PAL binds the aptamers 04 and 53 in cells. For this, a pulldown assay was performed.

V.2.1 mCherryPAL pulldown assay

The binding of cellularly expressed mCherryPAL to immobilized aptamers 04.17 and 53.19 was tested in mammalian cells.

Aptamers 04.17 and a corresponding mutant 04.17 M2, as well as the aptamer 53.19 and the corresponding mutant 53.19 M27 were ordered as biotinylated RNA sequences **(XI.2.3).** HeLa cells were transfected with pmCherryPAL, pNLS-mCherryPAL or pmCherry-C1 plasmid **(VIII.19.1)** and the cell lysate was loaded on the immobilized RNA sequences similar to a pulldown assay **(VIII.19.10)**. Since PAL was known to bind to these sequences in the light and was tagged with a fluorescent protein, the mCherry signal was used as an indirect measurement of PAL binding to

the aptamers in a fluorescence plate reader. The assay was developed and performed in cooperation with Anna Maria Weber.

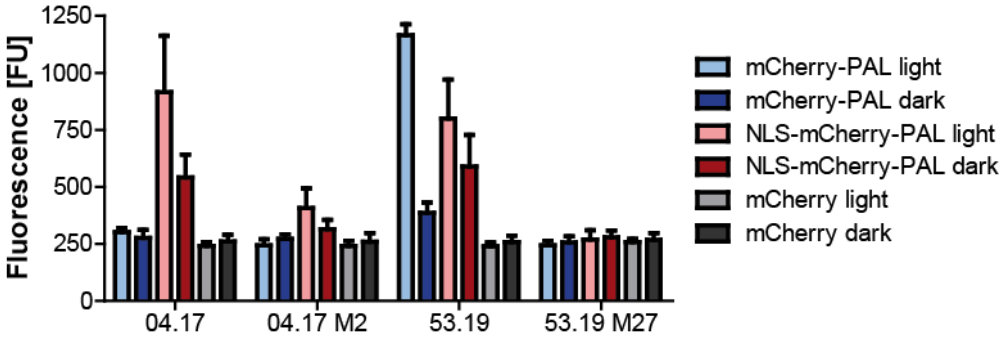


Figure 32: Recovered mCherry signal of the mCherryPAL pulldown assay. Biotinylated RNA aptamer variants were immobilized on plates. HeLa cell lysate containing mCherryPAL, NLSmCherryPAL or mCherry was added to the aptamers in the light or in the absence of light and after washing the recovered fluorescence intensity in the mCherry channel (Ex/Em: 587/610 nm) was measured. The data is presented as mean values ± s.d (n = 8 from 4 individual experiments with two cell culture replicates).

In **Figure 32**, a mCherry signal was only recovered for aptamer 53.19 in the presence of light but not for aptamer 04.17. For the NLS-tagged mCherryPAL variant mCherry fluorescence was recovered for both aptamers in the light and to a slightly smaller extend even in the dark. Point mutant 04.17 M2 only shows a weak signal for NLS-Cherry-PAL while the control 53.19 M27 shows no signal above background. Control mCherry was not detected by any sequence. Based on these results and since for aptamer 04.17 only a small light/dark fold-change was detected for the NLS-tagged mCherryPAL variant and no fold-change was detected for mCherryPAL, aptamer 53.19 was further used for the insertion in the sgRNAs.

V.2.2 Design of sgRNAs for transactivation of transcription experiments

Since the sgRNA 2.0 was shown to enable overexpression of endogenous genes using the SAM system[118] new sgRNAs were derived from sgRNA 2.0 and sgRNA 2.0 was used as positive control.

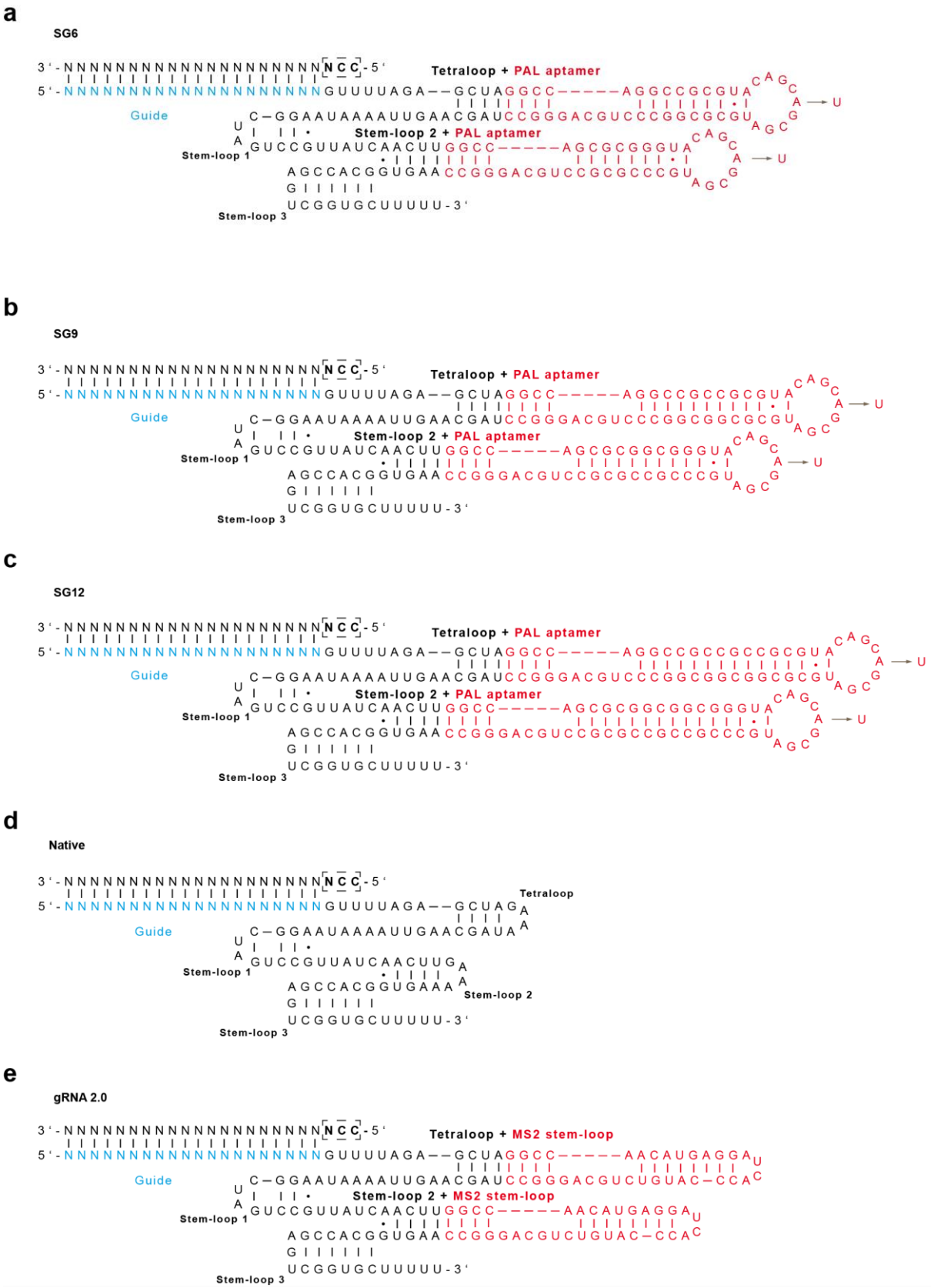


Figure 33: Design of sgRNAs. The sgRNAs are based on native sgRNA (control) and sgRNA 2.0 from which sgRNAs with PAL aptamer 53 (SG6, SG9 and SG12) were derived. a) sgRNA SG6 with a stem size of 6 nt. b) sgRNA SG9 with a stem size of 9 nt. c) sgRNA SG12 with a stem size of 12 nt. d) Native sgRNA without aptamers. e) sgRNA 2.0 from the SAM system.

The sgRNAs SG6, SG9 and SG12 were derived from sgRNA 2.0 by inserting the aptamers in the tetra-loop (TL) and stem loop 2 (SL2), while keeping parts of the MS2 aptamer stems and inserting new stem sequences **(Figure 33)**, defining their lengths and names. The introduced stem sequence for SG6 has a length of 6 nt, hence the sgRNA is dubbed SG6. SG9 and SG12 were designed, respectively. A better visualization of the insertion of the aptamers in the sgRNAs is given in **Figure 34**.

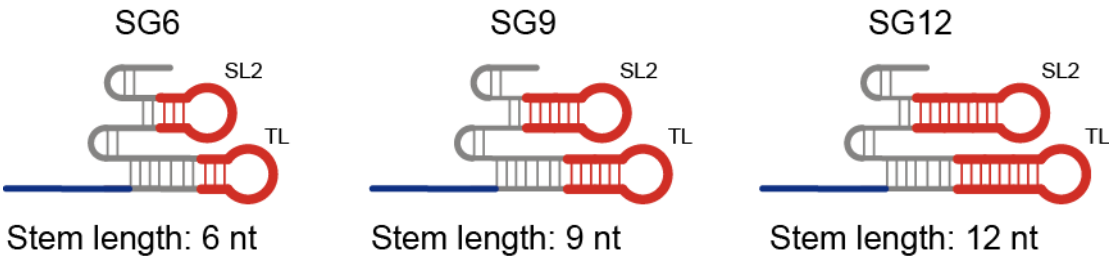


Figure 34: Schematic of the sgRNAs SG6, SG9 and SG12. The seed sequence is depicted in blue, while the aptamers are colored in red.

The three different stem sizes were chosen to avoid masking of the aptamer by the Cas9 protein, in case the aptamers are too short and not accessible for PAL. It was hypothesized, that the stem length might change the orientation of PAL and the effector domains towards the DNA, leading to a change in transcriptional activation efficiency. The seed sequences were either chosen to bind to the 8x gRNA binding site of pGL3-Basic-8x-gRNA (5'-AAAGGTCGAGAAACTGCAAA-3')[161] or as a non-binding negative control seed sequence (5'-AACAGTCGCGTTTGCGACT-3')[188]. The sgRNAs were tested for complex formation in a Cas9 cleavage assay.

V.2.3 Cas9 cleavage assay using new sgRNAs

The newly designed sgRNAs for light-dependent transcription activation (LITA) were tested in a Cas9 cleavage assay to verify their activity in complex with Cas9.

The sequences for the sgRNAs, targeting the 8xsgRNA binding site of pGL3-Basic-8x-sgRNA-eBFP were generated by hybridizing a sense and antisense DNA sequence by

subsequent PCR amplification using the sgRNA_T7_fwd and sgRNA_native_rev as described previously. The PCR products were visualized on agarose gel **(Figure 35)** and used for *in vitro* transcription to generate sgRNA.

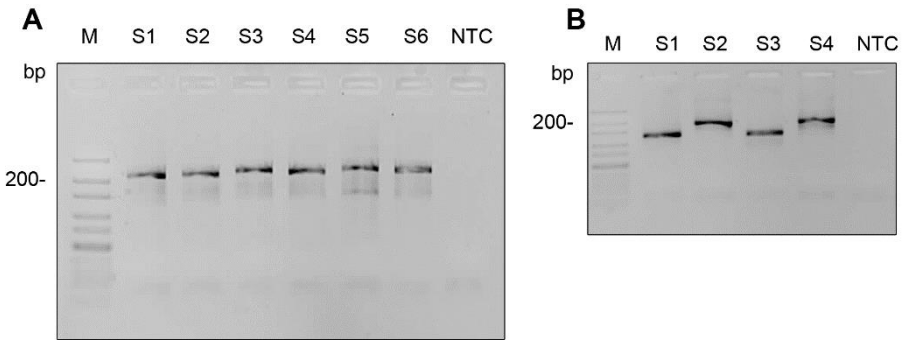


Figure 35: 4% agarose gels with PCR products of sgRNA templates for the Cas9 cleavage assay. M: Thermoscientific GeneRuler Ultra Low Range DNA Ladder. A) S1: Amplicon of sgRNA SG6 at 193 bp. S2: Amplicon of negative control sgRNA SG6 at 193 bp. S3: Amplicon of sgRNA SG9 at 205 bp. S4: Amplicon of negative control sgRNA SG9 at 205 bp. S5: Amplicon of sgRNA SG12 at 217 bp. S6: Amplicon of negative control sgRNA SG12 at 217 bp. NTC: no template control. B) S1: Amplicon native sgRNA at 108 bp. S2: Amplicon of sgRNA 2.0 at 168 bp. S3: Amplicon of negative control native sgRNA at 108 bp. S4: Amplicon of negative control sgRNA 2.0 at 168 bp. NTC: no template control. The samples were migrating for 10 min at 150 V and were stained with ethidium bromide.

The PCR products were purified and were used as templates for *in vitro* transcription **(VIII.12)**.

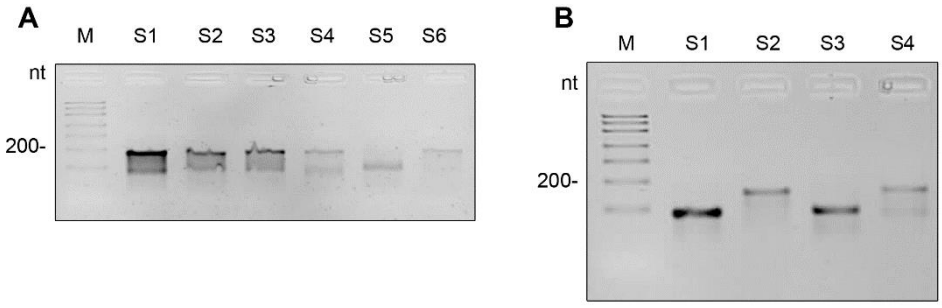


Figure 36: 4% agarose gels with *in vitro* transcribed sgRNA for the Cas9 cleavage assay. M: Thermoscientific RiboRuler Low Range RNA Ladder. A) S1: sgRNA SG6 at 166 nt. S2: Negative control sgRNA SG6 at 166 nt. S3: sgRNA SG9 at 178 nt. S4: negative control sgRNA SG9 at 178 nt. S5: sgRNA SG12 at 190 nt. S6: negative control sgRNA SG12 at 190 nt. B) S1: native sgRNA at 81 nt. S2: sgRNA 2.0 at 141 nt. S3: negative control native sgRNA at 81 nt. S4: negative control sgRNA 2.0 at 141 nt. The samples were migrating for 10 min at 150 V and were stained with ethidium bromide.

The sgRNAs were *in vitro* transcribed and analyzed by agarose gel electrophoresis. sgRNAs were generated for S1 to S4 and S6, however the amount of sgRNA for S5 was lower **(Figure 36, A)**. The control sgRNAs were successfully transcribed as depicted in **Figure 36 B**. The sgRNAs were then tested for binding to Cas9 in a Cas9 cleavage assay **(VIII.18).** The substrate was a 1360 bp PCR product of the pGL3-Basic-8xsgRNA-eGFP plasmid and was generated by Nemanja Stijepovic (LabBook Date 15.06.2018). The evaluation of the gel pictures was done using ImageJ software.

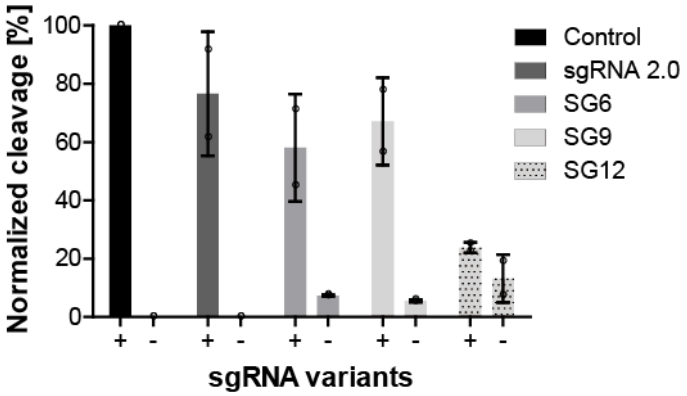


Figure 37: Normalized cleavage fraction of sgRNAs with and without aptamers. The new sgRNAs, which were designed incorporating the PAL-aptamer with varying stem sizes (SG6, SG9 and SG12) in the TL and SL2 accordingly to the design of sgRNA 2.0 were tested in a Cas9 cleavage assay. The data is presented as mean values ± s.d and was normalized to the native sgRNA (Control) using min-max scaling (n = 2 from 2 individual experiments). Raw image data in Supporting Figures 2.

The results from the Cas9 cleavage assay **(Figure 37)** indicate a reduced cleavage activity for all modified sgRNAs, whereas the sgRNA 2.0 achieves around 80% of cleavage activity and the sgRNA variants SG6 and SG9 achieve around 60% of cleavage activity. SG12 shows a strong reduction in cleavage activity which is around 20% compared to the native control sgRNA.

Based on the results it was found, that SG6 and SG9 are sgRNA candidates that form an active complex with Cas9, which is able to induce a cleavage reaction. Low activity was found for SG12. Next, the components for testing LITA in mammalian cells were cloned.

V.2.4 Light-dependent activation of transcription in mammalian cells

V.2.4.1 Light-dependent activation of eBFP and assay optimization

In order to test the PAL-dependent transcription activation system, a reporter system was generated, which allows for screening and optimization of the parameters, that modulate the transcription activation. Since plate-based secretable assays often require expensive reagents, for the beginning the fluorophore eBFP2 was used as a reporter in mammalian cells. The expression of a fluorescent protein enables cell analysis using flowcytometry. As plasmid backbone the published pGL3-Basic-8x-gRNA-eGFP[161] was used and amplified by inverse PCR **(VIII.14.6)**, excluding the eGFP insert. As primers the pGL3-Basic-8x-gRNA_fwd and pGL3-Basic-8x-gRNA_rev was used. For the PCR, Θ_{elong} was 60s and T_{ann} was 72°C and the PCR was running for 25 cycles. The eBFP insert was amplified from pU6-sgEx-Cas9-T2A-BFP with primer-extended overhangs for the insertion in pGL3-Basic-8x-gRNA **(VIII.3.1)**. As primers, the eBFP_fwd and eBFP_rev primers were used. For the PCR, Θ_{elong} was 20s and T_{ann} was 72°C and the PCR was running for 25 cycles. The PCR products of both reactions were identified on a 1% agarose gel **(Figure 38)**. The PCR products were purified **(VIII.6)**. The concentrations were measured for the eBFP insert as 27.9 ng/µl and for the pGL3-Basic-8xgRNA backbone as 23.1 ng/µl **(VIII.5)**.

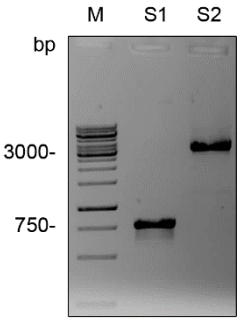


Figure 38: 1% agarose gel with PCR products of eBFP insert and linear plasmid backbone. M: Thermoscientific GeneRuler 1kb DNA Ladder. S1: eBFP insert running at 726 bp. S2: Linear plasmid backbone for eBFP insertion running at 3468 bp. The samples were migrating for 30 min at 130 V and stained with ethidium bromide.

The purified eBFP insert and plasmid backbone were used for an In-Fusion reaction **(VIII.14.1)** obtaining a clone with the correct sequence (SEQ_ID: 001).

The generated plasmid was used later on in the cellular assays as a reporter plasmid. Next, the transcriptional activators were cloned.

The published SAM system was used as a positive control in later transcriptional activation assays. The original plasmid, expressing the MS2-p65-HSF1-T2A-eGFP fusion protein was a lentiviral plasmid with a size of 11,409 bp. Since the transfection of large plasmids can be less effective the MS2-p65-HSF1-T2A-eGFP insert was cloned into the much smaller pEGFP-N1 backbone. The pEGFP-N1 backbone was also used to generate the PAL fusion protein, thus having both systems in the same plasmid backbone. First the pEGFP-N1 backbone was amplified by inverse PCR **(VIII.14.5)**, excluding the eGFP CDS. As primers, the pEGFP-N1_fwd and pEGFP-N1_rev primers were used. For the PCR, Θ_{elong} was 90s and T_{ann} was 72°C and the PCR was running for 25 cycles. The PCR products was identified on a 1% agarose gel **(Figure 39)**. The PCR product was purified **(VIII.6)**. The concentrations were measured for pEGFP-N1 backbone as 80.0 ng/µl **(VIII.5)**. The MS2-p65-HSF1-T2A-eGFP insert was amplified by Milena Mund using standard PCR and Pfu polymerase **(VIII.3.1)**. For the PCR, T_{ann} was 60°C, Θ_{elong} was 280s the final elongation time was also set to 280s and the PCR was running for 22 cycles. As primers the MS2-P1 and MS2-P2 primers were used.

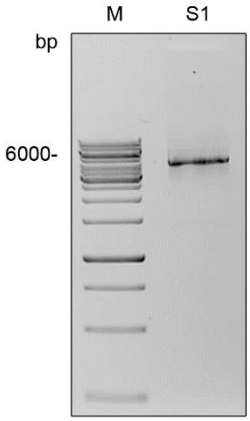


Figure 39: 1% agarose gel with PCR products of inverse PCR of pEGFP-N1. M: Thermoscientific GeneRuler 1kb DNA Ladder. S1: linear pEGFP-N1 backbone at 4022 bp. The samples were migrating for 30 min at 130 V and stained with ethidium bromide.

The insert of MS2-p65-HSF1-T2A-eGFP was successfully amplified and visible at 2259 bp on a 1% agarose gel **(Figure 40)**.

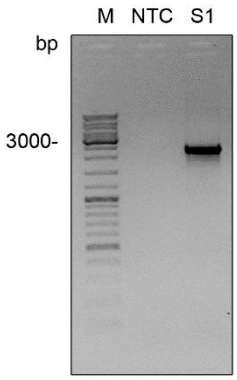


Figure 40: 1% agarose gel with PCR products of MS2-p65-HSF1-T2A-eGFP insert. M: DNA Leiter-Mix, peqGOLD. NTC: no template control S1: MS2-p65-HSF1-T2A-eGFP at 2259 bp. The samples were migrating for 30 min at 130 V and stained with ethidium bromide.

To clone the new plasmid pMS2-p65-HSF1-T2A-eGFP, both Fragments were used in a Gibson Assembly reaction **(VIII.14.3)**. Three clones were analyzed from which one clone was sequenced correctly (SEQ_ID: 002 - 003).

The generated plasmid encoding MS2-p65-HSF1-T2A-eGFP (MPH) shall be used as a positive control in cellular assays and to compare the PAL-based transcriptional activation to the published SAM system.

Since it is not known, that PAL exhibits transcriptional activation on its own, the established transcriptional activators p65 and HSF1 are fused to the photoreceptor protein. Due to the incorporated FMN molecule in the PAL protein, in darkness, PAL exhibits a weak dark fluorescence, which can be used to analyze the switching characteristics of the protein. In order to investigate the switching functionality of PAL, no eGFP should be co-expressed. Thus, mCherry was added as a transfection control, leading to the final plasmid pNLS-p65-HSF1-PAL-T2A-mCherry (PHP) and its variation pNLS-HSF1-p65-PAL-T2A-mCherry (HPP). The initial strategy to generate PHP was to split the fusion protein into three parts, that shall be spliced into one fragment. The fused full-length fragment then was to be inserted into the pEGFP-N1 backbone. The first fragment, NLS-p65-HSF1 (1041 bp), dubbed F1, was amplified using primers PAL-P1 and PAL-P2 in a standard Pfu PCR **(VIII.3.1)** and MS2-p65-HSF1_GFP plasmid as template. For the PCR, Θ_{elong} was 130s and T_{ann} was 60°C and the PCR was running

for 22 cycles. The second fragment, PAL (1158 bp), dubbed as F2, was amplified using the primers PAL-P3 and PAL-P4 also using standard Pfu PCR **(VIII.3.1)** and pmCherryPAL plasmid as template. For the PCR, Θ_{elong} was 145s and T_{ann} was 60°C and the PCR was running for 22 cycles. The PCR products were analyzed by agarose gel electrophoresis **(Figure 41, A)**. The third fragment, T2A-mCherry (780 bp), dubbed as F3, was amplified using primers T2A_mCherry_fwd and MS2-P2 in a Flash Fusion PCR reaction **(VIII.3.3)**, with pmCherry-C1 as template. For the PCR, Θ_{elong} was 60s and T_{ann} was 55°C and the PCR was running for 30 cycles. The PCR products of F3 were analyzed by agarose gel electrophoresis **(Figure 41, B)**.

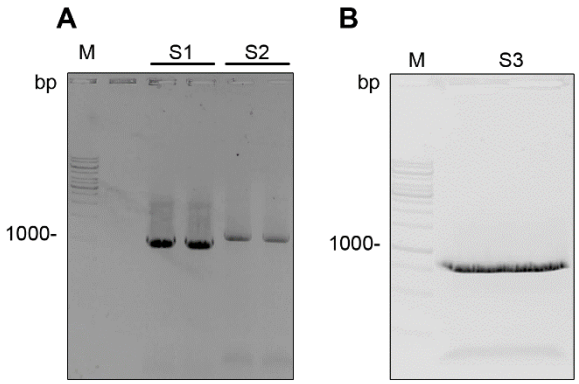


Figure 41: 1% agarose gel with PCR products of fragments F1, F2 and F3. M: Thermoscientific GeneRuler 1kb DNA Ladder A) S1: PCR products of F1 at 1041 bp. S2: PCR products of F2 at 1158 bp. B) S3: PCR products of F3 at 780 bp. The samples were migrating for 30 min at 130 V and stained with ethidium bromide.

For F1 in **Figure 41 A** S1 a PCR product was visible close to 1000 bp which is at the expected size of F1. For F2 in **Figure 41 A** S2 a slightly higher running PCR product compared to F1 was found, which in in line with the 1158 bp of F2. The third fragment was visualized in **Figure 41 B** between 1000 bp and 750 bp of the marker migrating at the expected size for F3 at 780 bp.

All three fragments were purified and the concentration of the fragments was determined using a photo spectrometer. In the next step the three fragments shall be spliced to a single 2925 bp fragment. In order to achieve this, Gibson Cloning was the first method of choice, since it already enabled the seamless cloning of the MS2-p65-HSF1-T2A-eGFP insert into the pEGFP-N1 backbone.

First the three fragments F1, F2 and F3 were incubated in a Gibson Assembly reaction **(VIII.14.3)**. From the final reaction a 1:10 dilution was made and used as template in a Flash Phusion PCR to amplify the full-length PHP insert **(VIII.3.3)**. For the Flash Phusion PCR the primers PAL-P1 and MS2-P2 were used. The PCR program was adjusted with Θ_{elong} as 60s and T_{ann} as 72°C running for 25 cycles. The Gibson Assembly reaction and the PCR products were loaded on a 1% agarose gel **(Figure 42)**.

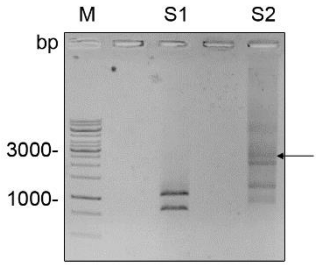


Figure 42: 1% agarose gel with reaction products of Gibson Assembly and PCR products of Flash Phusion PCR. M: Thermoscientific GeneRuler 1kb DNA Ladder. S1: Products of the Gibson Assembly reaction. S2: PCR products of the Flash Phusion PCR. The product at ~ 3000 bp, indicated by the arrow, was cut out. The samples were migrating for 25 min at 150 V and stained with ethidium bromide.

The Gibson Assembly reaction did not yield in a visible spliced full-length fragment **(Figure 42, S1)**. However, the initial fragments F1 and F2 are visible at 1000 bp and F3 a bit lower at 750 bp. After the amplification of Gibson Assembly products in a Flash Phusion PCR, multiple PCR products were visible **(Figure 42, S2)**. A product migrating at the size of ~ 3000 bp was cut out. However, it was noticed, that the concentration was too low for further usage in cloning.

In a new approach, 90 fmol of each fragment was mixed with 30 fmol of pEGFP-N1 backbone and incubated in another Gibson Assembly reaction **(VIII.14.3)** to synthesize not just the full-length fragment but the whole plasmid in one reaction. After the reaction, 3 µl were transformed into Stellar competent *E. coli* **(VIII.15)**. From twelve analyzed clones, none had the correct combination of inserts. Some clones were found to be background and some cloned were found to be re-circularized plasmid backbone.

Since Gibson Assembly did not result to the desired joined plasmid, another approach was performed using SOE-PCR **(VIII.14.4)**. As the fragments have overlapping homologous regions, within a PCR they can prime each other. To increase the chances of getting at least two fragments spliced, three reactions were set up using F1 + F2, F2 + F3 and F1 + F2 + F3. For the splicing reaction of F1 + F2 the primers PAL-P1 and PAL-P4 were used. For the splicing reaction of F2 + F3 the primers PAL-P3 and MS2-P2 were used. For the splicing reaction of F1, F2 and F3 the primers PAL-P1 and MS2-P2 were used. The amounts of fragments for the reaction were 74 fmol for F1, 66 fmol for F2 and 102 fmol for F3. For the PCR, Θ_{elong} was 60s and T_{ann} was 60°C and the PCR was running for 25 cycles.

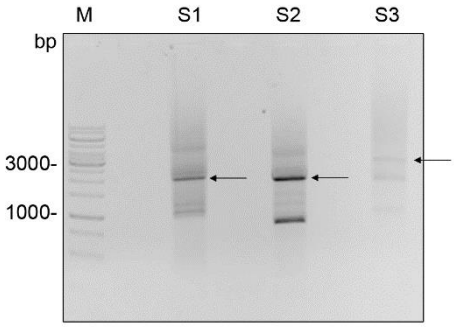


Figure 43: 1% agarose gel with reaction products of the SOE-PCR. M: ThermoScientific GeneRuler 1kb DNA Ladder. S1: SOE-PCR products using F1 + F2. S2: SOE-PCR products using F2 + F3. S3: SOE-PCR products using F1 + F2 + F3. The samples were migrating for 25 min at 140 V and stained with ethidium bromide.

The products of the SOE-PCR were analyzed by agarose gel electrophoresis **(Figure 43)**. For the first reaction of F1 and F2 a product migrating at ~2000 bp appeared **(Figure 43, S1, black arrow)**, which runs at the expected size for F1 + F2. The product was isolated from the gel. Also, further PCR products were obtained, which were discarded. For the second reaction **(Figure 43, S2)** also a product appeared at ~2000 bp but migrating a bit lower compared to S1. Also, a product at ~1000 bp was detected. The product at ~ 2000 bp was cut out **(Figure 43, S2, black arrow)**. For the third reaction multiple products were detected of which one was migrating at the expected size of ~3000 bp and was cut out **(Figure 43, S3, black arrow)**.

Since the concentration of the isolated PCR product of **Figure 43** S3 was too low, the full-length product could not be used for a cloning reaction. Then, another approach named AQUA cloning **(VIII.14.2)**, also called *in vivo* assembly[189, 190], was used. The three fragments are transformed together with the linear plasmid backbone into competent *E. coli* and the molecular machinery of the host organism assembles the fragments based on homology directed repair. For the AQUA cloning of PHP, the three fragments F1, F2 and F3 were mixed together with 60 fmol of each fragment. 20 fmol pEGFP-N1 backbone was added to the reaction. Clones were analyzed by sanger sequencing using the sequencing primers CMV-Fwd and pEGFP_C2-RP **(VIII.16)**. Two clones were sequenced correctly, from which clone 3 was used for further studies (SEQ_ID: 004 - 005). Before testing LITA in cells, the PAL fusion protein was expressed test wise in HeLa cells and the switching of the fusion protein was tested by microscopy analysis **(VIII.19.2)**. Since PAL fluorescence is detectable in the dark but not in the light conformation of the protein it was hypothesized that intact PAL should follow this characteristic.

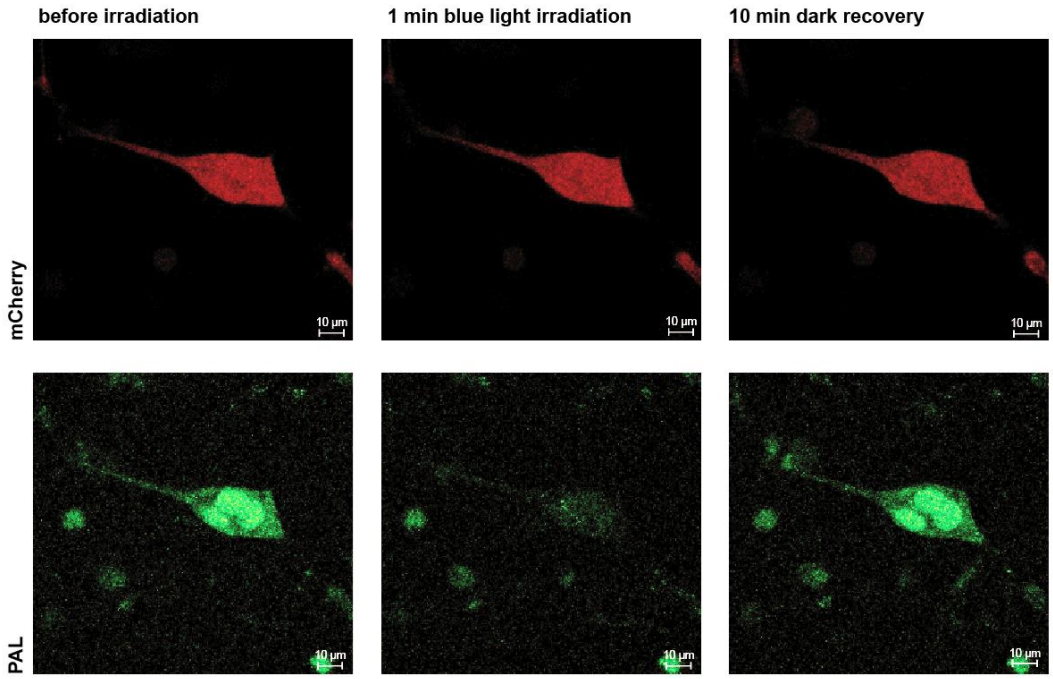


Figure 44: Switching activity of PHP in HeLa cells. Images were taken in the mCherry channel (ex/em: 543 nm/596-696 nm) and for the visualization of FMN visible in the PAL dark conformation ex/em: 405 nm/488-529 nm. First an image was taken when cells were kept for at least 10 min in the dark. Then, the cells were irradiated with blue light to change the conformation of PAL, which leads to a loss of PAL fluorescence. After 10 min recovery in the dark, PAL switches back to the fluorescent dark conformation indicating that the fusion protein can undergo a complete photo cycle. Scale bar = 10 µm.

PAL dark fluorescence based on the FMN chromophore was detected before irradiation and after 10 min of dark recovery **(Figure 44)**, whereas very little fluorescence was detected after 1 min irradiation with 465 nm. PAL was found to be mainly located in the nucleus of the cell. The mCherry signal is ubiquitous distributed in the cell and is not visually affected by the irradiation sequence. It was concluded, that the fusion protein is able to switch its conformation under blue light and is also able to switch back to the conformation in the dark.

The generated plasmids PHP and HPP express the PAL fused to transactivation domains. These plasmids are crucial for the light-dependent upregulation of target genes in mammalian cells. Next, the dCas9 expressing plasmid was modified.

In order to keep the assay design close to the published SAM system[118], the same dCas9 plasmid was used as in they used in their studies. However, the transcription activator VP64 is permanently fused to the dCas9 protein, which could cause unwanted activation of transcription in the dark. Thus, VP64 had to be removed from the dCas9VP64 plasmid, so VP64 will not constantly activate the gene, the complex is programmed to. The excision of VP64 from the plasmid was performed by inverse PCR cloning **(VIII.14.5)** using dCas9_-VP64_fwd and dCas9_-VP64_rev as primers and the dCas9VP64_eGFP as template plasmid. The PCR was running for 30 cycles as a 2-step PCR with 72°C of elongation, where Θ_{elong} was set to 6 min 30 sec. The PCR products were analyzed by agarose gel electrophoresis **(Figure 45, S1)**.

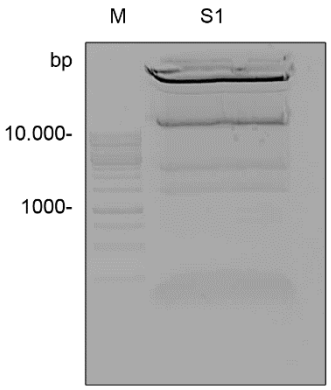


Figure 45: 1% agarose gels with reaction products of the inverse PCR of the dCas9 plasmid. M: Thermoscientific GeneRuler 1kb DNA Ladder. S1: PCR products of the inverse PCR of PHP at 14379 bp. The samples were migrating for 25 min at 130 V and stained with ethidium bromide.

Multiple PCR products appeared on the gel from which the product running close to 10,000 bp was isolated and purified by DNA purification kit **(VIII.6)**. After cloning, a total of 4 clones were sequenced from which 3 sequences matched the predicted sequence (SEQ_ID: 012). The generated plasmid expressing dCas9 without VP64 was used for the light-dependent assays. Next, the plasmids expressing the sgRNAs need to be cloned.

In order to apply the LITA assays, sgRNAs need to be co-expressed to direct dCas9 to a specific promoter region. For a proof of principle study, the sgRNAs were programmed to bind to the 8xsgRNA binding site upstream of the previously generated pGL3-Basic-8xsgRNA-eBFP plasmid, which contains 8 copies of the 5'-AAAGGTCGAGAAACTGCAAA-3' artificial binding sequence. The designed sgRNAs with varying stem-sizes were introduced into the pENTR.hU6.empty plasmid, which was kindly provided by Georg Pietruschka. AQUA cloning was used to insert the previously hybridized sgRNA DNA sequences into the pENTR plasmid backbone **(VIII.14.2)**. The linear pENTR backbone was generated by inverse PCR **(VIII.14.5)** using pENTR.hU6-F_BB and pENTR.hU6-R_BB as primers. The sgRNA sense and anti-sense DNA strands were first mixed and then were amplified using the terminal primers eBFP_sgRNA_in_pENTR_fwd and eBFP_sgRNA_in_pENTR_rev in a standard Pfu PCR. Prior to the PCR the sgRNA templates were diluted to 2-8 ng/µl and then of each sgRNA 1 µl template dilution was used in a 100 µl PCR. The PCR was running for 26 cycles. Θ_{elong} was 50s and T_{ann} was 63°C. The PCR products were analyzed by agarose gel electrophoresis.

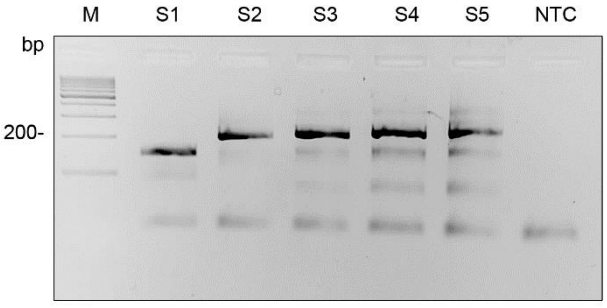


Figure 46: 4% agarose gel with sgRNA inserts targeting the 8x sgRNA binding site in pGL3-Basic-8x-sgRNA-eBFP. M: Thermoscientific GeneRuler 100 bp DNA Ladder. S1: native sgRNA (control) at 148 bp. S2: sgRNA 2.0 at 208 bp. S3: sgRNA SG6 at 224 bp. S4: sgRNA SG9 at 236 bp. S5: sgRNA SG12 at 248 bp. NTC: no template control. The samples were migrating for 20 min at 150 V and stained with ethidium bromide.

In **Figure 46**, the PCR products of the amplified full-length sgRNA inserts are analyzed by agarose gel electrophoresis. The sgRNAs were successfully amplified and not all of the sense and antisense strands were hybridized as these appear as lower products on the gel. Primers were visible as the lowest signal in each lane. The sgRNA inserts were purified **(VIII.6)**. The pENTR.hU6 plasmid backbone was amplified using a Θ_{elong} of 60s and T_{ann} was 68°C for 5 seconds and a total of 28 PCR cycles **(VIII.3.3)**. The pENTR.hU6 plasmid backbone was prepared according to the inverse PCR protocol **(VIII.14.5)**. The PCR products were visualized on a 1% agarose gel **(Figure 47)**.

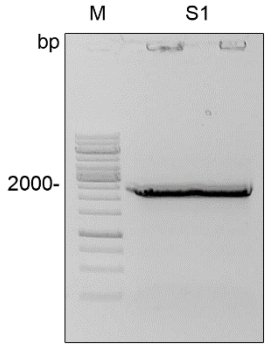


Figure 47: 1% agarose gel with PCR products of the inverse PCR of pENTR.hU6. M: Thermoscientific GeneRuler 1kb DNA Ladder. S1: linear pENTR.hU6 plasmid backbone at 2525 bp. The samples were migrating for 20 min at 150 V and stained with ethidium bromide.

The linear pENTR.hU6 plasmid backbone was successfully amplified **(Figure 47, S1)**. The product was isolated and purified.

The sgRNA inserts and the pENTR.hU6 plasmid backbone were used for the AQUA cloning approach. For the AQUA cloning reactions 90 ng of pENTR.hU6 plasmid backbone were used with a 10-fold molar excess of sgRNA inserts **(VIII.14.2)**. Clones were obtained for all sgRNA variants but SG12 (SEQ_ID: 013 - 017). For SG12 a clone (clone 7) with a point mutation was detected.

A Flash Phusion inverse PCR **(VIII.14.5)** was used to correct the mutation. For the PCR the primers M2S12_mut_fwd and M2S12_mut_rev was used. The pENTR.hU6.sgRNA.CMV.SG12 clone 7 plasmid backbone was amplified using a Θ_{elong} of 60s in a 2-step PCR with 27 PCR cycles. The PCR products were visualized on a 0.8% agarose gel **(Figure 48)**.

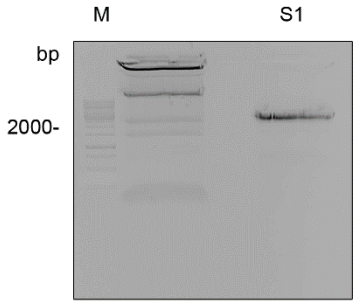


Figure 48: 0.8% agarose gel with PCR products of the inverse PCR of pENTR.hU6.sgRNA.CMV.SG12. M: Thermo Scientific GeneRuler 1kb DNA Ladder. S1: linear pENTR.hU6.sgRNA.CMV.SG12 plasmid backbone at 2729 bp. The samples were migrating for 25 min at 130 V and stained with ethidium bromide.

A PCR product appeared at the correct size of 2729 bp. The product was cut out and purified **(VIII.6)**. The linear plasmid was then further processed by inverse PCR cloning **(VIII.14.5)**. Sequencing revealed a correct clone for pENTR.hU6.sgRNA.CMV.SG12 (SEQ_ID: 018).

The generation of the point mutants was done as described above using a hybridized sense and anti-sense strand of the sgRNAs, which was amplified and cloned into the pENTR.hU6 backbone by AQUA cloning **(VIII.14.2)**.

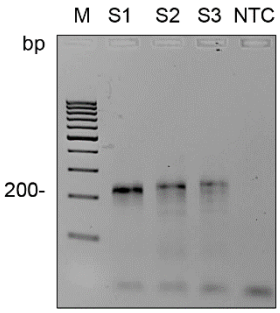


Figure 49: 4% agarose gel with negative control sgRNA inserts targeting the 8x sgRNA binding site in pGL3-Basic-8x-sgRNA-eBFP. M: Thermoscientific GeneRuler 100 bp DNA Ladder. S1: sgRNA SG6M21 at 224 bp. S2: sgRNA SG9M21 at 236 bp. S3: sgRNA SG12M21 at 248 bp. NTC: no template control. The samples were migrating for 20 min at 150 V and stained with ethidium bromide.

In **Figure 49**, the PCR products of the sgRNA inserts for the M21 point mutations are visualized by agarose gel electrophoresis. For all three sgRNAs a product at the expected size appeared. The PCR products were purified **(VIII.6)** and used for AQUA cloning **(VIII.14.2)**. For each construct a correct clone was sequenced (SEQ_ID: 019 - 021).

Since the sgRNAs were indirectly shown to bind to Cas9 and the PAL fusion protein is actively switching when expressed in HeLa cells, in the next step, LITA shall be tested within mammalian cell culture using HeLa cells. For a first assay the SAM system was used as a positive control. SG6 was tested first and also intermixed with activators p65-HSF1-PAL (PHP) and MS2-p65-HSF1 (MPH) with the sgRNAs, to investigate possible cross activation.

HeLa cells were transfected **(VIII.19.1)** using fixed amounts of ng plasmid and were incubated in a dark or a lit cell culture incubator. After 24 hours, the cells were analyzed by flow-cytometry **(VIII.19.5) (Figure 50)**.

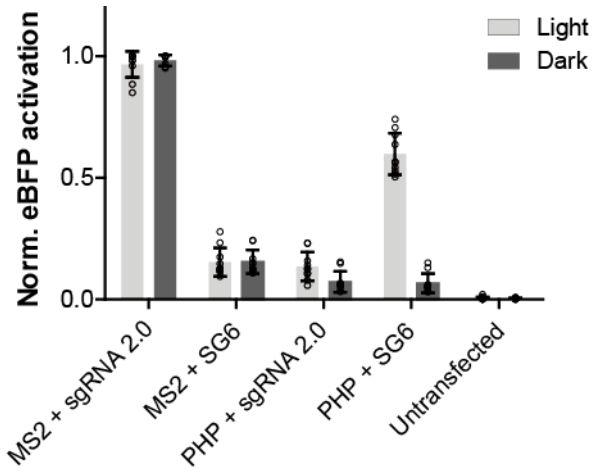


Figure 50: Normalized light-induced activation of eBFP expression measured by flow cytometry. HeLa cells were transfected with dCas9, MPH or PHP, eBFP reporter and sgRNA targeting the 8x gRNA binding sites upstream of the eBFP promoter. The incubation was for 24 hours in a dark or an illuminated cell culture incubator. The data is presented as mean values ± s.d and was normalized to gRNA 2.0 control using min-max scaling (n = 10 from 5 individual experiments with two cell culture replicates).

As expected, neither in the dark nor in the light the positive control MS2 + sgRNA 2.0 was reduced and shows in both setups the same amount of activation. When SG6 was used in combination with the MPH effectors only a light-independent residual activation was found. Vice versa, using the PHP activators combined with sgRNA 2.0 residual activation was observed with negligible difference between light and dark. When PHP was used in combination with SG6 ~60% of the activation compared to MS2 + sgRNA 2.0 was achieved in the light, whereas no activation was observed in the dark **(Figure 50)**. Since light-dependent activation of reporter gene expression was found

for SG6, the system was proven to function. Based on this result the system was optimized.

In a next step, the plasmid ratios between dCas9, sgRNA, activator and reporter plasmid were optimized. It was hypothesized, that the different plasmids have varying transfection efficiencies and also the promoters expressing the gens have varying strengths. Molar ratios between the plasmids were varied, whereas the setup above was used as a positive control.

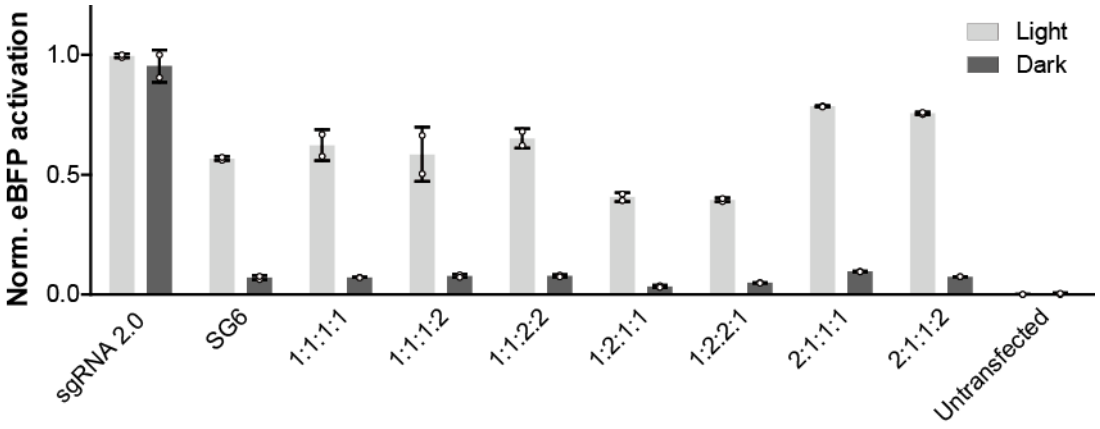


Figure 51: Normalized light-induced activation of eBFP expression using varying plasmid ratios measured by flow cytometry. HeLa cells were transfected with different ratios of dCas9, MPH or PHP, eBFP reporter and sgRNA targeting the 8x sgRNA binding sites within the pGL3-Basic-8xsgRNA-eBFP reporter plasmid. Ratios are given as dCas9:activator:reporter:sgRNA. The data is presented as mean values ± s.d and was normalized to the sgRNA 2.0 control using min-max scaling (n = 6 from 3 individual experiments with two cell culture replicates).

The variation of molar amount of plasmid influences the characteristics of the dynamic range of LITA. Again, sgRNA 2.0 was used as light-independent positive control and was shown to activate eBFP expression the most **(Figure 51)**. The results from the previous experiment regarding the activation using SG6 and PHP were confirmed. Plasmid ratios of 1:1:1:1, 1:1:1:2 and 1:1:2:2 show similar activation compared to the SG6 control, whereas ratios of 1:2:1:1 and 1:2:2:1 show a smaller dynamic range when comparing the expression of eBFP in the light and dark. However, ratios of 2:1:1:1 and 2:1:1:2 led to increased dynamic range compared to the previous setup SG6, with a ratio of 2:1:1:1 resulting in the highest dynamic range **(Figure 51)**. Thus, the plasmid ratio of 2:1:1:1 was used for further assays.

Next, the impact of the stem-length of the aptamers on the amount of eBFP activation was analyzed. The aptamers, which are integrated in the tetra-loop and stem loop 2 of the sgRNAs were introduced with stem-lengths of 6 nt (SG6), 9 nt (SG9) and 12 nt (SG12), which lead to a turn in the aptamer loop region and might change the orientation of the effectors towards the promoter and molecular interaction partners. As further control, aptamer point mutant M21 was included for each stem-length. The M21 mutation was introduced in both aptamers (tetraloop and stem-loop 2), hence further dubbed M21TL+SL2, or just simply M21. HeLa cells were transfected with the LITA components and eBFP as reporter plasmid **(VIII.19.5)**. After 24 hours the cells were analyzed by flow cytometry.

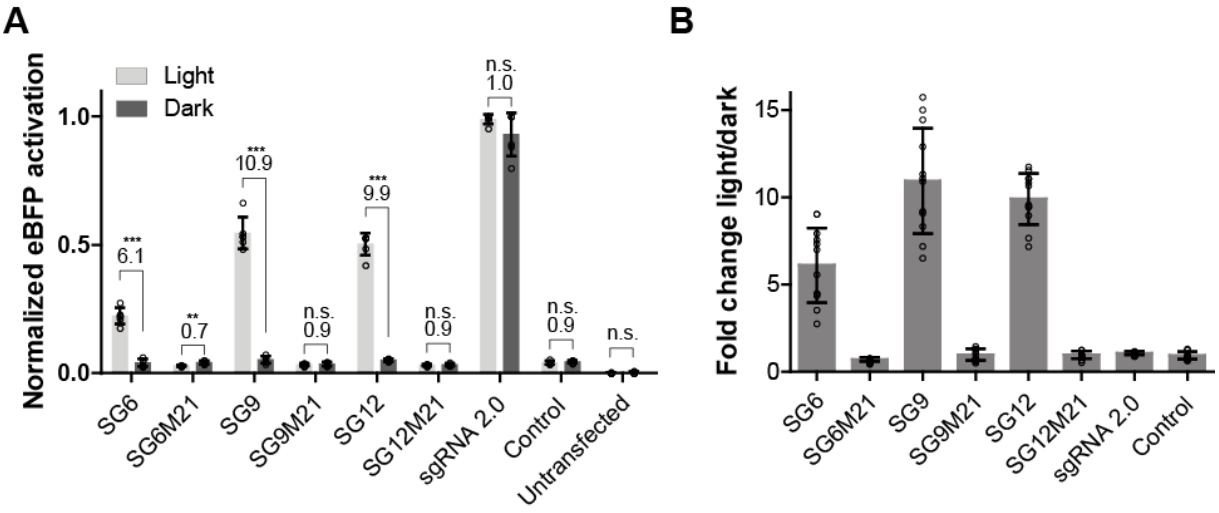


Figure 52: Normalized light-induced activation of eBFP expression testing varying aptamer stem-lengths and point mutants controls measured by flow cytometry. A) Light-dependent regulation of eBFP using PAL aptamer extended sgRNAs with varying stem-lengths (SG6, SG9 and SG21) or PAL-binding deficient mutants (SG6M21, SG9M21 and SG12M21). The data is presented as mean values ± s.d and was normalized to the sgRNA 2.0 control using min-max scaling (n = 6 from 3 individual experiments with two cell culture replicates. Welch's two-tailed t test was performed for the fold induction in the light versus the sample in the dark, resulting in p values for SG6 of p = 0.000004, for SG6M21 of p = 0.0059, for SG9 of p = 0.000008, for SG9M21 of p = 0.4794, for SG12 of p = 0.000002, for SG12M21 of p = 0.5292, for sgRNA 2.0 of p = 0.3876, for Control of p = 0.3251 and Untransfected of p = 0.1485, whereas the fold-changes of SG9M21, SG12M21, sgRNA 2.0, Control and Untransfected were non-significant (ns). *p <0.05, p** <0.01, ***p<0.005 versus the sample in the dark). B) Light-dependent regulation of eBFP using PAL aptamer extended sgRNAs with varying stem-lengths (SG6, SG9 and SG21) or PAL-binding deficient mutants (SG6M21, SG9M21 and SG12M21). The dynamic range was calculated by showing the fold-changes between samples incubated in the light and samples incubated in the dark. The data represents the results from A).

sgRNAs SG9 and SG12 show the largest differences comparing samples incubated in the light to samples in the dark **(Figure 52)** of 10.9-fold for SG9 and 9.9-fold of SG12, whereas the fold-change for SG6 was at 6.1. The positive control sgRNA 2.0 was

activating slightly stronger, leading to no significant difference between light and dark, whereas native sgRNA (Control), does not induce eBFP expression, neither in the light nor in the dark. All mutant plasmids, SG6M21, SG9M21 and SG12M21 do not induce eBFP expression **(Figure 52)**. Based on these results, SG9 and SG12 were found to be the most potent candidates for LITA.

During the experiments, it was identified, that a point mutation in the NLS region (from PKKKRKV to PKKRRKV) of PHP might cause a decreased efficiency of activation. The NLS was corrected using inverse PCR **(VIII.14.6)** and mutagenic primers. The primers used were NPHPTM_NLS_corr_fwd and NPHPTM_NLS_corr_rev. The PCR was running for 30 cycles, Θ_{elong} was 180s and T_{ann} was 68°C for 10s. The PCR products were analyzed by agarose gel electrophoresis **(Figure 53, S1)**. Two PCR products appeared, from which the lower one is migrating at a size of < 500 bp and the upper product at ~ 6000 bp. The higher migrating product was isolated from the gel and purified **(VIII.6)**. Subsequently, the plasmid was further processed by inverse PCR cloning **(VIII.14.5)**. From a total of four clones, all were sequenced correctly (SEQ_ID: 009 - 011).

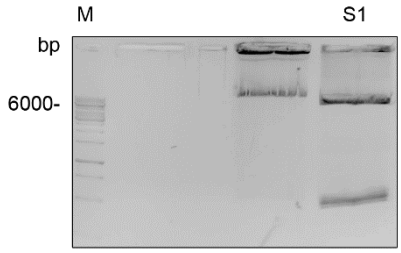


Figure 53: 1% agarose gels with reaction products of the inverse mutagenic PCR of PHP NLS correction. M: Thermoscientific GeneRuler 1kb DNA Ladder. S1: PCR products of the inverse PCR of PHP at 6326 bp. The samples were migrating for 30 min at 130 V and stained with ethidium bromide.

The mutation was corrected and three clones of the NLS corrected PHP were tested in LITA using SG9. As control a PHP variant lacking the NLS signal was used.

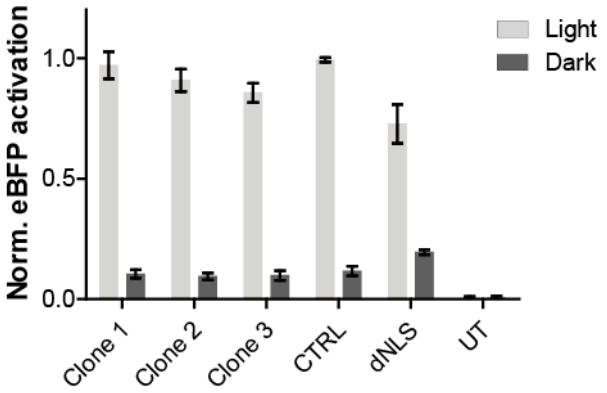


Figure 54: Normalized light-induced activation of eBFP expression testing PHP constructs with varying nuclear localization sequences measured by flow cytometry. Clone 1-3 have the corrected PKKKRKV NLS, whereas CTRL is the PHP, which was used before and having the PKKRRKV NLS. As control a dNLS variant of PHP was used to demonstrate the impact of the NLS sequence. The data is presented as mean values ± s.d and was normalized to CTRL using min-max scaling (n = 4 from 2 individual experiments with two cell culture replicates).

It was found, that the PHP variant used does not show a decrease in performance compared to the new clones with corrected NLS **(Figure 54)**. The PHP variant without NLS was performing slightly weaker compared to variants using the NLS. Based on this finding the previous PHP using the PKKRRKV was used in further assays.

Next, it was investigated, whether two aptamers within the sgRNA are necessary for robust gene activation or if a single aptamer is sufficient. Thus, SG9 was used as sgRNA and the M21 point mutation was introduced either in the tetraloop (M21TL) or stem-loop 2 (M21SL2). The M21 point mutation of aptamer 53 is characterized by an exchange of the Adenine at position 21 in the loop region of the aptamer to a uridine. This mutation renders the aptamer unable to bind PAL in the light nor in the dark as shown in Weber *et al.* 2019 Sup Fig 7 c[166]. The cloning was done using AQUA cloning. For each construct a sense and antisense DNA strand was hybridized and PCR amplified to generate the inserts.

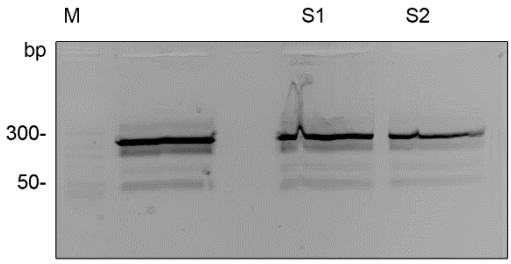


Figure 55: 4% agarose gel with sgRNA inserts targeting the 8x sgRNA binding site in pGL3-Basic-8x-sgRNA-eBFP and having one aptamer mutated. M: Thermoscientific GeneRuler 100 bp DNA Ladder. S1: sgRNA SG9M21TL at 236 bp. S2: sgRNA SG9M21SL2 at 236 bp. The samples were migrating for 20 min at 150 V and stained with ethidium bromide.

In **Figure 55**, both inserts of sgRNA versions with one aptamer functional and the other aptamer mutated were successfully amplified and were used for cloning as described previously. For both constructs clones with the correct sequences were identified (SEQ_ID: 041 - 042).

The new constructs were tested in a LITA assay **(VIII.19.5)** and the results were analyzed by flow cytometry.

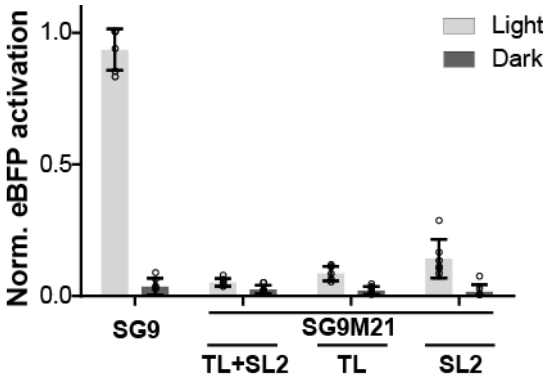


Figure 56: Light-dependent induction of eBFP expression using sgRNAs having one (SG9M21 TL; SG9M21SL) or both (SG9M21TL+SL) aptamer domains mutated. The data is normalized to SG9 in light using min-max scaling (n = 6 from 3 individual experiments).

Robust activation was found for SG9 in the light but not in the dark **(Figure 56)**. The corresponding M21 double mutant shows no activation neither in the light nor in the dark. When analyzing the single point mutations, both mutations massively decrease

the activation potency in the light, where levels in the dark remain at background level. These findings show, that two aptamers are essential for most efficient gene activation.

Next, the impact of light-intensity on the assay was tested. LITA was done using SG9 and using a LPA device **(VIII.19.3)**.

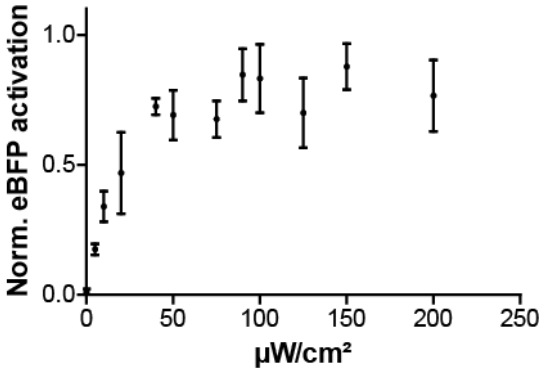


Figure 57: Impact of increasing light intensity on BFP expression using SG9. Given are mean values ± s.d normalized to the highest value using min-max scaling (n = 4 from 2 individual experiments).

When no light is applied to the cells, the eBFP expression remains at background **(Figure 57)**. Light intensity can be increased up to 100 µW/cm², while observing an increase in eBFP expression. For light-intensities > 100 µW/cm² eBFP expression remains stable or even decreases **(Figure 57)**. It was also found, that between 0 to 50 µW/cm² the eBFP expression increases nearly linearly with increasing light intensity.

To further optimize the LITA activation efficiency, the impact of the order of the effectors p65 and HSF1 within the PHP fusion protein was tested.

Based on the successful cloning of PHP, a variant pNLS-HSF1-p65-PAL-T2A-mCherry, dubbed HPP was cloned to investigate the impact of the order of the transcriptional activators within the fusion protein. The strategy to generate HPP from PHP was based on two steps. In the first step pNLS-HSF1-PAL-T2A-mCherry (dubbed HP), was generated by inverse PCR, re-circularized, transformed and sequenced. Linear HP was then generated by PCR. In a parallel PCR reaction p65 was amplified from PHP by Inverse-PCR **(VIII.14.5)** and AQUA cloning **(VIII.14.2)** was used to assemble the circular HPP in *E. coli*.

HPP was generated using NHPTM-fwd and NHPTM-rev in a Flash Fusion PCR reaction **(VIII.3.3)**. The PCR was running for 30 cycles, Θ_{elong} was 180s and T_{ann} was 68°C for 10s. Products of the PCR were analyzed by agarose gel electrophoresis. The PCR products were further processed according to **VIII.14.5**.

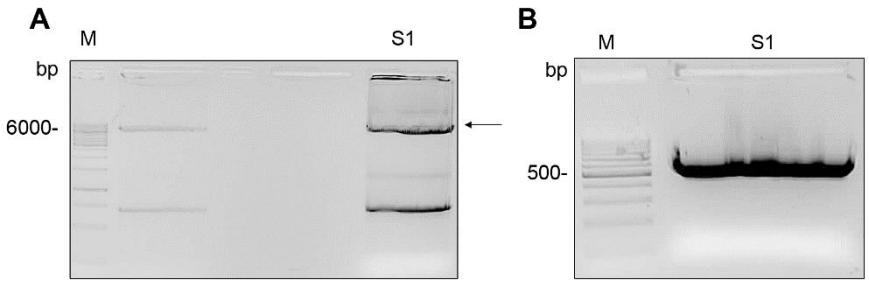


Figure 58: Agarose gels with reaction products of the inverse PCR of PHP and the PCR of p65. M: ThermoScientific GeneRuler 1kb DNA Ladder. A) 1% agarose gel. S1: PCR products of the inverse PCR of PHP at 6326 bp B) 4% agarose gel. S1: PCR products of the amplification of p65 from PHP at 543 bp. The samples were migrating for 30 min at 130 V and stained with ethidium bromide.

The PCR products of the amplification of PHP by excluding the p65 insert were visualized on a 1% agarose gel **(Figure 58, S1)**. Three products were visible from which one was migrating at ~750 bp, the second at ~1200 bp and the third at ~6000 bp. The third product was recovered from the gel and purified **(VIII.6)**, since it is the closest to the desired PCR product which would run at 6326 bp. The product was cloned into *E. coli* by inverse-PCR cloning and analyzed by sequencing (SEQ_ID: 006). Two clones were identified, from which one clone was mostly correct besides a single point mutation. The p65 insert was generated by Flash Phusion PCR **(VIII.3.3)** from PHP as template using p65-in-NHPTM-fwd and p65-in-NHPTM-rev as primers. The PCR was running for 30 cycles, Θ_{elong} was 25s and T_{ann} was 68°C for 10s. PCR products were visualized by agarose gel electrophoresis and are depicted in **Figure 58 B**. A lot of PCR product was obtained at ~ 500 bp, which was purified **(VIII.6)**. Next, p65 was cloned into the linear plasmid backbone of HP using AQUA cloning. First HP was linearized by inverse PCR using NHPTM-BB-fwd and NHPTM-BB-rev as primers in a Flash Phusion PCR **(VIII.3.3)**. The PCR was running for 30 cycles, Θ_{elong} was 180s and T_{ann} was 68°C for 10s. The PCR products were analyzed by agarose gel electrophoresis **(Figure 59, A)**.

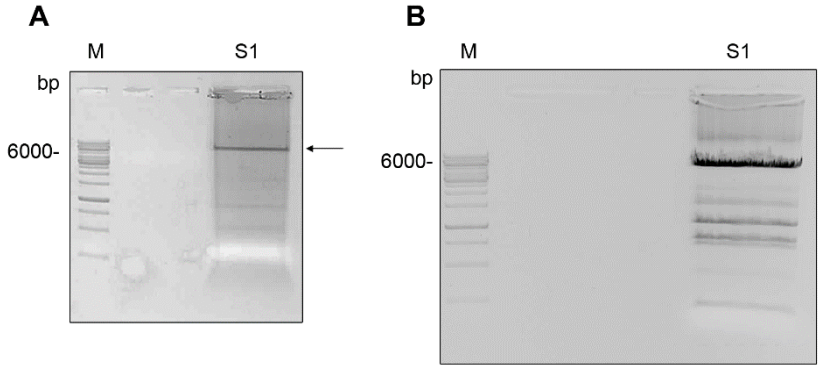


Figure 59: 1% agarose gels with reaction products of the inverse PCR of HP and the inverse PCR of HPP. M: Thermoscientific GeneRuler 1kb DNA Ladder. A) S1: PCR products of the inverse PCR of HP at 6326 bp B) 4% agarose gel. S1: PCR products of the amplification of HPP (mutation correction) at 6884 bp. The samples were migrating for 30 min at 130 V and stained with ethidium bromide.

In **Figure 59 A** S1, a PCR product migrating at ~ 6000 bp appeared and was cut out and purified **(VIII.6)**. The isolated plasmid backbone was used together with the isolated p65 for AQUA cloning, using 90 ng of plasmid backbone and a 10:1 molar excess of p65 insert **(VIII.14.2)**. Clones were analyzed by sequencing, from which two clones were positive (SEQ_ID: 007). Clone 1 was further used for the correction of the point mutation, which was still present. The point mutation of HPP clone 2 was corrected by inverse PCR using NHPPTM_mut_corr_fwd and NHPPTM_mut_corr_rev as primers **(VIII.14.6)**. The PCR was running for 30 cycles, Θ_{elong} was 180s and T_{ann} was 68°C for 10s. The PCR products were analyzed by agarose gel electrophoresis **(Figure 59, B, S1)**. The product at ~6000 bp was cut out and purified by DNA purification kit. Subsequently, the product was processed by inverse PCR cloning **(VIII.14.6)**. Three clones were sent for sanger sequencing, from which all three clones were positive and clone 1 was further used (SEQ_ID: 008).

Both fusion proteins, PHP and HPP, using SG9 and the corresponding M21 control were then tested in a LITA assay **(VIII.19.5) (Figure 60)**.

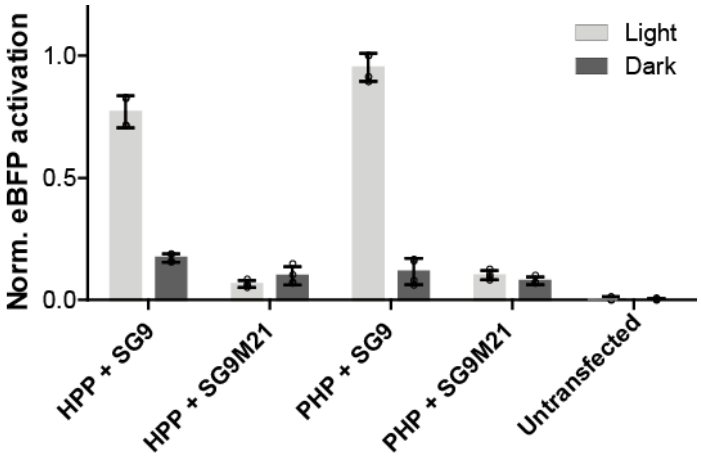


Figure 60: Influence of the arrangement of the activator domains on light-dependent eBFP expression (PHP: p65-HSF1-PAL; HPP: HSF1-p65-PAL). The data is presented as mean values ± s.d (n = 4 from 2 individual experiments).

It was found that the previous arrangement of PHP was superior to HPP when used for eBFP activation and using SG9 **(Figure 60)**. It was also observed, that for HPP a very small increase of eBFP activation in the dark was detectable. For the SG9M21 mutants, the eBFP expression levels remained at background level.

To prove the spatio-temporal component for LITA, a photo-mask experiment was performed. It was hypothesized, that a monolayer of transfected cells only would express the reporter gene, if light reaches the cells, which can be in a pattern using a photo mask. HeLa cells were seeded in a glass-bottom plate, transfected the cells using SG9 and tested three irradiation patterns. Either a slide photomask was used **(Figure 61, A)**, no photomask **(Figure 61, B)** or cells covered and kept in the dark **(Figure 61, C)**. The cells were incubated using the LPA device and the fluorescence signal of the cell was then analyzed by laser-scanning microscopy **(VIII.19.2)**. The images were then processed by a python script **(XI.6)**, which converts the image to grey scale and then vertically integrates the pixel values, to visualize regions of activated cells **(Figure 61)**. The process of evaluation is described in **VIII.19.4**.

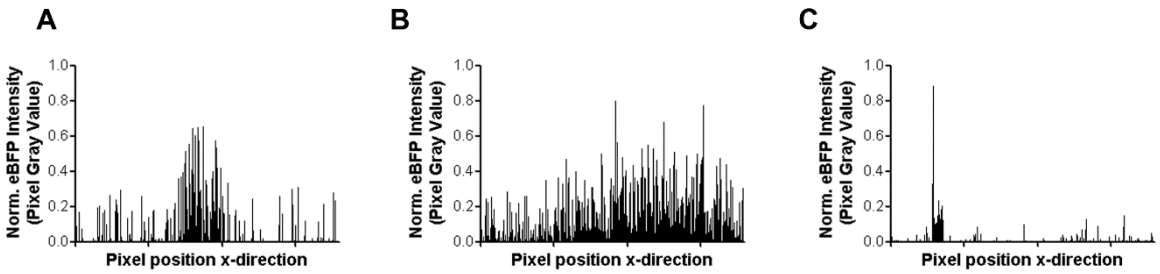


Figure 61: Integrated vertical pixel density of eBFP activated cells using laser scanning microscopy. The pixel position in x-direction indicates the vertically integrated pixel value. For the y-axis scaling was applied to normalize the pixel values to the highest value in the corresponding image. A) A slide photomask was used. B) no mask was applied. C) cells were fully covered by a photomask. Raw picture data in Supporting Figures 3.

A relative homogenous distribution of activated cells was found for HeLa cells, that were activated without a photomask, whereas the distribution of activated cells is slightly shifted to the right x-direction **(Figure 61, B)**. For cells covered with a photomask and thereby incubated in the dark, no activation was indicated. However, a spike signal was identified **(Figure 61, C)**. HeLa cells, that were incubated using a photomask with a slide opening show increased fluorescence at the position of the slide, as indicated by the region of higher fluorescence signal in **Figure 61 A**. Compared to **Figure 61, B**, the overall level of activated cells is much lower in **Figure 61 A**, where cells were in the dark. These findings show, that the system is capable of being precisely activated by user-defined areas.

V.2.4.2 Light-dependent activation of CXCR4 expression

It was hypothesized, that LITA can be tested on endogenous targets such as CXCR4. CXCR4 is a receptor, that is expressed on the surface of a variety of mammalian cells and thus can be addressed by antibody staining[191]. To select a cell line, HeLa, MCF7 (negative control) and A549 (positive control) cells[c] were seeded. After 24 hours the cells were analyzed by antibody staining and flow cytometry including cells that were not stained as negative controls to analyze the base line expression **(VIII.19.6) (Figure 62)**.

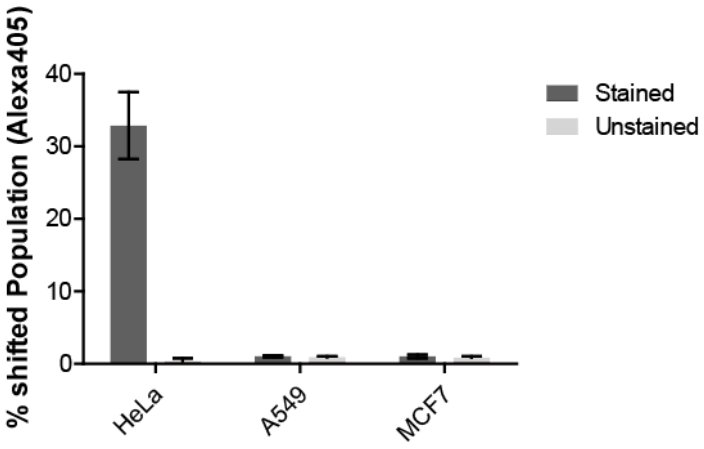


Figure 62: HeLa, A549 and MCF7 cells stained by antibody staining and measured using flow cytometry. The data is presented as mean values ± s.d (n = 4 from 2 individual experiments).

CXCR4 expression was detected for HeLa cells but not for A549 cells. MCF7 cells did not show CXCR4 expression.

To investigate the potential of LITA with endogenous promoters the CXCR4 promoter was targeted by SG9 sgRNAs. The promoter position 6 (CXCR4-#6) was tested first. The sgRNA sense and antisense DNA strands were hybridized and amplified in a

[c] proteinatlas.org/, 11.03.2019

standard PCR reaction as described previously using CXCR4_sgRNA_in_pENTR_fwd and sgRNA_in_pENTR_rev as primers.

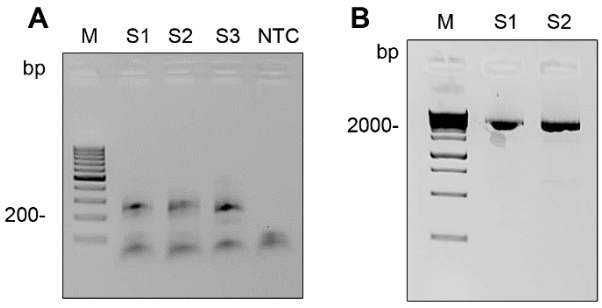


Figure 63: 4% agarose gel with sgRNA inserts targeting position 6 of the human CXCR4 promoter. A) 4% agarose gel. M: Thermoscientific GeneRuler 100 bp DNA Ladder. S1: CXCR4-#6 targeting sgRNA SG9 at 224 bp. S2: CXCR4-#6 targeting sgRNA SG9M21 at 236 bp. S3: CXCR4-#6 targeting sgRNA 2.0 at 208 bp. NTC: no template control. B) 1% agarose gel. M: TS GeneRuler 1kb DNA Ladder. S1: pENTR.hU6.sgRNA.CXCR4-#6.SG9 at 2749 bp. S2: pENTR.hU6.sgRNA.CXCR4-#6.sgRNA2.0 at 2733 bp. The samples were migrating for 20 min at 140 V and stained with ethidium bromide.

In **Figure 63 A**, the PCR products of the sgRNAs targeting CXCR4-#6 are visualized on a 4% agarose gel, with each product migrating at its predicted size. The products were purified **(VIII.6)** and were cloned into the pENTR.hU6 plasmid backbone by AQUA cloning **(VIII.14.2)**. After sequencing only one clone of sgRNA CXCR4-#6 SG9M21 was correct (SEQ_ID: 022). However, there were clones with a single point mutation for sgRNA CXCR4-#6 2.0 and sgRNA CXCR4-#6 SG9. The point mutations were corrected using mutagenic primers and inverse PCR **(VIII.14.5)**. The primers used for the sgRNA CXCR4-#6 2.0 reaction were CXCR4_2.0_mut_fwd and CXCR4_2.0_mut_rev and for the sgRNA CXCR4-#6 SG9 were CXCR4_M2S9_mut_fwd and CXCR4_M2S9_mut_rev. As templates for SG9 clone 1 was used and for sgRNA 2.0 clone 3 was used. The PCR products were analyzed by agarose gel electrophoresis **(Figure 63, B)**. PCR products appeared at their expected sizes and the plasmids were circularized and transformed into *E. coli*. A correct clone was obtained for sgRNA CXCR4-#6 2.0 (SEQ_ID: 023). For sgRNA CXCR4-#6 SG9 a new strategy was tested, using the 8xsgRNA binding site targeting sgRNAs of the pGL3-Basic-8x-sgRNA-eBFP plasmid and exchanging the seed-region by inverse PCR **(VIII.14.5)**. The primers used for the mutation were CXCR4_M2S9_new_fwd and

CXCR4_M2S9_new_rev. The PCR products were visualized on an agarose gel **(Figure 64)** indicating that the PCR generated a product at the expected size of 2749 bp.

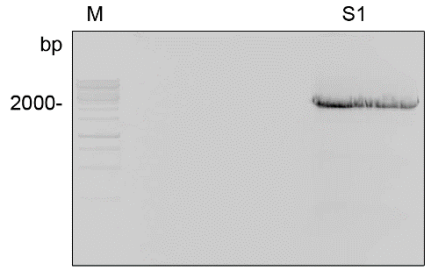


Figure 64: 1% agarose gel with PCR products of the inverse PCR of pENTR.hU6.sgRNA.CXCR4-#6.SG9M21. M: Thermoscientific GeneRuler 1kb DNA Ladder. S1: linear pENTR.hU6.sgRNA.CXCR4-#6.SG9M21 at 2749 bp. The samples were migrating for 20 min at 140 V and stained with ethidium bromide.

The plasmid was circularized and transformed into *E. coli*. After sequencing, a correct clone of pENTR.hU6.sgRNA.CXCR4-#6.SG9 was identified (SEQ_ID: 024). CXCR4 activation was tested in a LITA assay using HeLa cells. HeLa cells were seeded and transfected with dCas9, PHP and SG9 sgRNA targeting a region upstream of the human CXCR4 promoter. The seed-sequence was used from Zalatan *et al*., where it was shown to target the CXCR4 promoter using the SAM system[134]. sgRNA 2.0 was used as a positive control. After 24 hours, the cells were analyzed by flow cytometry **(Figure 65)**.

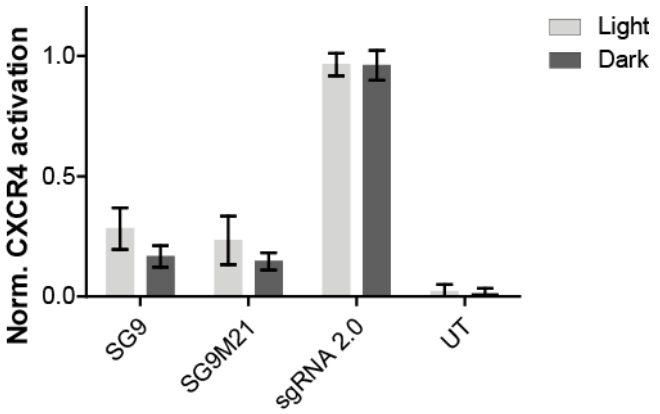


Figure 65: Light-dependent upregulation of CXCR4 in HeLa cells, stained by antibody staining and measured using flow cytometry. The data was normalized to sgRNA 2.0 using min-max scaling and is presented as mean values ± s.d (n = 6 from 3 individual experiments).

In **Figure 65**, CXCR4 upregulation was successful for light-independent sgRNA 2.0, whereas the CXCR4 expression using SG9 targeting the CXCR4 promoter was comparable to the SG9M21 point mutant. Small but no significant differences were found between light and dark indicating the expression levels to be close to the background **(Figure 65)**. It was hypothesized, that for PHP the seed-sequence used in SG9 and SG9M21 might not be optimal, thus an array of 10 different published seed-sequences from Zalatan *et al*. was tested, all targeting the CXCR4 promoter, from which position 6 (pos6) was the seed-sequence which was already tested before **(Figure 65)**. To evaluate the background level of CXCR4 activation, a control without gRNA was included. Since the generation of CXCR4-#6 sgRNAs was successful further sgRNAs of positions CXCR4-#1 to CXCR4-#10 of SG9 were generated using the inverse PCR strategy, where a template with the correct aptamer domain was used and only the seed-region was exchanged by extended primers. In this case, a universal reverse primer, gRNA_insert_universal_rev was used and the forward primer were modified according to the desired seed-region **(Table 2)**.

Table 2: CXCR4 target sites and forward primers

CXCR4 position	Forward primer
CXCR4-#1	gRNA_M2S9_sgCXCR4.1_fwd
CXCR4-#2	gRNA_M2S9_sgCXCR4.2_fwd
CXCR4-#3	gRNA_M2S9_sgCXCR4.3_fwd
CXCR4-#4	gRNA_M2S9_sgCXCR4.4_fwd
CXCR4-#5	gRNA_M2S9_sgCXCR4.5_fwd
CXCR4-#7	gRNA_M2S9_sgCXCR4.7_fwd
CXCR4-#8	gRNA_M2S9_sgCXCR4.8_fwd
CXCR4-#9	gRNA_M2S9_sgCXCR4.9_fwd
CXCR4-#10	gRNA_M2S9_sgCXCR4.10_fwd

As template the 8xsgRNA binding site targeting sgRNA SG9 was used. The products of the inverse PCR were visualized on a 1% agarose gel **(Figure 66)**.

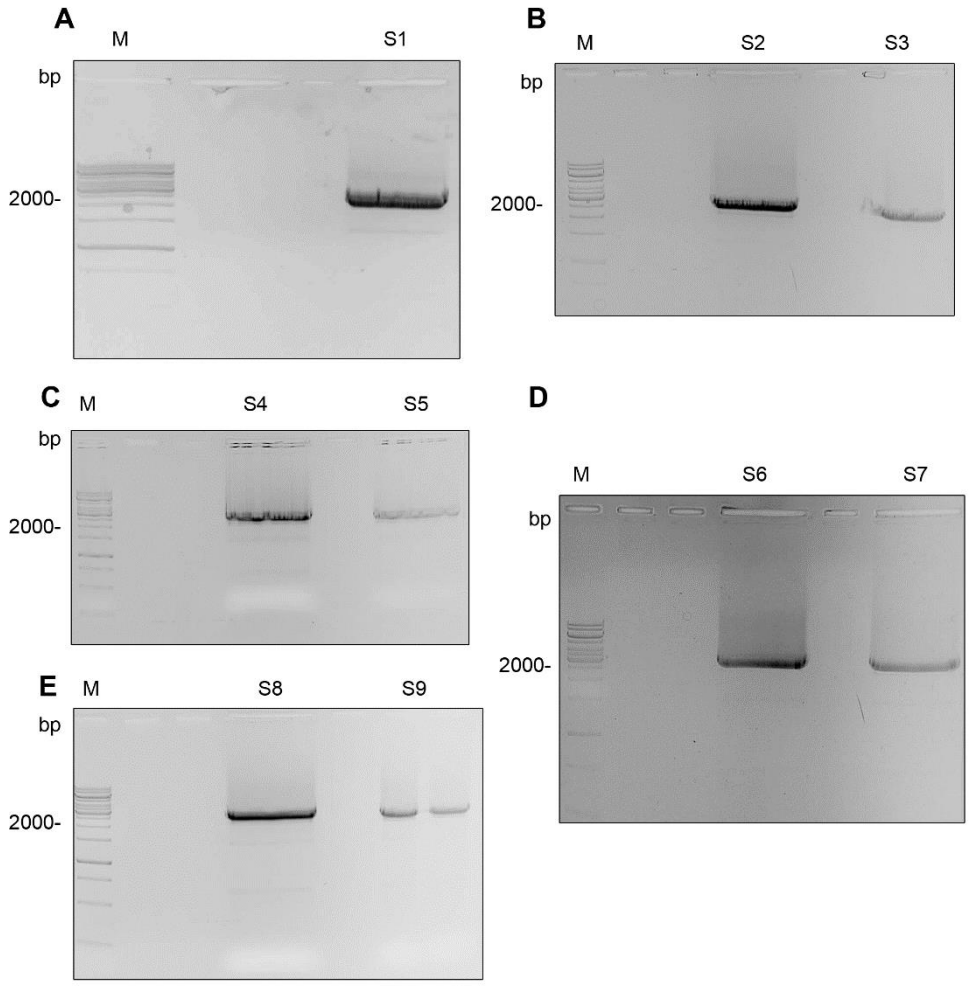


Figure 66: 1% agarose gels with PCR products of the inverse PCR of pENTR.hU6.sgRNA SG9 expressing sgRNAs targeting different positions on the human CXCR4 promoter. M: Thermoscientific GeneRuler 1kb DNA Ladder. A) S1: linear pENTR.hU6.sgRNA.CXCR4-#1.SG9 at 2749 bp. B) S2: linear pENTR.hU6.sgRNA.CXCR4-#2.SG9 at 2749 bp. S3: linear pENTR.hU6.sgRNA.CXCR4-#3.SG9 at 2749 bp. C) S4: linear pENTR.hU6.sgRNA.CXCR4-#4.SG9 at 2749 bp. S5: linear pENTR.hU6.sgRNA.CXCR4-#5.SG9 at 2749 bp. D) S6: linear pENTR.hU6.sgRNA.CXCR4-#7.SG9 at 2749 bp. S7: linear pENTR.hU6.sgRNA.CXCR4-#8.SG9 at 2749 bp. D) S8: linear pENTR.hU6.sgRNA.CXCR4-#9.SG9 at 2749 bp. S9: linear pENTR.hU6.sgRNA.CXCR4-#10.SG9 at 2749 bp. The samples were migrating for 20 min at 140 V and stained with ethidium bromide.

For all inverse PCRs the PCR products appeared at the correct size of 2749 bp **(Figure 66)**. The products were isolated, purified and used for inverse PCR cloning **(VIII.14.5)**. Sequencing revealed for each construct correct clones (SEQ_ID: 025 - 032).

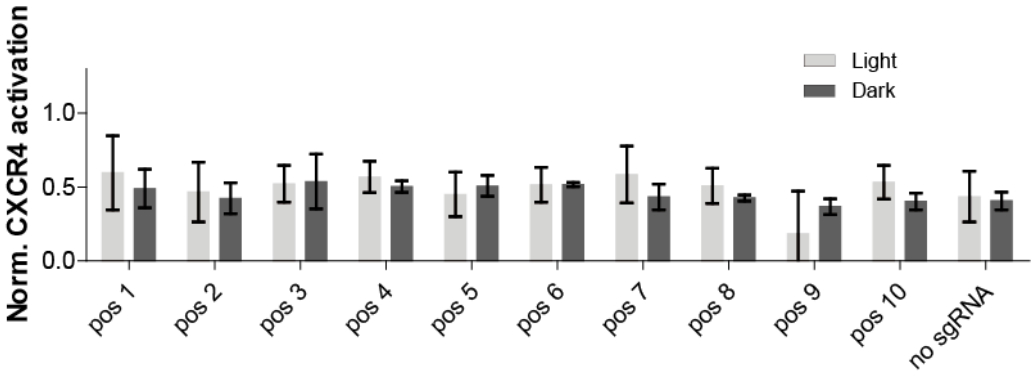


Figure 67: Light-dependent upregulation of CXCR4 in HeLa cells testing different seed-sequences, stained by antibody staining and measured using flow cytometry. The data was scaled to a shift of 2 using min-max scaling and is presented as mean values ± s.d (n = 4 from 2 individual experiments).

All samples, independent from the incubation in the dark or in the light, were at background level (no sgRNA), indicating no light-dependent activation of CXCR4 expression **(Figure 67)**.

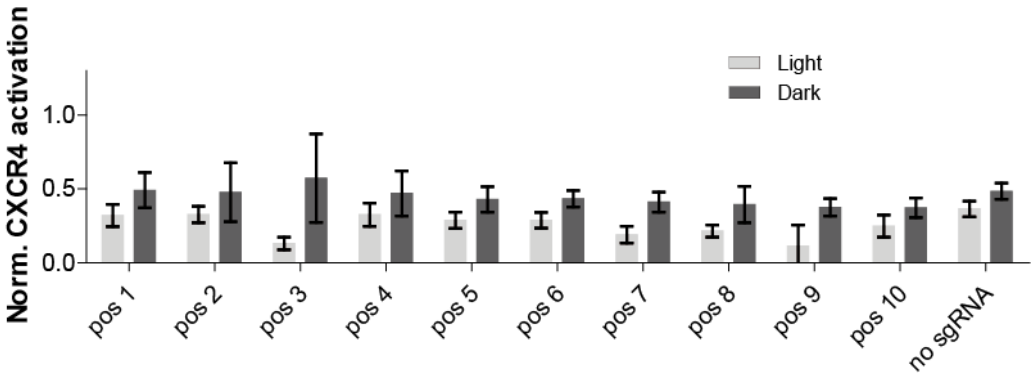


Figure 68: Light-dependent upregulation of CXCR4 in Hek293T cells testing different seed-sequences, stained by antibody staining and measured using flow cytometry. The data was scaled to a shift of 2 using min-max scaling and is presented as mean values ± s.d (n = 4 from 2 individual experiments).

Also, for HEK293T cells no activation of CXCR4 was detected and all samples at comparable levels of the no sgRNA background control **(Figure 68)**. With these findings it was decided to switch the target gene for the activation of endogenous genes using LITA.

V.2.4.3 Light-dependent activation of ASCL1 expression

Another often used target for CRISPRa is ASCL1. ASCL1 encodes a transcription factor, which is involved in neuronal differentiation. In HeLa cells the expression of ASCL1 is low, making it a suitable target gene for artificial upregulation leading to high dynamic ranges. Unlike CXCR4, ASCL1 is expressed in the nucleoplasm and additionally in the cytosol[192]. Thus, it cannot be measured using antibody staining in combination with flow cytometry. It was decided to analyze ASCL1 using qPCR, as this is a procedure, which is already described in literature[64].

The same technique was used to generate aptamer modified sgRNAs that target the human ASCL1 promoter. Again, as PCR template the 8xsgRNA binding site targeting SG9 sgRNA was used and as positive control the appropriate sgRNA 2.0. Primers were designed, whereas the reverse primer was always the gRNA_insert_universal_rev and the forward primer was according to the seed-region which was added to the sgRNA **(Table 3)**.

Table 3: ASCL1 target sites and forward primers.

ASCL1 position	System	Forward primer
ASCL1-#2	SAM	sgASCL1-2_2.0_fwd
ASCL1-#1	PAL	sgASCL1-1_M2S9_fwd
ASCL1-#2	PAL	sgASCL1-2_M2S9_fwd
ASCL1-#3	PAL	sgASCL1-3_M2S9_fwd
ASCL1-#4	PAL	sgASCL1-4_M2S9_fwd
ASCL1-#5	PAL	sgASCL1-5_M2S9_fwd
ASCL1-#6	PAL	sgASCL1-6_M2S9_fwd

All new plasmids were generated by inverse PCR as described above **(VIII.14.5)** and visualized on 1% agarose gels **(Figure 69)**.

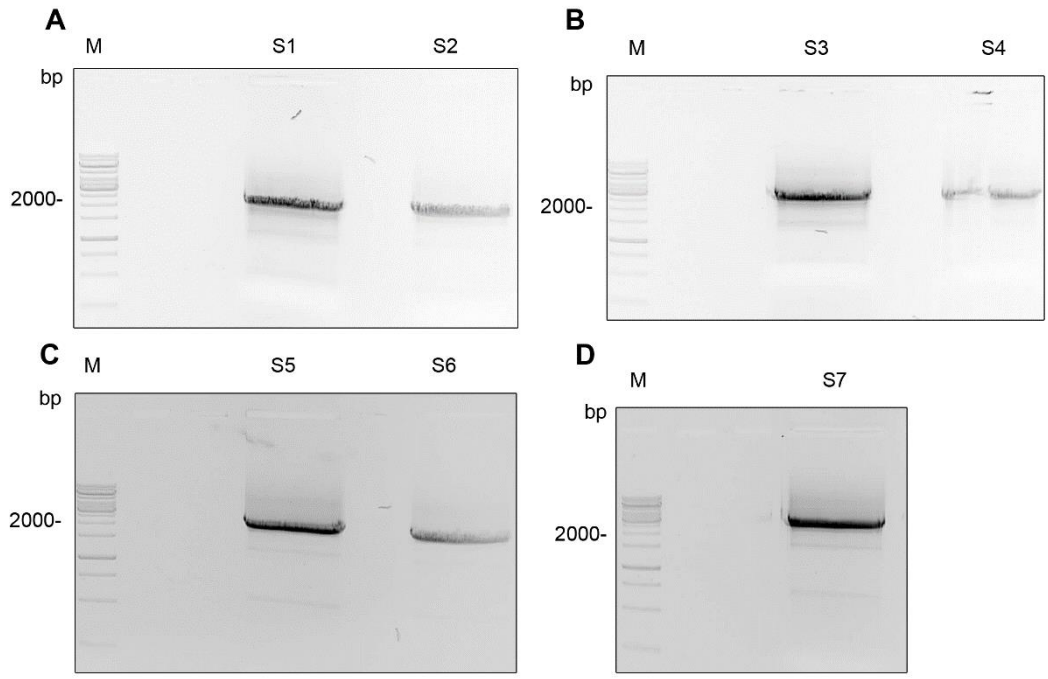


Figure 69: 1% agarose gels with PCR products of the inverse PCR of pENTR.hU6.sgRNA SG9 expressing sgRNAs targeting different positions on the human ASCL1 promoter. M: Thermoscientific GeneRuler 1kb DNA Ladder. A) S1: linear pENTR.hU6.sgRNA.ASCL1-#2.sgRNA2.0 at 2733 bp. S2: linear pENTR.hU6.sgRNA.ASCL1-#1.SG9 at 2749 bp. B) S3: linear pENTR.hU6.sgRNA.ASCL1-#2.SG9 at 2749 bp. S4: linear pENTR.hU6.sgRNA.ASCL1-#3.SG9 at 2749 bp. C) S5: linear pENTR.hU6.sgRNA.ASCL1-#4.SG9 at 2749 bp. S6: linear pENTR.hU6.sgRNA.ASCL1-#5.SG9 at 2749 bp. D) S7: linear pENTR.hU6.sgRNA.ASCL1-#6.SG9 at 2749 bp. The samples were migrating for 30 min at 140 V and stained with ethidium bromide.

All PCR products of the inverse PCR for the seed-sequence modification were identified migrating at the expected size **(Figure 69)**. The products were isolated from the gel and purified using DNA purification kit. The linear plasmids were cloned according to inverse PCR cloning **(VIII.14.5)**. For each construct a clone with the correct sequence was identified (SEQ_ID: 033 - 039).

In order to evaluate ASCL1 as a new target for LITA assay, the activation using sgRNA 2.0 and a non-binding sgRNA as negative control (SG9 targeting the 8xsgRNA binding site of pGL3-Basic-8xsgRNA-eBFP) was tested first to determine the dynamic range between fully activated and non-activated cells. In the same setup the impact of light/dark incubation of the samples on ASCL1 expression was tested. HeLa cells were seeded and transfected using dCas9, MPH and sgRNA 2.0 targeting ASCL1 promoter position 2. The downstream protocol for qPCR evaluation was as published and is described in **VIII.19.8**, using GAPDH as house-keeping gene. The Δct values of ASCL1

were normalized to GAPDH and were plotted in **Figure 70 A**. Additionally, for GAPDH a standard curve was recorded.

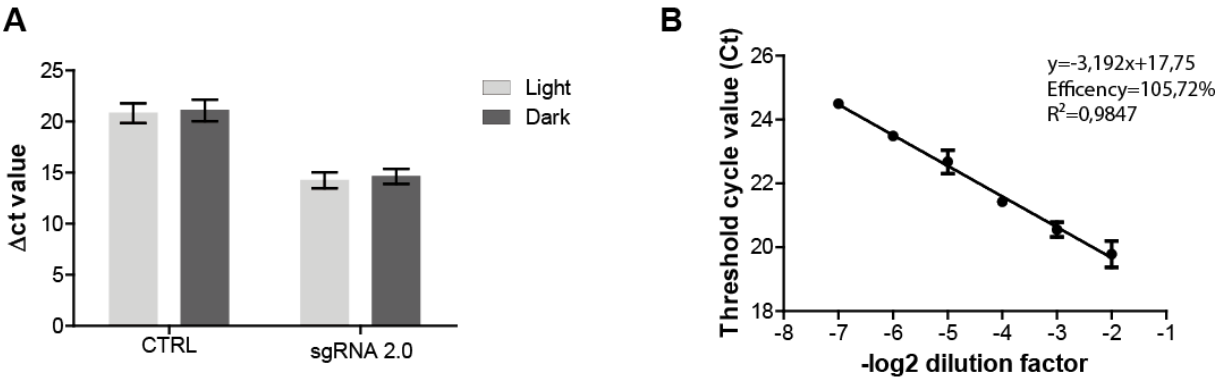


Figure 70: qPCR results from the ASCL1 test activation using MPH and sgRNA 2.0 targeting promoter position 2 of human ASCL1. A) Δct values of ASCL1 were normalized to GAPDH and plotted. B) A dilution series was performed to analyse the linearity and quality of the qPCR reaction. The data is presented as mean values ± s.d (n = 4 from 2 individual experiments). Raw qPCR data are in XI.4.1.

Initial experiments activating ASCL1 indicated, that using sgRNA 2.0 targeting promoter position 2 in combination with MPH and dCas9 was successfully activating ASCL1 expression independent from the state of irradiation **(Figure 70)**. A difference of more than 5 Δct values indicates an upregulation of >32x. It was found, that the raw Δct values of non-activated ASCL1 are at cycle 40, which is the background of what was measured with the qPCR. A calibration curve for GAPDH was recorded, indicating a PCR efficiency of 105.72% and R^2=0.9847 **(Figure 70, B)**.

We continued to test ASCL1 expression using LITA assay and designed sgRNAs binding to the ASCL1 promoter **(Figure 71)**

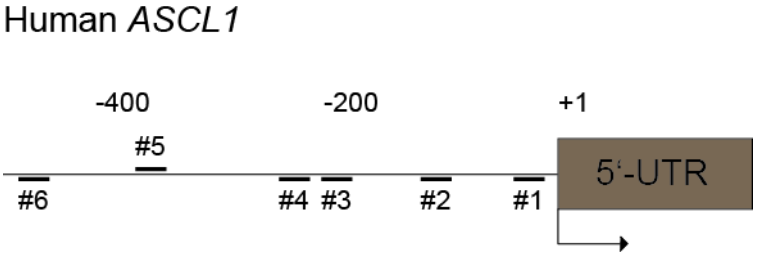


Figure 71: Scheme of the human ASCL1 promoter and the positions of the designed sgRNAs.

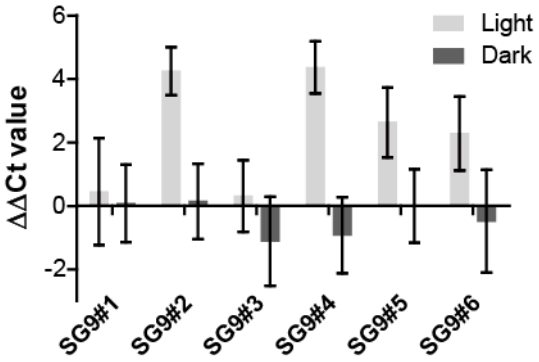


Figure 72: qPCR results from the ASCL1 activation using PHP and SG9 targeting varying promoter positions of human ASCL1. The raw Δct values were first normalized to the house-keeping gene GAPDH and then normalized to the negative control sgRNA (SG9 targeting the 8xsgRNA binding site in pGL3-Basic-8xsgRNA-eBFP). The data is presented as mean values ± s.d (n = 4 from 2 individual experiments). Raw qPCR data are in XI.4.2.

Screening of different sgRNA binding sites upstream of the ASCL1 promoter revealed that sgRNAs targeting position 2 (SG9#2) and sgRNA targeting position 4 (SG9#4) are able to induce ASCL1 expression in the light by ~16-fold **(Figure 72)**. No activation was found for position 1 (SG9#1) and position 3 (SG9#3), whereas only little activation was found for sgRNAs position 5 (SG9#5) and position 6 (SG9#6). These findings were solidified in a further LITA assay, focusing on position 2 and position 4 only. It was interesting to compare the performance to sgRNA 2.0 and if sgRNA multiplexing further increases ASCL1 expression in the light. Thus, sgRNA 2.0 was included into the experiment and also sgRNA SG9#2 and sgRNA SG9#4 were combined in a single multiplexed sample **(Figure 73)**.

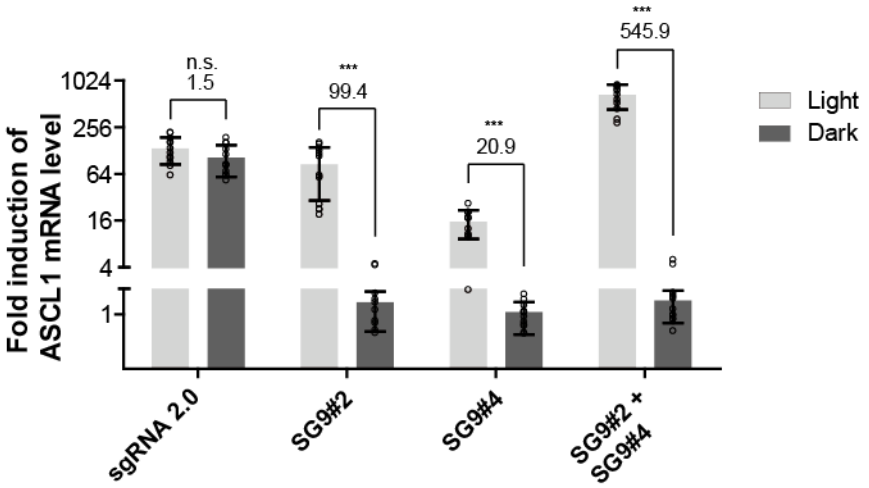


Figure 73: Light-dependent regulation of endogenous hASCL1 using sgRNAs SG9#2, SG9#4 and sgRNA2.0 (Δct-values were normalized to control sgRNA). For the fold induction of mRNA, 2^ΔΔct was calculated. Welch's two-tailed t test was performed for the fold induction in the light versus the sample in the dark, resulting in p values for #2 + #4 of p = 0.000001, for #4 of p = 0.000007 and for #2 of p = 0.0003, whereas the fold-change of sgRNA 2.0 was non-significant (ns).*p <0.05, p** <0.01, ***p<0.005 versus the sample in the dark. Values over the bar groups represent the light/dark fold-change of the sample pair. The data in b and c are presented as mean values ± s.d (n = 6 from 3 individual experiments with two cell culture replicates). Raw qPCR data are in XI.4.3.

Light-dependent overexpression of ASCL1 was found for SG9#2 with a light/dark fold-change of 99.4 and for SG9#4 of 20.9, respectively, whereas the activation of ASCL1 in the light using SG9#2 was comparable to the amount of light-independent activation of sgRNA 2.0 **(Figure 73)**. However, combining SG9#2 and SG9#4 a >1000x activation of ASCL1 expression was detected in the light and a light/dark fold-change of 545.9, indicating that multiplexing leads to an increased dynamic range in conditional ASCL1 expression.

These findings show, that the new system is capable of upregulating endogenous genes in mammalian cells.

V.2.5 Cloning of a single plasmid solution using the UniSAM plasmid

To simplify LITA in terms of applicability, dCas9, PHP and sgRNA could be combined in a single plasmid and thus increasing the transfection efficiency. In a previous publication it was shown that the SAM system was successfully cloned in a single plasmid called UniSAM[193]. Since the SAM system only differs in the sgRNA aptamers and the effector interface component (MS2), the MPH insert was replaced by PHP in a first step and then the sgRNAs were exchanged. To generate linear PB-UniSAM backbone inverse PCR was used **(VIII.14.6)**, to excise the MPH insert. PB-UniSAM was used as template and PB-UniSAM-Fwd and PB-UniSAM-Rev as primers. The PCR program was adjusted with Θ_{elong} as 3 min and T_{ann} as 72°C running for 25 cycles as a 2-step PCR. The PHP insert, which is tagged with T2A-mCHerry was amplified using Flash Phusion PCR, pNLS-p65-HSF1-PAL-T2A-mCherry plasmid as template and the primers NPHPTM-fwd and NPHPTM-rev. The PCR program was adjusted with Θ_{elong} as 60 s and using an extra annealing step with T_{ann} as 58°C for 10 s running the PCR for 30 cycles The PCR products of both reactions were visualized on a 1% agarose gel **(Figure 74)**.

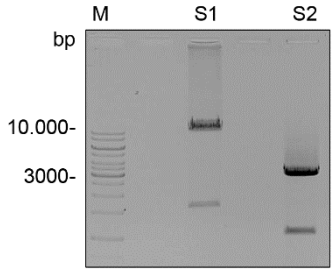


Figure 74: 1% agarose gel with reaction products of the PB-UniSAM plasmid backbone and the PHP insert. M: Thermoscientific GeneRuler 1kb DNA Ladder. S1: PB-UniSAM linear plasmid backbone at 10147 bp. S2: PHP insert at 2918 bp. The samples were migrating for 30 min at 130 V and stained with ethidium bromide.

Both PCR led to products at their expected sizes for PB-UniSAM at 10.147 bp and for the PHP insert at 2918 bp. The products were cut out and purified **(VIII.6)**. In a next step both products were used for AQUA cloning **(VIII.14.2)**. For sequencing of the clones, the primers NPHPTM_seq_1_fwd, NPHPTM_seq_2_fwd,

NPHPTM_seq_1_rev and NPHPTM_seq_2_rev was used. Two clones were successfully sequenced (SEQ_ID: 042 - 045) from which clone 3 was further used.

In a next step, the sgRNAs were inserted into the newly generated PB-UniSAM-PAL plasmid. The cloning was done using AQUA cloning **(VIII.14.2)**, where the PB-UniSAM and PB-UniSAM-PAL plasmids were linearized by inverse PCR **(VIII.14.5)**. The sgRNAs were PCR amplified from previous templates using sense and antisense strands. The sgRNAs were amplified using standard PCR and Pfu polymerase **(VIII.3.1)**. As primers eBFP_sgRNA_in_U6_fwd and eBFP_sgRNA_in_U6_rev was used and the PCR program was adjusted with Θ_{elong} as 50s and T_{ann} as 63°C running for 25 cycles. The PB-UniSAM-PAL and PB-Uni-SAM plasmids were amplified by inverse PCR **(VIII.14.5)**, using PB-UniSAM_U6_fwd and PB-UniSAM_U6_rev as primers with Θ_{elong} as 300s and T_{ann} as 65°C for 10s running for 28 cycles. The PCR products were visualized on agarose gels **(Figure 75)**.

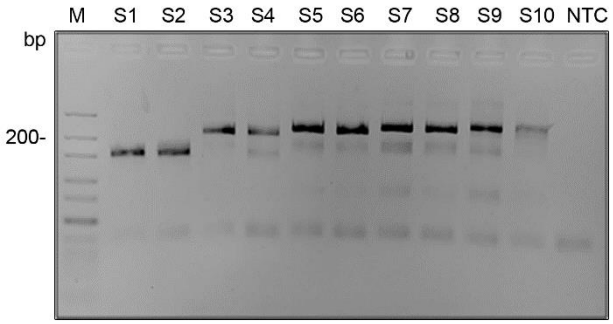


Figure 75: 4% agarose gel with sgRNA inserts programmed to bind to the 8x sgRNA binding site of pGL3-Basic-8xsgRNA-eBFP for PB-UniSAM. M: Thermoscientific GeneRuler ULR DNA Ladder. S1: native sgRNA (Control) at 131 bp, S2: non-targeting native sgRNA (Control) at 131 bp, S3: sgRNA 2.0 at 191 bp, S4: non-targeting sgRNA 2.0 at 191 bp, S5: sgRNA SG6 at 216 bp, S6: non-targeting sgRNA SG6 at 216 bp, S7: sgRNA SG9 at 228 bp, S8: non-targeting sgRNA SG9 at 228 bp, S9: sgRNA SG12 at 240 bp, S10: non-targeting sgRNA SG12 at 240 bp, NTC: no template control. The samples were migrating for 20 min at 150 V and stained with ethidium bromide.

In **Figure 75**, the PCR products of the sgRNA inserts are visualized on a 4% agarose gel. For all sgRNAs, a PCR product at the expected size was obtained. The sgRNAs were purified by DNA purification kit **(VIII.6)** and used for AQUA cloning **(VIII.14.2)**.

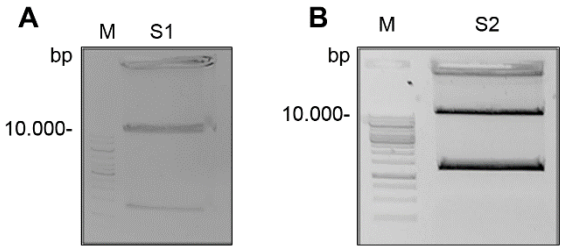


Figure 76: 0.8% agarose gels with linear PCR products of PB-UniSAM and PB-UniSAM-PAL. M: Thermoscientific GeneRuler 1kb DNA Ladder. A) S1: PB-UniSAM at 10147 bp. B) S2: PB-UniSAM-PAL at 13015 bp. The samples were migrating for 30 min at 130V and stained with ethidium bromide.

In **Figure 76** the PCR products of the inverse PCR of PB-UniSAM and PB-UniSAM-PAL were visualized. A PCR product at the expected size appeared, however in both cases a lower migrating product appeared. The PCR products migrating at 10.000 bp were cut out and purified **(VIII.6)**. The linearized and purified PB-UniSAM backbones were used together with the sgRNA inserts in AQUA cloning reactions **(VIII.14.2)**. Versions without sgRNA were also established as controls for MPH and PHP. In this case, the PB-UniSAM backbone was re-ligated using the techniques described in inverse PCR cloning **(VIII.14.5)**. Clones were sequenced using the hU6_seq_fwd primer. For sgRNA 2.0, SG6 and SG12 positive clones were identified, whereas for SG9 and all non-binding sgRNAs no correct clones were obtained (SEQ_ID: 048 - 050). For each of the no sgRNA controls a correct clone was obtained (SEQ_ID: 046 - 047).

The newly generated plasmids were tested in a cell-based assay, whereas HeLa cells were transfected using pGL3-Basic-8xsgRNA-eGFP and one of the PB-UniSAM plasmids **(VIII.19.1)**. After 24 hours of incubation, the cells were analyzed by flow-cytometry **(Figure 77)**.

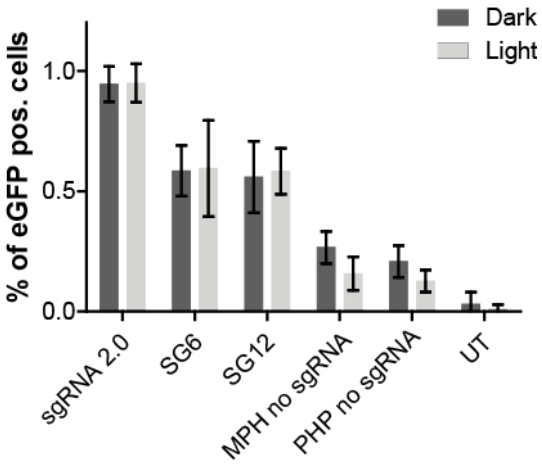


Figure 77: Light-dependent activation of eGFP transcription using pGL3-Basic-8xsgRNA-eBFP and PB-UniSAM derivates. The data is presented as mean values ± s.d (n = 8 from 4 individual experiments with two cell culture replicates).

The activation of eGFP transcription was successful using PB-UniSAM and sgRNA 2.0, whereas the corresponding MPH no sgRNA control only shows minimal induction of eGFP expression **(Figure 77)**. Using PB-UniSAM-PAL in combination with the expression of the sgRNAs SG6 or SG12, around 55% of the activation of sgRNA 2.0 was achieved, however without light-dependent characteristics. The corresponding negative control, which does not express sgRNA is comparable to the negative control of sgRNA 2.0 and close to the background. These findings indicate, that the design of PB-UniSAM is not yet suitable for LITA experiments, since no light-dependent activation was detected.

V.2.6 Replacement of aptamer 53 by aptamer 58 (NM60)

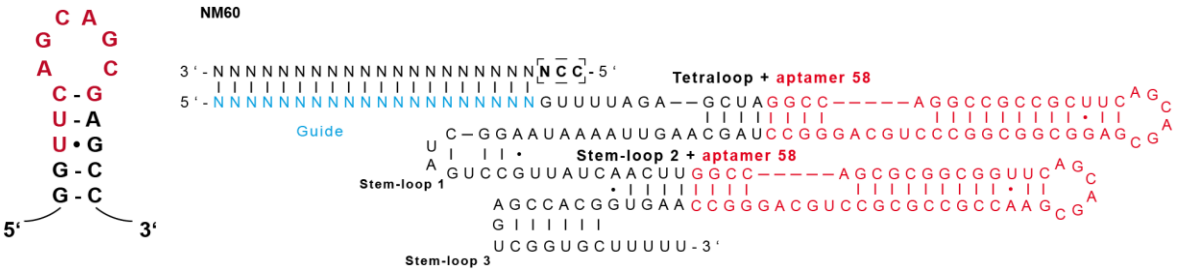


Figure 78: Aptamer 58 and the NM60 sgRNA. Sequence and motif (red) of the aptamer 58 (NM60) PAL aptamer variant on the left side. The sgRNA based on SG9 with aptamer 58 introduced in the tetra loop and stem-loop 2.

Since SG9 was the most potent sgRNA regarding the LITA assays, the SG9 scaffold was used and both aptamer 53 sequences were replaced by the aptamer 58 sequence (unpublished). The 5' GG and the 3' CC were omitted, since the aptamer is already embedded in the SG9 stem structure **(Figure 78)**. The name NM60 actually relates to a sequence from the original host *Nakamurella multipartita*, from which PAL was discovered. Here, the term NM60 is used equally to the term aptamer 58 as at the time the data was generated NM60 was also the name for the aptamer.

Analogously, to the cloning of SG9, a sense and an antisense DNA strand were ordered, hybridized and amplified to generate the SG9-NM60 insert **(VIII.14.5)**. The insert was cloned into linear pENTR.hU6.empty plasmid as described previously.

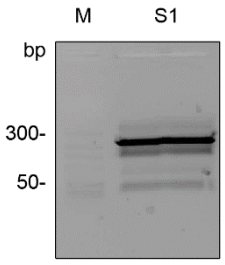


Figure 79: 4% agarose gel showing hybridized and amplified sgRNA insert SG9-NM60. M: Thermo Scientific GeneRuler ULR DNA ladder. S1: sgRNA insert SG9-NM60. The samples were migrating for 30 min at 140V and was stained with ethidium bromide.

In **Figure 79**, the PCR product of the full-length SG9-NM60 insert was visualized on an agarose gel. The PCR product was purified **(VIII.6)** and used with linear pENTR.hU6.empty for AQUA cloning **(VIII.14.2)**. A correct clone was obtained (SEQ_ID: 040).

The new sgRNA was tested in a LITA assay using dCas9, PHP and pGL3-Basic-8xsgRNA-eBFP with SG9 as control **(VIII.19.5)**.

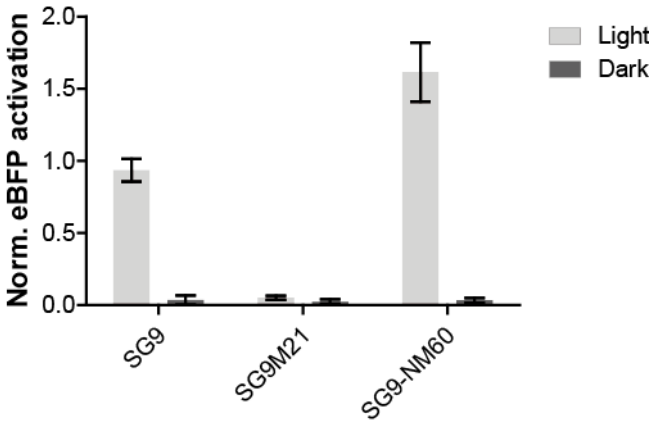


Figure 80: Light-dependent induction of eBFP comparing SG9 with SG9-NM60. The data is normalized to SG9 in light using min-max scaling (n = 6 from 3 individual experiments).

It was found, that the expression of eBFP for SG9-NM60 is 1.73x higher compared to SG9 in the light, while both sgRNAs remain at background level in the dark **(Figure 80)**. The control sgRNA SG9M21 does not induce eBFP expression neither in the light nor in the dark. These findings indicate, that SG9-NM60 might suited even better for future experiments as the up to now used sgRNA SG9.

V.2.7 Shuttle PAL

One idea of further optimizing LITA assays and to increase the dynamic range of the assay is to spatially separate the effectors (PHP) from the nucleus in the dark. Background activity of PHP can be explained by the random interaction of the transactivation domains of p65 and HSF1 with promoter regions, which might disturb gene expression. Also, since the dCas9:sgRNA complex is constantly bound to the target DNA strand, one has to account for background binding of PAL to the aptamers in the dark. To circumvent this, PAL could be engineered similarly to LINuS and LEXY[194]. An NLS-tag would be exposed from the protein in the light, directing the protein into the nucleus, whereas in the dark, the Jα-helix of PAL might mask the NLS signal. A weak nuclear export signal (NES) might direct PAL into the cytosol, where it cannot activate genes. Since eBFP activation was observed even when using a PHP variant which is not NLS tagged, it was hypothesized, that PAL by chance might contain some peptide signal, which acts in eukaryotic cells as NLS sequence.

MKVNRPAERASFGSFVLDAGSARFVGSDELALVLGFAPGDVVLTPAVVLA 50
HLHPDDRLEWQAGLQRCLATGRPVVVNHLLLTAEAEPRPAMTTLTALTEQ 100
DRVRAVTGVITDLSDRVRRATEAEIRQAVRAAAATRSEIDQAKGIVMAAF 150
DVDADQAFALLKWHSSQSNRKLRDLATGMIEGLAAANSALPLRRRLSTVF 200
TDMGCPAPSTKGWTVPVTDIGLPPTSGLIPTALLPGILTRAAHDASVAIT 250
VADVTAPDQPLVYANPAFERLTGYAAAEVLGRNCRFLQAESGDPHERSAI 300
RSAIANGDAVTTLIRNFRQDGHAFWNEFHLSPVRNGAGRVTHYIGYQLDV 350
TERVERDQQLEQLAS

Predicted bipartite NLS sequence:

Predicted bipartite NLS		
Pos.	**Sequence**	**Score**
169	NRKLRDLATGMIEGLAAANSALPLRRRLST	4.3

Figure 81: Amino acid sequence of PAL and results from the NLS-mapper analysis.

In **Figure 81**, the amino acid sequence of PAL was analyzed using the software tool NLS-mapper[195]. A bipartite NLS with a moderate NLS activity locating the protein in the nucleus and the cytoplasm was identified. To enhance the strength of the NLS

signal, known NLS sequences were searched in the literature and the amino acid sequence KRRRL was found to be a strong NLS with a algorithm cutoff score of 10.0. Thus, the L192K point mutation was introduced within PAL to increase the NLS-mapper cutoff score from 4.3 to 10.0. The NLS is located within a proline-rich adapter segment, which joins the ANTAR and LOV moieties together and also might be masked by the Jα-helix in the dark (see Weber *et al.* 2019, Fig. 2). The plasmid pmCherryPAL was used as a template in an inverse PCR reaction using PAL_L192K_mut_fwd and PAL_L192K_mut_rev as mutagenic primers. The PCR was set with Θ_{elong} as 180s and T_{ann} as 68°C for 10s running for 30 cycles. The PCR products were visualized on a 1% agarose gel **(Figure 82)**. The PCR product was reused for inverse PCR cloning. Clones were analyzed by DNA sequencing using the SV40-pA-rev primer (SEQ_ID: 051).

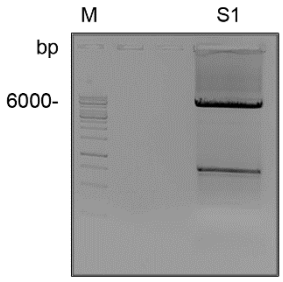


Figure 82: 1% agarose gel with PCR products of the inverse PCR to introduce the L192K mutation in pmCherryPAL. M: Thermo Scientific GeneRuler 1kb DNA ladder. S1: PCR products of pmCherryPAL_L192K at 5796 bp. The samples were migrating for 30 min at 140 V and was stained with ethidium bromide.

After the inverse PCR two PCR products appeared on the 1% agarose gel, from which the upper product runs at the predicted size of 5796 bp and was cut out and purified.

To test if the NLS was active in a light-dependent fashion, HeLa cells were transfected with pmCherryPAL and pmCherryPAL_L192K and analyzed the mCherry distribution in the light and in the dark by microscopy analysis **(Figure 83) (VIII.19.2)**.

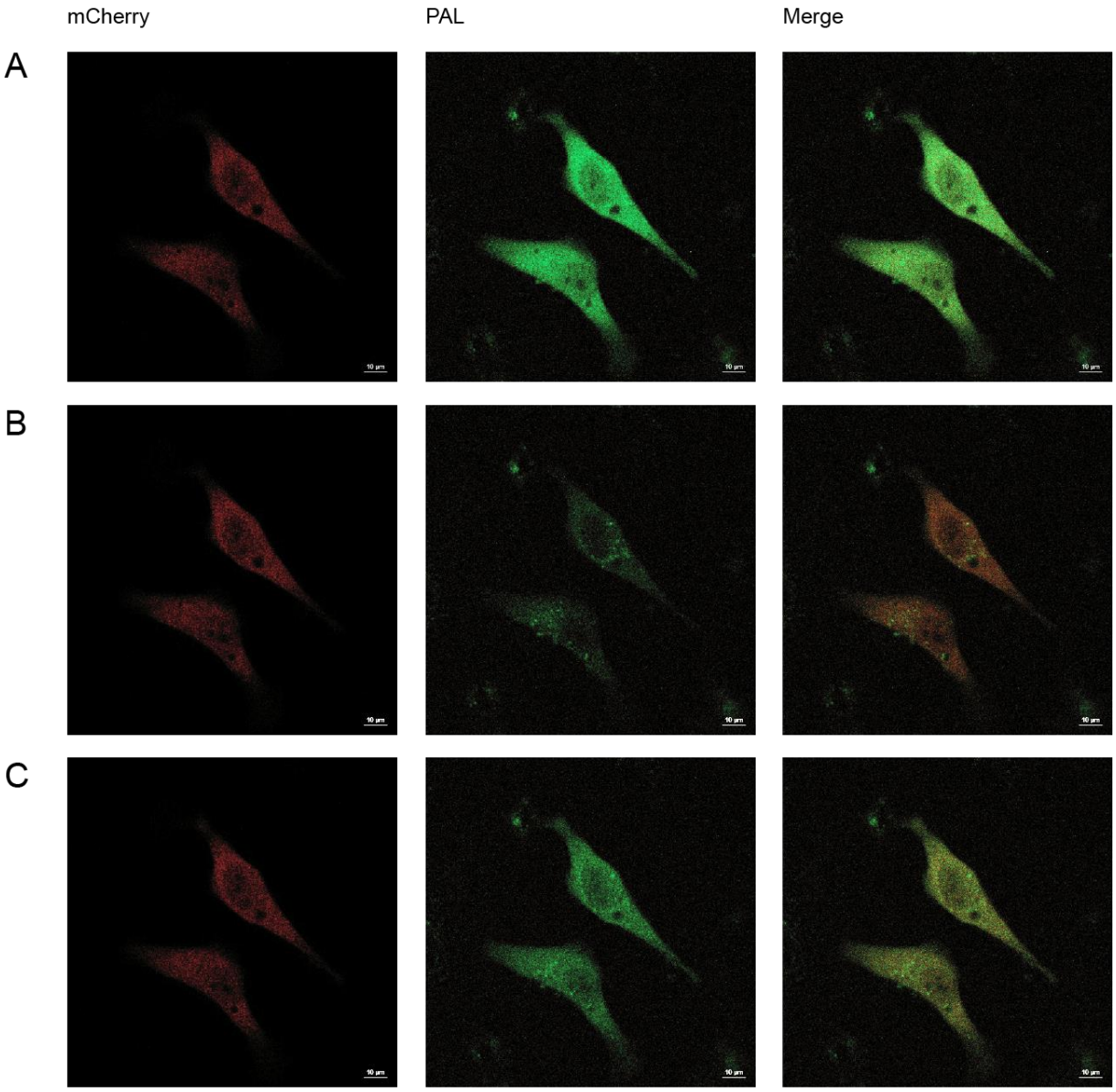


Figure 83: Switching activity and localization of mCherryPAL in HeLa cells. Images were taken in the mCherry channel (ex/em: 543 nm/596-696 nm) and for the visualization of FMN visible in the PAL dark conformation ex/em: 405 nm/488-529 nm. First an image was taken when cells were kept for at least 10 min in the dark(A). Then, the cells were irradiated with blue light to change the conformation of PAL, which leads to a loss of PAL fluorescence (B). After 10 min recovery in the dark, PAL switches back to the fluorescent dark conformation indicating that the fusion protein can undergo a complete photo cycle (C). Scale bar = 10 µm.

In **Figure 84** the switching activity and protein localization is depicted. The cellular location of the protein cannot be observed. However, most of the mCherryPAL protein is located in the cytosol.

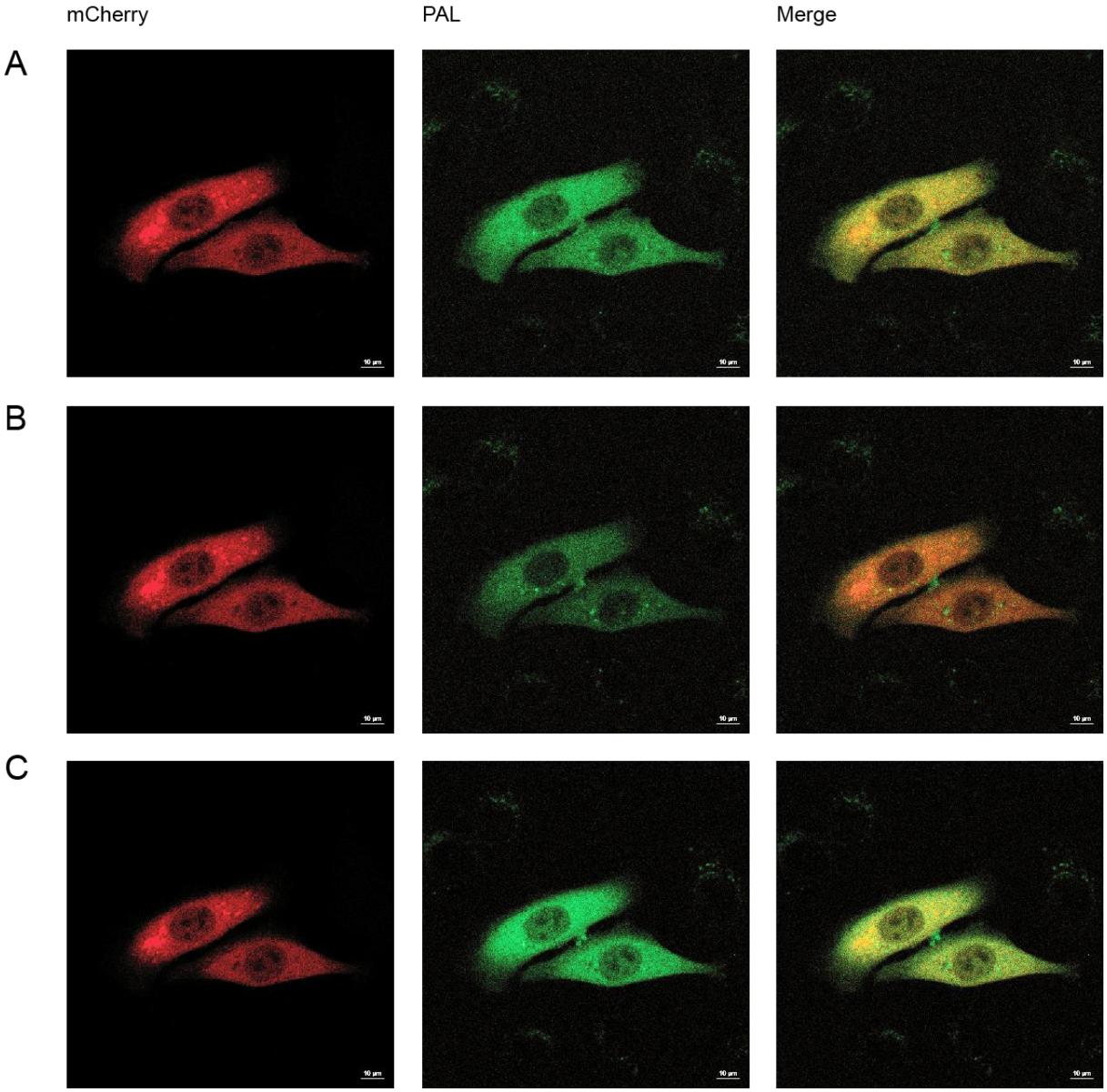


Figure 84: Switching activity and localization of mCherryPAL_L192K in HeLa cells. Images were taken in the mCherry channel (ex/em: 543 nm/596-696 nm) and for the visualization of FMN visible in the PAL dark conformation ex/em: 405 nm/488-529 nm. First an image was taken when cells were kept for at least 10 min in the dark (A). Then, the cells were irradiated with blue light to change the conformation of PAL, which leads to a loss of PAL fluorescence (B). After 10 min recovery in the dark, PAL switches back to the fluorescent dark conformation indicating that the fusion protein can undergo a complete photo cycle (C). Scale bar = 10 µm.

The switching activity and protein localization of mCherryPAL_L192K is shown in **Figure 84**. Also, for mCheryPAL_L192K no change in cellular localization was found, which could be detected by eye. However, the switching of the protein was visible.

VI Discussion

Since the successful selection of aptamers 04 and 53, which bind only to the light-conformation of PAL, a direct and conditional fusion between RNA structures and proteins was possible. This interaction enables for a whole set of light-dependent tools, that control cellular functions such as translation[166], short hairpin RNAs[196] or Cas9 activity and transcription as described in this work.

Light-dependent control of cellular functions enable to study dynamic processes such as gene expression, protein localization or cell signaling[197]. In order to alter genetic information within cells, programmable nucleases such as Cas9 are well suited tools, since they can be programmed to bind to specific locations within the target genome and optionally can be expressed these or can be transiently transfected[198, 199]. However, the constant expression of Cas9 can lead to unwanted effects, such as off-target effects in which a Cas9:sgRNA complex can bind to sequences within the genome, which are similar to the target sequence and induce cleavage in these regions[200]. Thus, the modulation of Cas9 activity, when expressed in human cells is highly beneficial[201, 202]. The nuclease function of Cas9 is dependent on the complex formation with the sgRNA, which can be modulated to alter Cas9 activity[89]. The activity of the Cas9 protein can be modulated using chemical [163, 203-208] or optical [11, 209, 210] inducers. This approach always modulates all Cas9 molecules and does not allow for a target specific modulation if multiplexing is applied. Another approach is the modulation of sgRNAs, which can be target specific. Systems using ligand-induced[211, 212] or ribozyme-based[213] modulators were already published. Also, proteins, that modulate sgRNA/Cas9 complexation[205] and anti-sense technology[214] were used. However, none of these systems enable for light-dependent, reversible and fully genetically encodable control of Cas9 activation, while maintaining the usage of multiple sgRNAs at the same time.

Native sgRNAs were extended with the aptamer 04, which binds specifically to the light-conformation of the photoreceptor protein PAL[166]. It was hypothesized, that the binding of PAL to the aptamers will prevent the sgRNAs from complexation with Cas9 and thus blocking the function of Cas9 activity. The system is genetically encodable

and sgRNAs can be multiplexed with and without aptamers, to allow for partially multiplexed and light-dependent Cas9 activity.

VI.1Light-dependent Cas9 repulsion

Cas9 protein was expressed *in vitro*, which was needed for *in vitro* assays, confirming the light-dependent binding of aptamer modified sgRNAs in Cas9 cleavage assays **(Figure 14)**. Cas9 was identified at a molecular weight of 160kDa. However, also impurities were detected. These impurities most likely origin from *E. coli* host proteins, which show affinity towards the Immobilized Metal Affinity Chromatography[215]. The expression of Cas9 was mainly based on the protocols published by Carolin Anders and Martin Jinek[186], while it was decided to only express his-tagged Cas9 and resign on removing affinity tags from the protein and further purification by cation exchange chromatography, since it was not planned to inject the protein into mammalian cells. Rosetta cells might not be the optimal choice for the expression of Cas9[216]. Instead of using Rosetta 2 DE3 cells, DE3 RIL cells were used, that express extra tRNAs for arginine, isoleucine and leucine. The extra tRNAs are beneficial, since the native host of Cas9, *Streptococcus pyogenes*, has a GC content of 38.5 ± 0.1%[217] compared to *E. coli* with a GC content of ~50%. It was hypothesized, that the minor impurities will not affect the cleavage reaction and the effect would be equal to all samples, thus a further purification was not considered. Further, it was found, that still a remarkable amount of Cas9 protein remains in the pellet after sonication and centrifugation **(Figure 14 A, lane Pe)**, which was seen as an option for a later extraction, if more protein was needed. Elution fractions 1 and 2 were pooled **(Figure 14 B)**. Cas9 was detected by immune precipitation **(Figure 15)**. It was found to bind several proteins from the lysate and the pellet. However, these reactions probably occur from an overload of protein onto the SDS polyacrylamide gel. The load fraction, which contains diluted lysate does not reveal unspecific binding of the antibody to cellular proteins, while a clear band at a molecular weight of160 kDa is detected. The detected protein migrates in line with the positive control Cas9 (PC). Comparing the performance of the antibody with published western blots, reveals similar results as found for the

expression in *P. aeruginosa*[218], where the same antibody was used. Comparing for cross reactions of the antibody towards other proteins is not possible, since it is a common practice that western blots are cut above and below the Cas9 signal at 160 kDa and do not allow for an insight in unspecific binding of the antibody. The protein concentration was determined using a Coomassie® Brilliant Blue G-250 dye-based assay, resulting in a total concentration of 400 µg/ml **(Figure 16)**. As reference, a BSA standard solution was made. The assay used in this study depends on charged amino acid side residues but also hydrophobic amino acids like tryptophan residues and leads to an increase in light absorption from 465 nm to 595 nm when binding to protein occurs. Although, BSA is a cheap and readily available protein, the binding sites for hydrophobic substances might alter the assay response and give more intensive staining than it should[219]. The detection range of the assay is at 0.2 to 0.9 mg/ml as stated in the manufacturer's manual. Alternatively, the Bradford assay has a detection range around 0,2 – 20 µg/ml[220]. As the elution samples were diluted 1:10 a total of 40 µg/ml was measured, which is below the detection range given by the manufacturer. These findings indicate, that an uncertainty of the concentration can be expected. To examine the amount of error for the measured concentration two factors come into play. On the one hand, within the cleavage assay, the protein can be compared to commercial Cas9. On the other hand, the cleavage reaction is a dead-end reaction, in which higher Cas9 concentrations will increase the speed of the reaction but not change the final result of the assay, since the cleavage reaction is not reversible and thus, slightly higher Cas9 concentrations should not affect the results at first.

It was decided to target the CDS of mCherry **(Figure 17)**, since this would also allow for a knock-out of the fluorescent protein in transfected mammalian cells. For optimal target strand selection the mCherry CDS was analyzed by the ChopChop algorithm [221] **(Figure 18)**. The selection of a good seed sequence is based on the composition of the PAM site and the seed sequence itself. Whereas for PAM usually NGG is used, the first 5-8 nucleotides next to the PAM site are the most important nucleotides for binding specificity[222]. The sequence identified by ChopChop starts with a guanidine, which makes it suitable for *in vitro* transcription using T7 RNAP[223].

We also included a non-binding control sgRNA, which was scrambled in the stem loop 1 region **(Figure 18)**. The analysis of the crystal structure of Cas9/sgRNA and target

DNA shows, that stem loop 1, which is primarily recognized by the REC lobe and the PI domain[89], is crucial for the function of the complex[89]. The DNA templates for native sgRNA and scrambled sgRNA were amplified by PCR leading to the expected products at a size of 127 bp **(Figure 19)**. After transcription and purification, the sgRNAs both with a size of 103 nt were obtained **(Figure 20)**. Next, the cleavage substrate was generated. As cleavage substrate pmCherry plasmid **(Figure 21)** or PCR amplified mCherry CDS was used **(Figure 22)**. The plasmid was first linearized using a restriction digest as suggested by Anders and Jinek[186]. In order to perform the Cas9 cleavage assay two protocols were used, from which sgRNAs as RNA component were used and agarose gels were used to evaluate the cleavage fractions[86].

Then, the sgRNAs, Cas9 and the cleavage substrate were tested in a Cas9 cleavage assay in order to verify the activity of Cas9 enzyme **(Figure 23)**. Since aptamer 04 was planned to get inserted into the sgRNAs, the cleavage activity was tested in various buffers. One of the buffers is the intracellular buffer, which was used to select the aptamers for PAL[166]. The other two buffers are the commercial Cas9 buffer provided with the commercial Cas9 protein from NEB and a 5x Cas9 cleavage buffer from Anders and Jinek[186]. Cleavage was detected in all three buffers **(Figure 24)**, indicating, that ICB can be used for future Cas9 cleavage assays.

The design of the sgRNAs with incorporated PAL aptamers is crucial for the correct folding and binding to the Cas9 protein. Aptamers like the MS2 aptamer or the PP7 aptamer were already introduced in sgRNAs[118, 224, 225]. The strategy to insert aptamers into PAL was adapted from structure-guided approaches like the SAM system[118]. It is shown, that stem loop 1 has several interactions with the REC lobe and the PI domain of the Cas9 protein. Stem loop 3 has several but compared to stem loop 1 fewer interactions with the RuvC domain of Cas9, whereas stem loop 2 has only two interactions at the base of the stem formation, namely A68 with His1349 and Ser1351 and G81 with Lys33 and Tyr1356[226]. Since the tetra-loop is an artificial linker[86], it is not known to interact with the protein scaffold of Cas9. It is also reported, that stem loop 1 is essential for the correct and functional formation of the Cas9-sgRNA complex, whereas stem loops 2 and 3 support the stability of the complex[89]. Thus, the modification of stem loop 2 and the tetra-loop are preferred sites for aptamer insertion and stem loop 1 was not touched. As aptamers, the 04 PAL aptamer (5'-UUGAAGCAGACGA-3') and a non-binding variant (5'- *AACUUC*CAGACGA-3')

(Figure 25 and Figure 26)[166] were used. The aptamers were introduced as a monomer or a dimer into the stem-loops of the sgRNAs. For example, the stem loop 3 was replaced by a 04 aptamer dimer (sg04.19di7). The sgRNAs were extended in the linker region (instead of 19 nt, 23 nt were used) and another aptamer 04 was added in the tetra-loop (sg04.23t7). The design approach of the sgRNA 2.0 and the SAM system was used as a basis and monomeric aptamers were introduced into the tetra-loop and stem loop 2 (sg04.23di), in the tetraloop only (sg04.23) or additionally attached it to the 3' end (sg04.23tri). Further, a construct was tested, where the tetra-loop remains unchanged but aptamers were inserted into stem loop 2 and the 3' end. To have at least two aptamers in the sgRNA is explained by the dimeric characteristic of PAL in the light conformation[166]. All sgRNA DNA templates were successfully amplified **(Figure 27)** and used for *in vitro* transcription. The *in vitro* transcription resulted in the transcribed sgRNAs, including native sgRNA and scrambled sgRNA as controls **(Figure 28)**. The amount of transcribed RNA varied between the constructs with sgRNA sg09.23t7 having the lowest concentration. It is known, that transcripts with a strong secondary such as sgRNAs can inhibit their own transcription[227]. However, the amount of inhibition will vary between all constructs. Next, the sgRNAs were tested in a Cas9 cleavage assay **(Figure 29)**. It was found, that cleavage for the positive control **(Figure 29, S1)** but not for the negative control scrambled sgRNA **(Figure 29, S2)** occurred**.** Cleavage was detectable for sgRNAs sg04.19di7, sg04.19tri, sg04.23di, sg04.23tri and sg04.23, as well in a very low amount for sg04.19dmu. Interestingly, the mutated variants do not induce cleavage, although the mutation should only prevent PAL binding to the aptamer. Additionally, the sgRNA sg04.23t7 did not induce cleavage. The sgRNA sg04.23t7 is a derivate of sg04.19di7, which induced cleavage. The difference is the insertion of the aptamer in the tetra-loop. The insertion of an aptamer in the tetra-loop in combination with a stem loop 3 substitution might lead to a destabilization of the sgRNA secondary structure, which renders the complex with Cas9 unable to cleave a substrate. This hypothesis is backed by findings, that modified 3' ends of sgRNAs can lead to a reduced activity in genome editing[228]. However, some constructs induce cleavage and are candidates for a light-dependent cleavage assay.

The cleaving sgRNAs were tested in a light dependent cleavage assay including PAL. Since it was hypothesized, that PAL binding to the sgRNAs will prohibit sgRNA binding to Cas9, it was expected to see a difference in the cleavage patterns, when samples incubated in the light were compared to samples incubated in the dark. All in all, no

differences in the cleavage patterns were observed when comparing the cleaved fractions from the light and the dark **(Figure 31 and Figure 32)**. A control gel without PAL was used, to analyze and compare cleavage differences, which might depend on the presence of PAL **(Figure 30)**. A hypothetical sgRNA stabilizing effect for sgRNA sg04.19tri was found, which shows slightly enhanced cleavage in the light **(Figure 31 A, S4)**, compared to the control gel without PAL **(Figure 30, S4)**. However, in a previous experiment, the full activity of sgRNA sg04.19tri was demonstrated **(Figure 28)**, which was not found in the control gel **(Figure 30)**. This assay-to-assay variation was also detected for sg04.19dmu, which did not induce cleavage before **(Figure 28)**, but shows cleavage for **Figure 31 A**. Since the mutant was tested to not interact with PAL (Dissertation Anna Maria Weber, 2021), it was concluded, that the presence of PAL is not the reason for the different characteristics in cleaving but is more probably caused by folding problems of the RNA. Having a deeper look into the sgRNA/Cas9 kinetics, it was found that Cas9 typically has an equilibrium dissociation constant (K_D) of of 10 ± 2 pM when binding to the sgRNA handle[183]. For aptamer 04.21 a K_D of 253 ± 12 nM was determined at 37 °C in the light and measured by surface plasmon resonance[166]. These findings indicate, that PAL would probably not be capable of blocking Cas9 from binding to the sgRNA or re-pulse Cas9, when already in complex with sgRNA. In order to do so, the complexation of sgRNA and Cas9 needs to be disturbed by the mutation of sgRNA nucleotides. A system using the theophylline aptamer was published, in which the binding of theophylline to the aptamer extended sgRNA reconstitutes the folding of the sgRNA and rendering it capable of binding to Cas9[229]. Similarly, small molecule-activated allosteric aptamer regulating (SMART)-sgRNAs can be used, where a ligand restores the folding of the sgRNA[105]. Also, the inclusion of the aptamer at the 5' end of the sgRNA, where a blocking strand prevents the sgRNA seed region to hybridize with the target DNA might be a strategy. Upon binding of the ligand to the aptamer a communication module will unlock the blocking strand and make the seed sequence accessible. This approach was examined by Bachelor student Nemanja Stijepovic in our laboratory (Nemanja Stijepovic, Bachelor Thesis, 2019). Finally, the project idea to repulse Cas9 from the sgRNA was not continued, mainly due to K_D differences between the aptamer/PAL complex and the sgRNA:Cas9 complex. A new strategy, where PAL can be used in coexistence of the Cas9/sgRNA complex and still is able to offer a light dependent functionality for the complex was found.

VI.2 Light-dependent activation of transcription

A new aim was to generate a tool, that allows for light-dependent transactivation of transcription. Instead of using Cas9, which induces cleavage (and thus is an uni-directional reaction), a cleavage deficient Cas9, dubbed dCas9, was used. The cleavage deficient Cas9 contains two silencing mutations, one in the RuvC1 and one in the HNH nuclease domain (D10A and H841A)[86]. dCas9 and sgRNA can be used as molecular navigation system to direct molecular functions to pre-programmed sites within the genome. One possible molecular function that acts on DNA is transcription, which can be achieved by recruiting transcriptional activators in proximity of promoter regions[116, 127]. The technique, using aptamer-modified sgRNAs in combination with dCas9 and transcriptional activators was first used for the SAM system[118]. In order to choose between aptamer 04 or aptamer 53 and to confirm the binding of cellular expressed mCherryPAL to aptamer 04 and aptamer 53, both aptamers were assessed in a mCherry pulldown assay **(Figure 32)**.

Cellular expressed mCherryPAL binds to aptamer 53.19 in the light but not in the dark, while aptamer 04.17 does not bind to mCherryPAL neither in the light nor in the dark. Interestingly NLS-tagged mCherryPAL binds to aptamers 04.17 and 53.19 in the light, however, background binding the dark was also enhanced. The slightly lower binding of NLS-mCherryPAL can be explained by an uncomplete sonication process, leaving nuclei of mammalian cells intact, which contain NLS-mCherryPAL protein. These intact nuclei are removed during a centrifugation step, leading to a lower concentration of NLS-mCherryPAL compared to mCherryPAL. It is described, that incomplete sonication can be used to collect intact mitochondria[230] and it was hypothesized, that similar effects may be true for the nuclei of mammalian cells. The finding, that mCherryPAL does not bind to 04.17 in the light was not expected and it can only be hypothesized, that either the concentration of mCherryPAL was too low or the protein was in the wrong conformation. In SPR measurements an aptamer 04 derivate 04.21 was found to bind to PAL with a K_D of 253 ± 12 nM at 37 °C and illuminated conditions. In contrast, aptamer 53.19 was measured to have a a K_D of 153 ± 14 nM at the same settings. These differences are probably not responsible for the findings obtained in the pulldown assay. It might be, that mCherryPAL was in a wrong conformation, when

performing the measurement. A problem with the RNA can be excluded, since binding of NLS-mCherryPAL was detected for aptamer 04.19. Based on these findings, aptamer 53 was further considered for insertion in sgRNAs, since the dark/light difference was strongly pronounced for the mCherryPAL without NLS.

In an adapted approach, the sgRNAs were modified with aptamer 53.19[166] and PAL was fused to the transcriptional activators p65 and HSF1. The adapted and light-dependent, aptamer-based transcription activation system shall be used to conditionally activate the transcription of endogenous genes within mammalian cells. The published SAM system was used as a model and as positive control and benchmark[118].

The sgRNAs were designed in order to bind to the 8xsgRNA binding site of pGL3-Basic-8x-sgRNA-eBFP (5'-AAAGGTCGAGAAACTGCAAA-3')[161]. For the *in vitro* transcription a GGG trinucleotide was added in front of the sequence, though already a GG dinucleotide greatly enhance the transcription[231]. The 5' added G's are not detrimental for performance in mammalian cells, since the sgRNA is trimmed after transcription[232] and exceeding 20 nucleotides of seed was shown to not increase the efficiency of the system[233]. Since it was planned, to express the sgRNAs in mammalian cells, guanine needs to be added to promote transcription by the U6 promoter[234], which is also true for the mouse U6 promoter[235]. Using the H1 promoter, however, would enable the usage of guanine and adenine as a starting nucleotide[236].

A broader variation of sgRNAs was screened, thus, it was decided to construct three sgRNAs with varying stem lengths. The sgRNAs were derived from the sgRNA 2.0[118], by replacing the MS2 aptamer with the aptamer 53[166]. To modify the stem length GC regions were introduced, that stabilize the aptamer as shown in **Figure 33** and **Figure 34**. It was hypothesized, that stems, which are too short might be too much imbedded in the Cas9 protein scaffold and thus inhibiting PAL from binding to the aptamer by electrostatic repulsion of the amino acid side residues[237]. The correct secondary structure folding of the sgRNAs was confirmed using the mfold software[238-240]. Since PAL forms a dimer, when irradiated with 447 nm[166], always two aptamers were introduced into the sgRNAs, whereas one aptamer was inserted in the tetra-loop and the other aptamer in stem loop 2. The stem length of the sgRNAs also led to a turn in the orientation of the aptamer. An A-form RNA helix completes a full turn every 11-12 nt, while extending the stem by three nucleotides leads to a quarter turn[241]. The

sgRNA templates were amplified by PCR **(Figure 35)**. Minor residues of sense or antisense strands were detected in the gel, however they were not removed, since during the RNA purification a DNase treatment is performed and correct sized sgRNA is isolated by gel extraction. The transcription of half-sized sgRNAs is unlikely, since the T7 promoter is more active when a double strand template is used, which is only true for the full-length sgRNA template and ssDNA template might tend to form inhibiting secondary structures. All sgRNAs were successfully transcribed **(Figure 36)**, however low yields were found for negative control sgRNA SG9, sgRNA SG12 and negative control sgRNA SG12. For sgRNA SG12 also a shorter product was detected which might be a result of secondary structure formation **(Figure 36 A, S5)**.

We continued, testing the sgRNAs in a Cas9 cleavage assay and normalized the cleavage activity to the native sgRNA **(Figure 37)**. A slightly reduced cleavage activity for sgRNA 2.0, SG6 and SG6 and heavily reduced cleavage activity for SG12 was observed, whereas the negative control sgRNAs remained at background level. The highest activity for native sgRNA can be explained by having not introduced any modifications to the sgRNA handle. It is hypothesized, that over the millions of years and by mutation, the interaction between Cas9 and the tracrRNA handle is optimized due to natural selection, leading to a best fitting complex formation. This might be still true, although the sgRNA is a synthetic concatenation between the crRNA and the tracrRNA. Although the modifications were introduced at the tetra-loop and stem-loop 2, there are still minor interactions (Lys33 with the RuvC domain and His1349 with the PI domain) of the stem region of stem-loop 2, which might be disturbed by extending the stem[89]. The weak cleavage efficiency of SG12 is probably cause by wrong folding or incomplete transcription, which might be valid explanation by looking at the gel picture of **Figure 36 A S5**. The RNA is probably not a secondary structure of SG12 but a fragmentary product of transcription. The *in vitro* transcription was incubating at 37 °C, however to transcribe sequences with strong secondary structures it might be an option to run the transcription at a lower temperature[d].

To investigate and optimize the characteristics of light-dependent transcription activation using sgRNAs modified with aptamer 53, it was decided to use a flow cytometry-based assay. Flow cytometry allows for the rapid measurement of

[d] thermofisher.com/de/de/home/references/ambion-tech-support/large-scale-transcription/general-articles/practical-tips-for-in-vitro-transcription.html, 21.09.2017

transfected single cells and thus reveals information such as the transfection efficiency and the population-based homogeneity of fluorescent cells[242]. These features cannot be measured using cell extracts and western blot analysis. As reporter protein, eBFP2 was chosen, since PAL might induce a false signal in the eGFP channel and mCherry was already blocked for co-expression with PAL.

First, the reporter plasmid was modified, in order to express eBFP instead of eGFP. This was due to the fact that the PAL fluorescence might interfere in the eGFP channel. The eBFP insert and the pGL3-Basic backbone were successfully amplified, visualized on an agarose gel **(Figure 38)** and cloned. Similarly, the MS2-p65-HSF1-T2A-eGFP cassette from the lentiviral plasmid backbone was cloned **(Figure 40)** into the plasmid backbone of pEGFP-N1 **(Figure 39)**, in order to have the photoreceptor proteins and the MS2-based construct in the same plasmid framework.

To keep the cloning of PHP easy, the same order of transactivation domains (p65-HSF1) as found in MPH was used for PAL. In CRISPRa, a whole set of transactivation domains was already used to boost the expression of genes. The most important CRISPRa systems to date are SAM[118], VPR[130] and SunTag[128, 243]. In a direct comparison, for human ASCL1 and NEUROG, it was shown that SAM is the most potent system[244]. Since the presented system is adapted from the SAM system, the transactivation domains of p65 and HSF1 were used. PAL was attached to the C-terminal end of p65-HSF1, since the C-terminus of PAL holds the LOV domain and the Jα helix[166], which are involved in photo switching of the protein. T2A-mCherry was attached to the C-terminal end, hypothesizing, that the self-cleaving T2A peptide will separate mCherry from the Jα helix of PAL. 2A-peptides are viral peptides, which cleave multicistronic sequences subsequently after their translation[245]. Additionally, an NLS was added to the N-terminus of the fusion protein to direct the transactivating components into the nucleus of the cell. The construction of PHP was divided into three fragments, fragment 1 (F1) as NLS-p65-HSF1, fragment 2 (F2) as PAL and fragment 3 (F3) as T2A-mCherry. As for MS2 and p65, between HSF1 and PAL, a 6xGS flexible linker was introduced to enhance the orientation of the transactivation domains towards promoter elements[246]. All three fragments were designed to have a 30 bp overlap for homology-based cloning and were successfully PCR amplified **(Figure 41)**. However, some by-products in the gel for F1 and F2 **(Figure 41, S1)** were detected. Cloning methods were prioritized, which do not rely on restriction sites and recombine

sequences based on homology. One approach is Gibson Assembly, in which homologous ends recombine by using a 5' exonuclease, a DNA polymerase and a DNA ligase[247]. It was hypothesized, that a small fraction of correctly assembled F1, F2 and F3 can be amplified in a PCR using terminal primers **(Figure 42, S2)**. The PCR led to multiple PCR products with a weak product at ~ 3000 bp. The Gibson Assembly reaction was analysed and un-spliced F1 and F2 was identified. These results indicate, that the Gibson Assembly reaction was not optimal. Probably the high GC content in the overlapping regions, especially the GS linker, which has a GC content of 67% might form secondary structures[e] and prevents the 5' exonuclease from digesting the DNA from the 5' end. The isothermal incubation at 50°C allows complementary overlaps with high GC contents to form mismatched pairings, which reduce the assembly efficiency[248]. The PCR product in **Figure 42 S2** was isolated, however the concentration was too low to continue working with it. Another approach is splicing by overlap extension (SOE)-PCR[249]. First the mixture of the fragments is denatured and while cooling, the homologous regions of the fragments can anneal with each other. In a second step, terminal primers amplify the full-length spliced construct. F1 + F2, F2 + F3 and F1 + F2 + F3 were combined **(Figure 43)**. For all three reactions, by-products were observed, resulting from mismatched hybridization during the annealing step. However, products with the size of the fused fragments were found for all reactions as indicated by the arrow **(Figure 43)**. These products were isolated, however the concentration of isolated sample S3 was too low for further processing. Thus, a third cloning method called advanced quick assembly (AQUA) cloning was tested[189]. In AQUA cloning, all fragments and the plasmid backbone are mixed in a single reaction without enzymes. The reaction is transformed into competent *E. coli* and host repair enzymes use the homologous regions of the fragments to splice them into a single construct. Sequencing resulted in two correct clones.

Before testing all components in cells, the expression, localization and functionality of PHP was analyzed in mammalian cells. PHP was expressed in HeLa cells and PAL was located in the nucleus of HeLa cells **(Figure 44)** by its dark fluorescence[166], while mCherry was ubiquitously distributed within the cell. This result was expected, since mCherry is separated from PHP by a self-cleaving T2A peptide. Switching activity was confirmed, by recording a photocycle of PAL. The HeLa cell shown in the pictures is

[e] biocat.com/bc/files/Gibson_Guide_V2_101417_web_version_8.5_x_11_FINAL.pdf, 23.01.2021

trinucleated. Multinucleation in HeLa cells is described in literature[250]. Although a clear separation of PAL and the pretended cytosol is detected, a final evidence using a nucleus staining agent such as 4′,6-diamidino-2-phenylindole (DAPI)[251] or Hoechst 33342[252] would support our findings.

To remove constant activation of transcription provided by VP64, which is fused to dCas9, the VP64-tag was successfully removed from the fusion protein dCas9-VP64-T2A-eGFP by inverse PCR **(Figure 45)**.

The sgRNAs SG6, SG9 and SG12, together with the controls native sgRNA and sgRNA 2.0 were cloned into pENTR.hU6 plasmids. The sgRNAs were first fused together from two fragments **(Figure 46)**. The by-products are non-fused sense and antisense strands and the lowest band probably belongs to the primers used in the reaction, since they also appear in the NTC. Hybridization was not complete, since the reaction was not optimized. However, during the PCR only fully hybridized strands can be amplified due to the primers, binding at the terminal sites of the assembled sgRNAs. pENTR.hU6.empty was successfully amplified by inverse PCR **(Figure 47)** and the sgRNAs were inserted by AQUA cloning. AQUA cloning, or assembly cloning in general, can be prone to mutations, especially recombination, when the overlaps are not carefully designed. Ideally the overlaps contain a length between 16 bp and 32 bp[189]. However, this does not consider the GC content and the melting temperature of the homologous regions, since it was found, that altering the binding strength of the homologous overlaps results in an increased recombination[190]. Finally, the design of the homologous regions is always restricted to the sequences present in the locus. The fusion of GS linker regions for example is prone for recombination, since the amino acids are highly repetitive and GC rich. As an example, the linker sequence in PHP used between HSF1 and PAL has a GC content of 77%. Thus, even clones that might have only a single point mutation can be corrected using inverse PCR, without introducing new recombination events as for example shown in **Figure 48**. For each PAL aptamer modified sgRNA, a negative control for the aptamer domain was generated by exchanging a loop residing Adenine to Uridine leading to the M21 point mutation which was shown not to bind to PAL anymore **(Figure 49)**[166].

Then, dCas9, sgRNAs, pGL3-Basic-8xsgRNA-eBFP and PHP or MPH were tested in a light-dependent activation assay in HeLa cells **(Figure 50)**. As expected, the MS2 system (indicated by MS2) in combination with sgRNA 2.0 showed light-independent

activation of eBFP expression as detected by flow cytometry. The results indicate, that this combination can be used as a positive control in further assays. The combination of SG6 and PHP shows ~60% activation in the light compared to MS2 and sgRNA 2.0, whereas no activation was detected in the dark. The cross-specificity of these systems was tested by pairing MS2 with SG6 and PHP with sgRNA 2.0. For MS2 and SG6 no activation was found, indicating, that the MS2 protein does not specifically interact with aptamer 53 inserted in SG6. The MS2 aptamer consists of a minimum 21 nt element, composed of a stem region including a single bulged nucleotide, a four nucleotide loop and a single stranded tail[253], while only a conserved adenine at position -4 is required for tight binding[254]. For the MS2 coat protein no RNA binding domain was found, when a ProSite Scan was applied[255]. For PAL, the already identified ANTAR domain, which is an RNA binding domain, was identified. Thus, it is hypothesized, that the interaction of the MS2 aptamer and MS2 is specific in terms of the characteristic amino acid side residues of the protein without interacting to a designated RNA binding domain, whereas aptamer 53 likely binds to the ANTAR domain of PAL. The interaction between PAL and MS2 aptamers can also be explained, since ANTAR domains have a higher affinity to dual hairpin RNA structures[256], which is the case for the sgRNA harboring two aptamers. Initial experiments for PAL showed a attenuated and unspecific interaction towards RNA, which is mainly independent of light[166]. Usually, RNA binding proteins have only a single RNA substrate, which they bind with high affinity[257]. A slightly enhanced binding of PAL to MS2 aptamers in the light can be explained by the unmasking of the ANTAR domain upon conformational change in the light[166]. Our findings indicate, that SG6 in combination with PHP is capable of activating the expression of eBFP in the light, whereas eBFP expression in the dark is at background level.

Next, the plasmid ratios were optimized **(Figure 51)** and the best ratio was identified for 2:1:1:1 (dCas:effector:reporter:sgRNA). Comparable ratios were found in literature, where mainly the dCas9 component was transfected using a higher plasmid ratio[11, 64]. This is probably due to the large size of dCas9, which needs longer for its expression and thus over time leads to a smaller molarity of protein within the cell. This lack of molecule density is compensated by a higher amount of dCas9 plasmid, which allows for an increased parallel expression of the protein. However, the amount of eBFP activation compared to sgRNA 2.0 is still not equal, which indicates, that further optimization is necessary.

The effect of the aptamer stem-length was analyzed in a LITA **(Figure 52)**, and it was found, that the amount of eBFP expression in the light is highest for a stem-size of 9 nt (SG9), whereas the amount of eBFP expression for SG6 and SG12 were lower, with SG6 showing the lowest amount of eBFP expression in the light. The corresponding point mutants were included, which, as expected, do not induce an increase of eBFP expression. The different amounts of eBFP expression in the light and in relation to the stem length are probably caused by protein repulsion and positioning effects, as described earlier. SG6 is probably too short, which does not allow PAL for tight enclosure of the aptamer. SG9 induces the highest amount of eBFP expression in the light, whereas SG12 shows a slightly decreased activity. These findings indicate, that not only an increase of aptamer stem length is necessary for optimizing the eBFP expression in the light. It was hypothesized, that the benefit of SG9 in terms of eBFP expression in the light might result from stabilizing protein-protein interaction between PAL and dCas9, whereas a stem length of 12 nt generates a gap between PAL and dCas9, which is too big for the stabilizing protein-protein interaction. Since the RNA undergoes a ~¼ turn with every stem extension of 3 nt, for SG6 and SG12 the protein-protein interaction might even be contra productive, resulting to lower amounts of eBFP expression in the light. It is unlikely, that the effector orientation towards the DNA is altered by the varying stem-length, since the GS linkers offer full rotational freedom[258]. It was also recognized, that the relative amount of eBFP expression in the light for i.e. SG6 compared to sgRNA 2.0 was lower (~25%) in **Figure 52** as in the previous experiment shown in **Figure 51** and a ratio 2:1:1:1 (~60%). Analyzing the raw shift data of the flow cytometry experiments, it was noticed, that the shifts within the experiment vary by a factor of 4. These variations can come from different sources. The first variation can come from the counting of the cells in the Neubauer chamber. Errors associated with using the Neubauer chamber are described in literature and range from handling errors[259], the amount of measurements and cells[260] and the distribution of the cells during measurement[261]. Further, the quality of the plasmid DNA, which can vary between plasmid batches, impacts the efficiency of the transfection. After plasmid isolation, bacterial host cell contamination can remain the sample[262, 263]. In order to minimize the effect of these errors, experiments are always conducted multiple times with replicates. Further, the results were normalized to make them comparable to each other. Another influence can originate from the transfection itself. For the assay a quadruple transfection is applied. There are certain chances, that cells are transfected

with each plasmid in a perfect molar ratio, which do not occur in all cells. Since doublet cells are removed by gating, highly eBFP expressing cells might still origin from cells, that enter the apoptotic stage and thus change their morphology by reducing their cellular volume and thus increasing the intracellular concentration of eBFP[264]. Further differences in the amount of eBFP expression can be explained by the underlying kinetics between PAL and MS2. Whereas the MS2 coat protein binds to the MS2 aptamer with a K_D of 0,2 – 3 nM[265], PAL binds to the aptamer 53 with a KD of 153 ± 14 nM (at 37 °C in the light, measured by SPR) and 17 ± 2 nM (measured by fluorescence anisotropy)[166]. Further, PAL can change between two conformations, which means during the equilibrium in an illuminated setup, PAL will switch back to its dark conformation, not binding to the aptamer until hit by a photon again, resulting in a new conformational change to the aptamer binding light conformation. Ultimately, the cellular concentrations of these proteins are unknown and it was not measured if one is expressed better compared to the other one. Thus, it was not a further concern, that MS2 activation was always superior compared to light-dependent activation by PAL. SG9 was used for further assays, since it showed the best performance in the eBFP activating LITA.

A point mutation in the NLS, N-terminally of p65 in PHP was identified and leading to a change of amino acid composition from PKKKRKV to PKKRRKV. When PHP was sequenced, the mutation was at the beginning of the sequencing and the mutation was believed to be an artefact. These artefacts are sequencing device specific and origin from a constant separation matrix, which does not resolve all fragments in the same way. Very short fragments (at the beginning) or very long fragments (at the end) affect the readout of these systems. Since PHP needs to be located in the nucleus for transactivation, a mutated NLS can reduce the activity of PHP and lead to a mixed distribution of the protein within cells[266]. Also the binding constant of the peptide to importin-alpha is lowered from 9 nM to 360 nM in this specific case[267]. It was hypothesized, that the mutation also might be the reason for a lower performance of PHP, when compared to the SAM system. On the other hand, it was observed, that PHP was located in the nucleus of transiently transfected HeLa cells **(Figure 44)**. The NLS was corrected by inverse PCR **(Figure 53)**, again spotting by-products in the gel in the loading pocket, probably associated to residual *DpnI* enzyme and a lower band, which might originate from unspecific hybridization of the primers to similar sequences. Since the plasmid was isolated before re-circularization, formation of by-products was

not further analysed. Three clones with a corrected NLS were identified. The corrected PHP variant was analyzed in a LITA experiment **(Figure 54)**. No major changes in eBFP expression were detected in the light, thus it was concluded, that the mutation does not affect the nuclear localization in a significant manner. Further, a PHP version without NLS was tested and light-dependent activation of eBFP was observed, indicating that other components of the fusion protein also might promote the transport of the protein into the nucleus. The transactivation domains p65 and HSF1 were analyzed using NLS Mapper[195], however no NLS was detected for both of the domains, opening the question, whether PAL mediates a certain amount of nuclear localization. The original PHP version was used for further assays, which was also used in previous experiments, since it showed the best performance among all samples **(Figure 54)**.

It was interesting to know, whether both aptamers were necessary for robust transactivation of transcription. For sgRNA 2.0 the MS2 aptamer was inserted only in the tetra-loop or only in the stem-loop 2 (sgRNA 1.1 and sgRNA 1.2). Insertion in the tetra-loop only resulted in a weaker amount of transcription activation compared to a single insertion of the aptamer in stem loop 2[118]. The highest amount of activation was always found for having both aptamers inserted in the sgRNA. To investigate the according characteristics for aptamer 53, it was decided to mutate one of both aptamers and analyze the effect in a cellular assay, overexpressing eBFP. Two constructs were designed, SG9M21TL, having a M21 mutation in the tetra-loop aptamer and SG9M21SL2, having a M21 mutation in only stem loop 2. Both constructs were synthesized using a sense and antisense strand, as introducing a mutation in the GC-rich stem regions of the sgRNAs was unlikely to succeed **(Figure 55)**. Again, the hybridization of the sense and the antisense strand was not 100%, however full-length sgRNA insert was generated and successfully cloned into the pENTR.hU6 plasmid.

It was hypothesized, that dimeric PAL prefers a dual hairpin as previously discussed for ANTAR domains, thus SG9 and a single mutated version were used in a LITA experiment **(Figure 56)**. In fact, two aptamers are crucial for robust light-dependent activation of gene expression. Interestingly and in contrast to the SAM system, SG9M21SL2 shows a slightly stronger activation of eBFP expression as SG9M21TL, which is exactly the opposite finding of the results of MS2-based experiments[118]. It might be that protein specific conformational circumstances lead to these effects. MS2 coat proteins form dimers and bind to the MS2 aptamer, independent of light. For PAL

it was hypothesized, that before binding to aptamer 53 it might be possible that the dimerization process needs to occur before binding of the dimer to the aptamer. However, until now these mechanics are not yet fully revealed. On the other hand, the low activation is close to the background signal of the assay and thus our findings might be within the error of transfection and measurement.

Next, the light intensity needed for light-dependent transcription activation using PHP was analyzed **(Figure 57)**. The amount of light applied to the cells can be optimized by levelling two parameters: maximum effect of gene activation and cellular stress. For the maximization of the amount of light, the photon density of the light is crucial. When PAL is irradiated with light, the intensity of the light modulates the photon density. The photon density has a direct influence on the likeliness of an FMN incorporated in the LOV domain of PAL getting hit by a photon. The higher the photon density, the higher the likeliness for a hit. Since PAL is in an equilibrium between the dark and the light conformation, even when light is applied to PAL, after a while it will switch back to the dark conformation. The time, which is needed until statistically a photon hits an FMN again and a new photo cycle is induced modulates the influence of the light intensity towards the gene expression. On the other side, if the photon density is too high, the host cell might get damaged, since it is known for blue light to induce cellular stress[44, 268]. It was found, that using a light intensity of 100 µW/cm² leads to a maximum of transcription activation and a further increase of light intensity does not increase the transcription efficiency. The same intensity was also described for similar systems[64, 166]. However, also lower intensities for PAL in bacteria are described[166]. For single-chain photo switchable Cas9 light intensities of up to 10 mW/cm² are described[111]. Based on our findings, the light intensity was kept at 100 µW/cm².

As further optimization, the orientation of transactivation domains was tested. A variant of PHP, NLS-HSF1-p65-PAL(-T2A-mCherry), HPP, was generated from PHP to investigate the effect of the order of the transactivation domains. For the SAM system it was shown, that the choice of effectors domains modulates the amount of transactivating activity[118]. Since in the MPH construct, the order of transactivating domains from the interfacing molecule (which is MS2) is p65 and HSF1, the order is not directly comparable to the PHP construct, in which PAL is the interfacing molecule, followed by HSF1 and p65. HPP was generated in a 2-step process, in which first p65 was excised from PHP and then re-inserted between HSF1 and PAL. The excision of

p65 by inverse PCR led to multiple PCR products **(Figure 58 A)**, from which the correct product, indicated by a black arrow, was isolated and purified. The top product is probably DNA, which is stuck to protein and thus cannot travel the gel. This might be residual *Dnp*I enzyme, which was used to digest the template plasmid coding for PHP after the PCR and before the gel electrophoresis. However, it also cannot clearly be excluded that it is genomic DNA from the PHP plasmid purification. The lower PCR product at ~ 700 bp is a PCR by-product originating from unspecific binding of the primers. The primers were designed to bind after p65 (forward) and before p65 (reverse). The forward primer might include some linker sequence, which is used to separate the components of fusion proteins. These kinds of linkers also exist between other components of the whole fusion protein and sequence homology might be the cause here. The insert, p65, was successfully amplified from PHP **(Figure 58 B)**. The PCR product of HSF1-PAL (HP) was re-circularized and sequenced. A clone having a single point mutation was used for further processing. As discussed previously, point mutations can occur when GC-rich homologous ends form secondary structures during recombination. This is also believed to be the reason for the point mutation here. Then, p65 was inserted C-terminally towards HSF1 by inverse PCR **(Figure 59 A)**. The PCR reaction led to some by-products as well, since the primers were directed N-terminally and C-terminally of HSF1, where linker regions are inserted as described before. The mutation from the first inverse PCR was corrected by mutagenic PCR **(Figure 59 B)**, again leading to by-products of the PCR. However, in each case the correct sized PCR products were cut out and purified before further processing. Finally, a correct clone for HPP was identified. PHP and HPP were tested using SG9 and the eBFP reporter plasmid in a LITA. PHP was found to generate the strongest amount of activation **(Figure 60)**. Since the molecular interactions were not directly analyzed a comparison of the hypothetical binding sites of both transactivation domains towards RNAP II and the DNA was made. The transcription activation domain of p65 is present only in the carboxyl terminus, while phosphorylation of the protein promotes binding towards the DNA[269]. It forms interactions with transcriptional regulatory proteins such as TFIIB, TFIIH, CREB-binding protein (CBP)/p300 and TAFII31. The TAD of p65 consists of two subdomains. TA2 is located between residues 428 and 521 and TA1 is located between residues 521 and 551[270]. Both subunits contain the Nine-amino-acid-transactivation domain, which is the smallest domain promoting activation of transcription for a broad range of genes. Comparing the K_D's of the TAD of p65 to TBP

(340 pM) and 23 nM for TFIIB and 26 nM for E1A (measured by SPR), leads to the speculation that the high affinity of p65 TAD towards TBP is the driving force in regard to direct the transcription factor in proximity to the DNA, while the intervening DNA stretch is looped out. Weaker binding TFIIB and E1A 13S proteins are believed to contribute positively to the already ongoing transcription[271]. This hypothesis is supported by the fact, that TBP is crucial for the correct function of p65 TAD[272, 273]. The TADs for HSF1 are described in yeast. HSF1 has two activation domains, that modulate activation of transcription during heat shock response from which one is N-terminal and the other is C-terminal of the protein, whereas each TAD activates another set of heat inducible genes[274]. The difference in both domains is believed to be found within the recruitment of TBP-associated factors (TAFs), since only for the C-terminal TAD TAF-recruiting activity is reported[275]. From these findings, it can be hypothesized, that p65 and HSF1 both interact with TBP, whereas the C-terminal TAD of HSF1 can recruit TAFs. Since the C-terminal TAD of HSF1 would be the only obvious benefit, its orientation within the fusion protein might make a difference. However, for both constructs PHP and HPP, the C-terminal TAD of HSF1 is within the fusion protein and its TAF recruiting function remains elusive. Probably the crystal structure data of the fusion proteins interacting with a promoter in context of RNAP II assembly might resolve these questions. PHP was the candidate which was further used based on the data set shown in **Figure 60**, also because of further re-combination of the three most important systems in CRISPRa and their transactivating components did not lead to an improvement of transcription induction[276].

In order to confirm the spatio-temporal nature of light-inducible transcription activation a setup using a photomask was constructed. Stiched whole-well images of HeLa cells were taken using the fluorescence channel of the eBFP reporter and the activated cells were analyzed regarding their spatial activation pattern. A python script was used, which converts pixel data to integrated pixel intensities in x-direction. It was observed, that spatial control of eBFP expression occurs, when comparing the photomasks with each other **(Figure 61)**. For the photomask, applying only a small slide of light to the cells, a signal in the central area of the well was detected, which was expected **(Figure 61 A)**, whereas the positive control, without blocking the light leads to a more homogenous distribution of cells that were activated. However, it can be seen, that in **Figure 61 B**, the pattern shifts to the right. This can be explained by the fact, that the LED in the LPA device is not located in the middle under the well but slightly shifted.

For the negative control in **Figure 61 C** a spike occurs on the left side of the graph. This can be explained by dust, which is located at the position on the image **(Supporting Figure 3)**. Based on these results, spatial control of gene expression using LITA was confirmed to be possible. It was decided to evaluate the stitched images using a python script, as the quadruple transfection was used and the number of transfected cells was not high enough to observe the patterns easily by eye.

Having optimized the stem-length, effector order, plasmid ratios and light-intensities next, the system was tested to activated endogenous targets such as CXCR4. CXCR4 is a receptor, that is expressed on the surface of a variety of mammalian cells and thus can be addressed by antibody staining[191]. It was already used as a target gene in combination with the SunTag system[128]. In our approach, HeLa[191], MCF7[277] (negative control) and A549[278] (positive control) cells were used. Antibody staining was only successful for HeLa cells **(Figure 62)**, while the positive control A549 cells remained CXCR4 negative. It remains unclear, why A549 did not show fluorescence, after antibody staining it can only be hypothesized that either the staining was not sensitive enough or during the procedure of staining an error occurred. However, HeLa cells were used for further experiments. Based on previous results, it was decided to use SG9 along with SG9M21 and sgRNA 2.0 as controls for upregulation of endogenous CXCR4. The seed sequences were adapted from Zalatan et al[134], since in the publication a MS2 aptamer-based approach was used to induce CXCR4 expression. The sgRNA inserts were PCR amplified **(Figure 63 A)** and cloned into the pENTR.hU6 backbone. Point mutations were fixed by inverse PCR **(Figure 63 B)**, however cloning revealed that correcting the mutation for sgRNA CXCR4-#6 SG9 was not possible. This might be due to poor primer binding efficiency after denaturation, since the ssDNA sgRNA template strands probably tend to form secondary structures. This hypothesis is backed by the fact that the sgRNA SG9 (without seed sequence) has already a GC content of 63%. Secondary structures, that inhibit or slow down PCR typically occur in GC-rich regions[279] and ssDNA hairpins can arrest DNA synthesis during amplification[280]. A much simpler strategy was revealed to be the exchange of the seed sequence only and leaving the highly structured Cas9 handle, in which the aptamers are inserted untouched. SG9, targeting the 8xsgRNA binding site of the pGL3-Basic-8x-sgRNA-eBFP plasmid was then used as template the seed sequence was exchanged by inverse PCR **(Figure 64)**, which after the re-circularization and sequencing of the PCR product led to a correct clone. Since HeLa cells were

successfully stained, it was concluded that the staining protocol was fine and it was continued to try to upregulate CXCR4 in HeLa cells using sgRNA 2.0 as positive control **(Figure 65)**. An upregulation of CXCR4 using sgRNA 2.0 was observed. The amount of CXCR4 activation remained at background level when SG9 or SG9M21 were used, independently from the state of illumination. These findings indicate, that the position of the complex binding to the promoter might be crucial and is not yet optimized for PAL-based transcription activation. Thus, further sgRNAs were cloned. Using inverse PCR for the seed sequence replacement was described previously[281]. A total of 10 varying seed sequences were generated for the SG9 variant, whereas for sgRNA 2.0 the seed sequence #6 was used, since it showed the highest amount of CXCR4 activation[134]. All seed sequences were successfully introduced **(Figure 66)**.

A total set of 10 different sgRNAs was screened, all targeting the CXCR4 promoter of human cells to identify candidates, which induce robust light-dependent gene regulation **(Figure 67)**. None of the positions lead to an increase of CXCR4 expression, relative to the negative control in which no sgRNA was present. It was hypothesized, that the transfection efficiency might be a problem in this case and decided to repeat the experiment using HEK293T cells, which are known to have a high transfection uptake **(Figure 68)**. However, similar results were obtained and no increase of CXCR4 expression was detected. Based on the results in HeLa cells and HEK293T cells it was concluded, that the sgRNAs might not properly targeting the upstream promoter area of CXCR4 since in both cases, the activity remains at background level. A known problem is the inaccessibility of DNA, i.e., heterochromatin is known to be targeted by Cas9 much more difficult[282] and even chromatin-disrupting treatment is not effective enough to solve this issue[283]. Further, sgRNAs should be ideally designed to target a site between 50 bp and 400 bp, upstream of the promoter[f]. A comparison of the sgRNA seed sequences, which were used for the study presented in this work and which were published for human CXCR4[134] and from the database of sam.genome-engineering.org[g] was made. None of the published sgRNAs was found in the database. To further investigate if the sgRNAs should bind to the CXCR4 promoter, the genomic locus of the CXCR4 promoter was searched using the Genome Data Viewer[h]. Analyzing the assembly of GRCh38.p13 (GCF_000001405.39) at Chromosome 2 (Chr

[f] bitesizebio.com/44248/crispr-based-activation-crispra-of-genes-a-how-to-guide/, 23.01.2021
[g] sam.genome-engineering.org, 23.01.2021
[h] ncbi.nlm.nih.gov/genome/gdv/, 23.01.2021

2 NC_000002.12), the CXCR4 gene was identified and the promoter region (TSS and up to 1000 bp upstream of the promoter) was isolated and analyzed in regard of the sgRNA binding sites **(X.3)**. In our study, a total of four sgRNAs were identified to bind upstream of the CXCR4 promoter, whereas the other 6 sgRNAs do not bind within this region. The sgRNAs from the database request were also analyzed in the same regard and from a total of 12 sgRNAs displayed, only 4 of them were identified to bind in this region. These differences might be caused by different transcripts. For the analysis the transcript NM_003467 was used. Thus, it can be confirmed, that a few of the sgRNAs used in the assay should have bind to the CXCR4 promoter. Further reasons might be, that none of the positions tested was optimal for efficient transcription activation, or the positions were blocked by regulating host proteins. Thus, it was decided to test LITA for a new endogenous target, which was Achaete-scute homolog 1 (ASCL1). ASCL1 was multiple times successfully used in CRISPRa experiments[64, 118, 284-286]. ASCL1 is a 25.5 kDa protein that belongs to the basic helix-loop-helix (BHLH) family of transcription factors and is involved in neuronal differentiation and mediates autonomic nervous system formation[287]. It is weakly expressed in HeLa cells and thus provides a good dynamic range when activated while providing a low background signal in basal state. Thus, sgRNAs were used, which were already tested for ASCL1 activation in HeLa cells before using split-CPTS2.0[64]. The SAM system was used as light-independent control as before. In contrast to CXCR4, ASCL1 is not expressed on the cell surface but in the nucleoplasm and additionally in cytosol. Further, it was decided not to analyze the protein level after expression but the mRNA level of ASCL1 measured by qPCR. Although, the protein level of ASCL1 could have been measured using immune precipitation, the low basal concentration of ASCL1 in HeLa cells might exceed the sensitivity of the method. To simplify the process of detecting ASCL1 using qPCR pre-defined TaqMan primers and a processing kit were used, which enables fast preparation of RNA extracts. This strategy was previously used successfully in detecting ASCL1 mRNA[64]. In order to overexpress human ASCL1, sgRNA seed sequences were adapted from Nihongaki et al[64], using seed sequence #2 for sgRNA 2.0 as positive control. The seed sequences of sgRNA 2.0 and SG9 were modified using inverse PCR **(Figure 69)** and correct clones were obtained. To establish sgRNA 2.0 as a positive control and to investigate the effect of incubation in the light or in the dark on ASCL1 expression, HeLa cells were transfected and ASCL1 expression was analyzed using qPCR **(Figure 70, A)**. An increase of ASCL1 mRNA level in samples,

that were activated using sgRNA 2.0 was observed, whereas no differences was found for the illumination setups, indicating that sgRNA 2.0 targeting position 2 induces the expression. A difference of more than 5 Δct values indicates an upregulation of >32x, which was less compared to what was achieved in the literature[64]. Low level of activation might origin on transfection-based effects which were discussed earlier. The mixing of transfected cells and untransfected cells in a cell lysate leads to a dilution of the mRNA from activated cells, lowering the total amount of activation, when measured by qPCR. A qPCR efficiency curve for GAPDH was performed, to test, whether the settings of the qPCR are in an acceptable range **(Figure 70, B)**. Ideally, the efficiency of the qPCR is between 90% and 110%, which is the case for our study. PCR efficiencies of more than 100% can occur, due to polymerase inhibition, which often results from impurities within the templates[i]. In fact, due to the low transfection efficiency and the low basal expression of ASCL1, cells were concentrated and the maximum number of cells was used, which is suitable using the lysis reaction. Also, the machine used for the experiments introduces a variation into PCR efficiency[288]. An efficiency test for ASCL1 was not performed for two reasons. First, the basal amount of ASCL1 mRNA in HeLa cells is close to the detection minimum. Second, the kit used in this study and the primers were commercially produced and specifically designed and validated for the corresponding gene[j]. PAL-aptamer modified sgRNAs targeting different positions of the ASCL1 promoter were tested **(Figure 71)** in a LITA to test the light-dependent upregulation of human ASCL1 **(Figure 72)**. The strongest amount of light-dependent upregulation was found for promoter positions 2 (SG9#2) and 4 (SG9#4). SG9#1 might be too close to the core promoter of ASCL1, and its binding might be inhibited by general transcription factors, which are associated with the core promoter[289]. Other sgRNAs might be too far from the core promoter, although the DNA can bend to assist in enhancers acting in trans[290]. Interestingly, for sgRNA 2.0 similar results were obtained using the same sgRNAs[64]. This leads to the hypothesis, that the sgRNA design for our system can be adapted from SAM-based sgRNA design, which is available as an online platform at sam.genome-enginering.org. To verify this hypothesis further genes need to be tested along sgRNA 2.0. SG9#2 and SG9#4 were tested in a multiplexed setup and in comparison, to sgRNA 2.0 **(Figure 73)**. Multiplexing can further boost the amount of gene activation due to synergistic

[i] biosistemika.com/blog/qpcr-efficiency-over-100/, 23.01.2021
[j] thermofisher.com/content/sfs/brochures/taqman-gex-brochure.pdf, 23.01.2021

effects[118, 291]. For SG9#2, comparable amounts of activation to sgRNA 2.0 were achieved, which might be explained by the same target locus used, whereas SG9#4 has a lower activity in the light. It would be interesting to know the total diameters of MPH and PHP and to compare them with each other. For the MS2 coat protein a diameter of 26 nm was measured[292]. The protein diameters for PAL, p65 and HSF1 are not known, however they could be measured by calculating the hydrodynamic diameter from the molecular weight[293]. Another technique, to identify the exact location of the transcription activators recruited by MS2 or PAL is DNA immunoprecipitation, which would allow to determine the correct interaction site on the DNA[294]. However, this does not take into account for protein-protein interactions. For sgRNA 2.0 a total upregulation of ~500 fold was found, when two sgRNAs were multiplexed[118]. Comparing the effect of a single sgRNA with CPTS2.0, after 24 hours the ASCL1 mRNA level was increased by ~200 fold and for split-CPTS2.0 ~600 fold[64]. mRNA levels of ~64 fold for SG9#2 and ~16 fold for SG9#4 were found in the light. Differences in regard of transcription activation might be explained with the fact that in our studies lipofection was used, whereas Nihongaki *et al.* used lentiviral transduction. It was shown, that the PAL dependent approach of light dependent transactivation of transcription is at least comparable to what is described in literature.

VI.3 Generation of a single plasmid solution based on the UniSAM plasmid

Since dealing with a multi transfection, which decreases the amount of signal, the combination of sgRNA, dCas9 and PHP expression in a single plasmid is advantageous. One example for the unification of all components of the SAM system is a plasmid called UniSAM[193]. PB-UniSAM plasmid backbone was amplified together with the PHP insert **(Figure 74)** and PCR products migrating at the expected sizes were obtained. By-products were obtained for both PCRs, however their appearance was discussed previously and due to the gel extraction of the correct fragments, they did not disturb the downstream process. Plasmid backbone and PHP insert were successfully cloned in an AQUA cloning reaction. The sgRNAs SG6, SG9 and SG12, including native sgRNA as control were PCR amplified **(Figure 75)** and migrate at their expected sizes. For the AQUA cloning, the plasmid backbones had to be amplified by

PCR, which again lead to by-products as previously described **(Figure 76)**. Cloning was only successful for some of the clones, probably due to recombination events that can be caused by the AQUA cloning method. The correct plasmids were tested in a cellular assay and eGFP activation was found for sgRNA 2.0 and attenuated eGFP activation for SG6 and SG9 **(Figure 77)**. The activation was lowest, when no sgRNA was expressed. The eGFP activation for SG6 and SG12 was at the same level, independent of the incubation in the light or the dark. The activation in the dark might origin from a non-cleaving peptide expressed by the multi-cistronic mRNA. The PB-UniSAM plasmid uses a P2A and T2A self-cleaving peptide[193, 295]. However, the activity of these peptides is modulated by their position and terminal amino acid sequence[296]. It might be, that the separation between dCas9VP64 and PHP is not successful, thus attaching the transcriptional activators covalently to dCas9VP64, leading to a loss of light-dependent control. Interestingly, for the light-independent sgRNA 2.0, this effect might have never been noticed. The impact of VP64 on transcription activation alone was described to be very low[118]. Successful separation of PHP from mCherry by T2A was described and discussed before. At this stage, the approach to combine the expression of sgRNA, PAL, dCas9 and a reporter was dropped.

VI.4 Replacement of aptamer 53 by aptamer 58 (NM60)

A new aptamer candidate, dubbed NM60 (aptamer 58), was identified by Anna Maria Weber, which might be superior in terms of affinity, compared to the aptamer 53.19 **(Figure 78)**. SG9 targeting the 8x sgRNA binding site of pGL3-Basic-8xsgRNA-eBFP should be compared with the aptamer modified NM60 variant of the sgRNA. The cloning of pENTR.hU6 expressing SG9-NM60 was successful as sequencing resulted in a correct clone **(Figure 79)**. The new aptamer motif was tested in a LITA **(Figure 80)**, where it led to higher eBFP activation in the light compared to the canonical SG9. Several reasons can lead to these findings. The altered secondary structure can alter the rate of transcription of the sgRNAs and also will change their stability in a cellular context. When analyzing the stabilities of SG9 and SG9-NM60 with mfold under inclusion of the 5'-AAAGGTCGAGAAACTGCAAA-3' seed sequence, for SG9 only a single structure is calculated with a free energy of ΔG = -100.10

kcal/mol[238]. SG9-NM60 can fold into six different structures with a free energy from ΔG = -81.00 kcal/mol to -77.60 kcal/mol[238]. In this example, the stability of the sgRNA is not crucial to the activity, thus it was hypothesized, that the differences might occur directly at the aptamer moiety. The stem-sizes are matched. Thus, effects occurring from a stem extension were excluded. However, the size of the hairpin-loop differs from 9 nt for SG9 to 6 nt for SG9-NM60. These changes might influence the binding to PAL and the bigger loop size of SG9 provides an enhanced flexibility, which still contributes to efficient binding to PAL[297]. If SG9-NM60 is superior to SG9 in terms of binding efficiency needs to be confirmed using for example SPR or further cellular assays. However, these results could not yet be reproduced by colleague Tejal Patwari.

VI.5 ShuttlePAL

Finally, it would be interesting to optimize LITA by mutating PAL into a light-dependent shuttle between the cytosol and the nucleus. Since PHP is located in the nucleus and eukaryotic transcription factors vary strongly in their non-specific DNA interaction[298], PHP might also be responsible for light-independent unspecific gene activation. To overcome these effects at least in the dark, a PAL variant was engineered, that locates in the nucleus only in the light conformation and trans locates to the cytosol when kept in darkness. A similar approach was published for *Avena sativa* phototropin-1 (AsLOV2), where a NLS and NES was fused to the protein and conformational change lead to masking of the signal peptides[57]. Based on previous observations, it was hypothesized, that PAL might already have epitopes, that mediate nuclear transport. A bipartite NLS with moderate NLS activity was identified around amino acid position 169 **(Figure 81)**. At amino acid position 193 three repeating arginine residues were found, which might mediate nuclear localization of PAL. In order to have a strong NLS assembled, a lysine needs to be attached N-terminally to the three arginine residues, leading to the peptide KRRR[299]. In fact, it was demonstrated, that the Jα N terminus forms hydrogen bonds with R193 and R195[166], thus this epitope might be suitable for being masked by the Jα helix in the dark, but exposed in the light, enabling light-dependent shuttling. The mCherryPAL plasmid was used to mutate the leucine, which is N-terminal of the three arginine residues to lysine to generate the strong NLS,

generating mCherryPAL_L192K **(Figure 82)**. The PCR product at 5796 bp was cut out, cloned and sequenced. Microscopy analysis of mCherryPAL_L192K was performed as it was hypothesized, that the protein might shuttle between the nucleus and the cytosol based on the conformation. As expected, no shuttling was observed for mCherryPAL, which was used as a control **(Figure 83)**. Although mCherryPAL_L192K did exhibit switching activity, no shuttling was observed **(Figure 84)**. First of all, it might be that the time scales used for shuttling were too short. For previous studies using AsLOV2, the protein did shuttle within a couple of minutes[57]. An optimization of shuttle times might be helpful here. Further and in contrast to the AsLOV2 study, a NLS within the protein scaffold and not at the N- or C-terminus was generated, which makes these systems hard to compare. Finally, the dimeric nature of PAL might lead to a constant masking of the NLS, thus rendering it inaccessible for the proteins involved in nuclear import. However, redesigning the strategy more towards the AsLOV2 based construction might help, to generate a shuttlePAL, from which also the LITA might benefit.

VII Outlook

At the end of this work a Cas9-based optogenetic tool, that allows for spatio-temporal activation of transcription was built and tested for expression of endogenous genes in mammalian cells. Endogenous ASCL1 was successfully overexpressed in HeLa cells up to 546-fold in the light compared to basal levels in the dark. On reporter level, the system was slightly improved by updating the aptamer 53 to aptamer 58. It was shown, that the system can be fine-tuned by regulating the intensity of the light, applied to the cells and that in summary all components need less genetic coding space compared to comparable published systems to date.

VII.1 Improvements of the aptamer and protein domain

The presented results indicate, that using aptamer 58 might lead to an improvement in dynamic upregulation of target genes. These findings should be further investigated and tested for endogenous genes, to confirm the superiority of aptamer 58 compared to aptamer 53. Also, an aptamer, that only binds to the dark conformation of PAL might be interesting, since it would allow for transcriptional overexpression in the dark and spatio-temporal control of transcription attenuation in illuminated areas. This could be interesting for applications in retinal neurons, since these are very sensitive to blue light[300]. Blue light can disrupt the function of phospholipids bound to the plasma membrane[301]. Positioning and number of aptamer-repeats within the sgRNAs could be varied as well, since more aptamers could recruit more transcriptional activators, which would increase the local concentration of effectors.

On the protein level, PAL could still be engineered, to reduce background gene activation in darkness. The initial idea of the shuttle PAL was to translocate PAL between the nucleus and the cytoplasm. Since transcriptional activity is only expected in the nucleus, the absence of PHP from the nucleus in the dark would reduce unspecific interactions with promoter regions and thus reduces background activity. This mechanism was shown for an engineered variant of AsLOV2[302], where the authors fused a NLS to the Jα-helix of the protein, allowing the protein to translocate

to the nucleus when irradiated with blue light. Accordingly, a second variant which was fused with a Nuclear Export Signal (NES), was able to translocate from the nucleus into the cytoplasm when irradiated with blue light[57]. This system could also be fused to dCas9, to reduce off-target binding events of dCas9/sgRNA complexes in the genome in darkness.

The transactivation domains used in our approach are p65 and HSF1. These were successfully used by Konermann et al within the SAM system in activating protein coding and non-coding elements[118]. Another set of transactivation domains is VP64-p65-rtA (VPR)[130]. For most cell types and species, the SAM system was shown to enable a high degree of transcriptional activation, however, in certain cases VPR might be the better choice. The SAM system for example was shown to be more effective in activating the Hemoglobin Subunit Gamma 1 (*HBG1*) and Titin (*TTN*) genes in HeLa cells, while in U-2 OS and MCF7 cells, VPR was more effective[276]. Thus, a PAL variant fused to VPR might be more potent, when used for studies in U-2 OS or MCF7 cells.

VII.2 Packaging in virus particles

In order to use the presented system *in vivo*, transfection of cells is not an option. The gold standard for gene delivery to native tissues and for gene therapy are Adeno Associated Virus (AAV). AAVs are ssDNA viruses that belong to the parvovirus family and depend on a co-infection of a second virus - mainly adeno viruses - in order to replicate. They are preferred over other viruses since they do not cause cancer and have a low immunogenic profile. They have been shown to be safe and effective in clinical applications[303]. Comparable systems like the CPTS2.0 system are large in size and exceed the packaging capacity of AAVs, which is at 4.7 kb[304]. AAV gene therapy was already successfully used for example the usage of dCas9-VPR with two AAVs to overexpress cone photoreceptor-specific M-opsin (Opn1mw) in a rhodopsin-deficient mouse model for retinitis pigmentosa. The authors could show that after one year, the retinal degeneration was attenuated and retinal function was improved[305]. While not being a light-dependent system, the approach demonstrates the potential of AAV based gene therapy in combination with dCas9-based transactivation systems. Other viral vectors such as adeno virus or lentivirus offer a higher packaging capacity,

however this advantage comes with challenges[306]. Adeno viruses offer only a transient expression of the payload and pre-existing antibodies will induce a strong immune response[307]. Lentiviruses are prone to insertional mutagenesis[308].

VII.3 Further size reduction

One of the biggest difficulties to use CRISPR/Cas9 in virus particles is the size of the Cas-protein. The protein is ~4.2kb in size and nearly completely demands the whole packaging capacity offered by a single AAV. Smaller variants have been discovered such as the *Streptococcus aureus* Cas9 (saCas9) with a size of ~3.2kb[309] or CasX, a hybrid Cas-protein with similarities to Cas9 and Cas12a with a size of ~3kb[101, 310]. Recently, Cas proteins were identified in megabacteriophages, dubbed Cas-Φ, with sizes between 2.1-2.4kb[102] **(Figure 85)**. These hypercompact genome editors could be used to direct transactivators towards endogenous promoters and due to their small size have the potential to be encoded in a single AAV together with the transactivation domains. However, the gRNAs for Cas-Φ are different from the sgRNAs, which can be used in combination with Cas9. Thus, new gRNA designs using the PAL aptamers need to be designed and tested first.

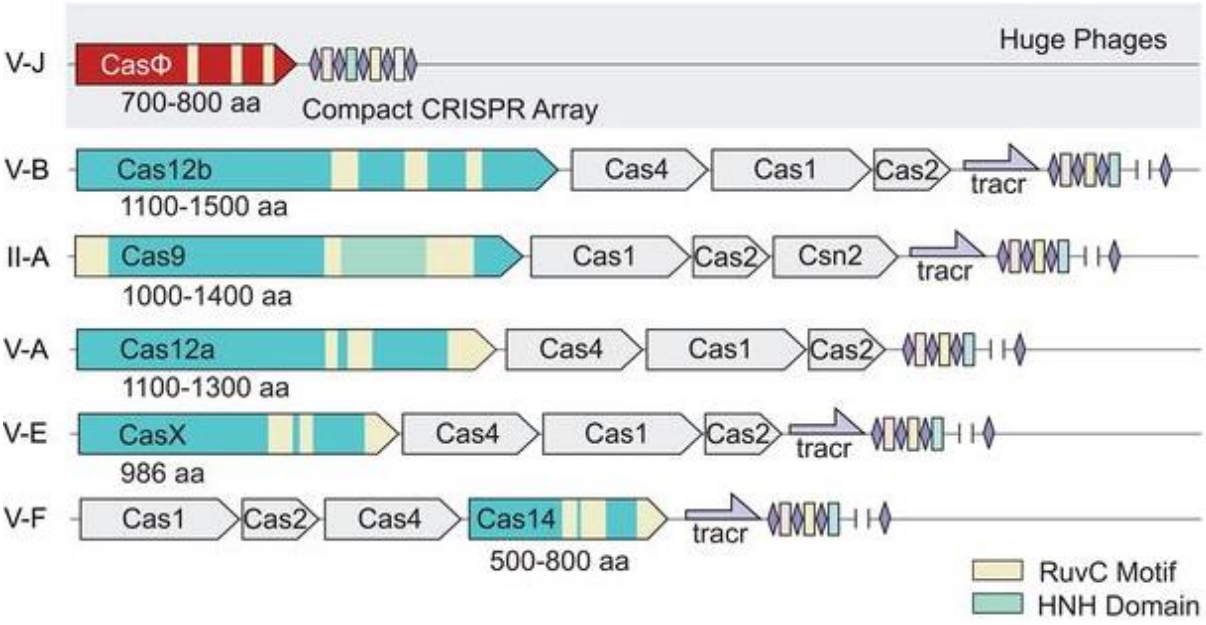


Figure 85: Size comparisons between different Cas-proteins. Image taken from [102].

Cas-Φ in combination with PHP and a small gRNA expression cassette has the potential to fit into the 4.7kb payload of a single AAV, making it the smallest light-dependent hypothetical Cas-based transcription activator to date.

Further, the usage of Cas-Φ would circumvent immunogenic reactions when used *in vivo*, which are documented for Cas9[311]. Since Cas-Φ is derived from bacteriophages, the risk of being recognized as a bacterial protein is lower compared to Cas9.

VII.4 Further applications

Besides the transactivation of genes, several other operations, that act on DNA are potentially possible **(Figure 86)**. In general, instead of using PHP, a KRAB domain would silence targeted genes, when a light stimulus is applied to the system[312]. Also, histone modifying effectors could be recruited, such as histone methyltransferases, histone demethylases, histone acetyltransferases or histone deacetylases[313]. The PAL/aptamer 53 or aptamer 58 pair could further contribute for dCas9 based imaging systems or light-dependent chromatin immunoprecipitation (ChIP).

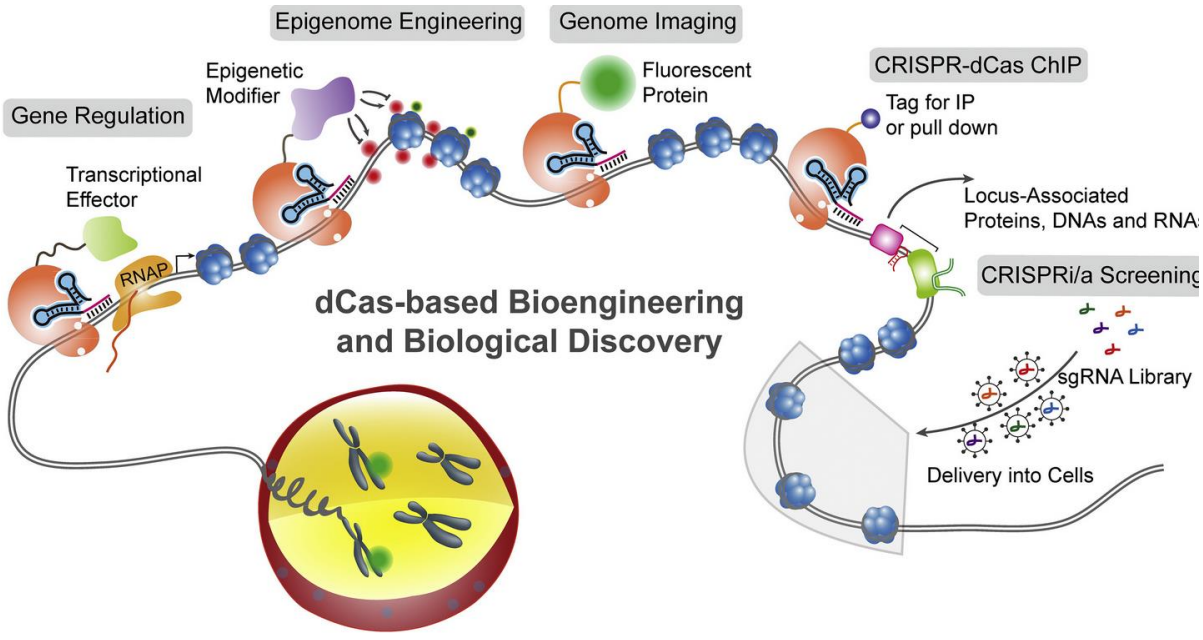


Figure 86: Overview of dCas9 applications that also could involve PAL in the future. Image taken from[313].

Finally, the system presented in this work contributes to the toolbox of synthetic gene circuits and adds a valuable light-dependent interface. It could get tested together with MS2 and PP7, both of them under the control of a conditional promoter and all three of them labelled with a different fluorophore, resulting in a variety of potentially 6 different colors, based on logic decisions.

In summary the basis of a robust light-dependent transactivation system is presented, which can be used to control transcription of endogenous genes. It offers the potential to get modified to a variety of functions such as gene silencing, epigenetic modifications and synthetic circuits. Although, the system was shown to work well in mammalian cell culture, its application *in vivo* needs to be investigated. One important step to successfully use the system *in vivo* might be the compaction of the system to make it fit into a single AAV.

VIII Methods

VIII.1 Primers and Oligonucleotides

Oligonucleotides and Primers were ordered and synthesized at Ella Biotech (Martinsried, München). The primers arrived as lyophilized pellet, which was dissolved in individual amounts of ddH_2O to gain 100 mM stock concentrations. All oligonucleotides were stored at -20 °C. The sequences of all oligonucleotides used in this thesis are presented in the appendix.

VIII.2 Plasmids

pmCherry-C1 and pEGFP-N1 were ordered from Clontech. CIB81-pmGFP was kindly provided by Dagmar Wachten (CAESAR, Bonn). pmCherryPAL was kindly provided by Anna Maria Weber. The modified version of pENTR.hU6hH1 was kindly provided by Georg Pietruschka.

The following plasmids were gained from Addgene:

Table 4: Plasmids from Addgene

Plasmid	Provided by	Publication	Addgene ID
pET-Cas9-6xHis	David Liu	*10.1038/nbt.3081*	62374
pGL3-Basic-8x-gRNA-eGFP	Charles Gersbach	*10.1038/nchembio.1753*	60718
dCAS9-VP64_GFP	Feng Zhang	*10.1038/nature14136*	61422
pU6-(BbsI)_CBh-Cas9-T2A-BFP	Ralf Kuehn	*10.1038/nbt.3198*	64323

pENTR.hU6hH1	Ming-Sound Tsao	*10.1186/1472-6750-11-24*	32686
MS2-P65-HSF1_GFP	Feng Zhang	*10.1038/nature14136*	61423
PB-UniSAM	Lesley Forrester	*10.1038/s41598-017-06468-6*	99866

VIII.3 Standard PCR

Within a polymerase chain reaction (PCR), nucleic acid fragments are amplified using a thermostable DNA polymerase and primers. For amplicons with a size < 1000 bp, homemade Pfu polymerase was used. For larger amplicons the Phusion Flash High-Fidelity PCR Master Mix was used.

VIII.3.1 Standard PCR Pfu

Table 5: Standard Pfu PCR

Component	Stock concentration	Final concentration	Volume
Pfu PCR buffer	10x	1X	2.5 µl
$MgCl_2$	25 mM	2 mM	2 µl
dNTP mix	25 mM	200 µM	0.4 µl
Forward primer	100 µM	1 µM	0.25 µl
Reverse Primer	100 µM	1 µM	0.25 µl
Template	-	C_{temp}	V_{temp}
Pfu Polymerase	2.5 U/µl	0.025 U/µl	0.25 µl
ddH_2O			ad 25 µl

VIII.3.2 Standard PCR program

Table 6: Standard PCR program

PCR step	Temperature	Time
Initial denaturation	95 °C	180 s
N cycles:		
Denaturation	95 °C	30 s
Annealing	T_{ann}	30 s
Elongation	72 °C	Θ_{elong}
Final elongation	72 °C	Θ_{elong} + 60s
Hold	10 °C	∞

VIII.3.3 Flash Phusion PCR

Table 7: Flash Phusion PCR

Component	Stock concentration	Final concentration	Volume
Template	-	C_{temp}	V_{temp}
Forward primer	10 µM	0.5 µM	3 µl
Reverse primer	10 µM	0.5 µM	3 µl
Phusion Flash High-Fidelity PCR Master Mix	2X	1X	30 µl
ddH_2O			ad 60 µl

VIII.3.4 Flash Phusion PCR program

Table 8: Flash Phusion PCR program

PCR step	Temperature	Time
Initial denaturation	98 °C	30 s
N cycles:		
Denaturation	98 °C	5 s
Annealing	67 °C	5 s
Elongation	72 °C	Θ_{elong}

Final elongation	72 °C	Θ_{elong} + 60s
Hold	10 °C	∞

VIII.4 Cultivation of microorganism

For the preparation of plasmid DNA, previously transformed e. coli bacteria were inoculated on LB-agar plates. The LB-agar plates were incubated over night at 37 °C. Liquid cultures were usually prepared as 10 mL cultures in Falcon tubes. The tubes with inoculated cultures and appropriate antibiotics were incubated in a rack with an angle of ~70° and shaking at 140 rpm.

For the expression of Cas9 first, a 10 ml preculture was made. On the next day the main culture with 200 ml was inoculated with the preculture (usually 1:100) and at an OD600=0,7 the main culture was induced with a final concentration of 1 mM IPTG. The temperature was lowered to 18 °C and the expression culture was incubated for 18h.

Mammalian cells were incubated at 37°C and 5% CO_2 in cell culture incubators and split every 2-3 days. HeLa cells and HEK293T cells were cultivated in DMEM + 10% FCS, supplied with 1% PenStrep, 1% NEAA and 1% sodium pyruvate.

In case of illuminating condition in the incubator, cells were irradiated with 100 µW/cm² using an alternating pulsing pattern with 30s on/off cycles.

VIII.5 DNA concentration measurement

Concentration and purity of DNA was determined by measuring the absorbance of 260 nm and 280 nm using the NanoDrop 2000c photometer. The ratio of 260/280 nm and 260/230 nm allow for the determination of DNA sample purity. Ratios for 260/280 nm of >1.8 and < 2.0 were considered as pure samples.

VIII.6 Extraction of plasmid DNA from *E. coli*

For the extraction of DNA from e. coli, the DNA isolation and purification kit “NucleoSpin Plasmid” (Macherey-Nagel) was used. All centrifugation steps were done at RT. The extraction was done according to the manufacturers protocol and the purified DNA was eluted in ddH_2O.

VIII.7 SDS Polyacrylamid-Gelelektrophoresis (SDS-PAGE)

SDS polyacrylamide-gelelectrophoresis enables the identification of the protein composition within a complex sample by separating the single components by size. The charge of the proteins is masked by the adsorption of negatively charged SDS.

If not stated differently SDS-PAGE gels were at 10%.

Table 9: Composition of SDS polyacrylamide gel

Separating gel (6 ml)	Volume [µl]	Stacking gel (4 ml)	Volume [µl]
ddH2O	4420	ddH2O	2440
1.5 M Tris pH 8.8	1500	1M Tris pH 8.6	1000
10% SDS	60	10% SDS	20
30% Bis-Acrylamide	20	30% Bis-Acrylamide	540
TEMED	6	TEMED	5
10% APS	60	10% APS	20

Generally, 8 µl sample was mixed with 2 µl 5x SDS Loading Buffer and afterwards denatured in a heating block for 5 minutes at 100 °C. The samples were centrifuged and the full volume of 10 µl was pipetted onto the gel. As a marker 5 µl „Thermoscientific PageRuler Prestained Marker” was used.

Electrophoresis was carried out at 160 V for 70 minutes.

VIII.8 Coomassie Staining

After SDS-PAGE, the proteins were stained with Coomassie Brilliant Blue G 250.

De-staining was done by heating the gel in water using a microwave and repeating this procedure for three to four times.

VIII.9 BioRad Protein Assay

The BioRad Protein assay was carried out following the instructions in the manual[k].

VIII.10 Western blot

To identify specific proteins on the lanes of the SDS-PAGE, the separated proteins are transferred onto nitrocellulose membranes. For blotting a semi-dry western blot was used. Two filter papers and the nitro cellulose membrane were equilibrated for 5 min in anode buffer 2. One filter paper was incubated for 5 min in anode buffer 1. Three filter papers and the polyacrylamide gel were equilibrated for 5 min in cathode buffer. After incubation the blot was composed beginning from the bottom to the top. First three filter papers which incubated in cathode buffer were placed upon the cathode. On top the polyacrylamide gel was placed. Upon the gel the nitrocellulose membrane was placed. Subsequently the 2 filter papers which incubated in anode buffer 2 were added on top of the membrane. Finally, the filter paper that incubated in anode buffer 1 was added on top. Afterwards, the anode was placed on top of the blot. Blotting was done using 2 mA/cm² for 30 min.

Blocking was done incubating the blot for 1 h in 5% BSA. The blot was washed with 1x TBST. Primary antibody was the Qiagen PentaHis antibody (mouse), secondary antibody was goat-anti-mouse Dylight 800 (Thermofisher). Before usage, the

[k] bio-rad.com/webroot/web/pdf/lsr/literature/LIT33.pdf

antibodies were diluted in DPBS. Incubation of each antibody was 1h. In between, washing steps were included 3x 5min with TBST. The evaluation of the blot was done using Odyssey Fluorescence Imaging System (LI-COR Biosciences).

VIII.11 Cas9 expression

For the expression of his-tagged Cas9, one day prior to the expression a 10 ml preculture was inoculated with *e. coli* BL21 DE3 RIL expressing pET-Cas9-6xHis. The preculture was incubated over night at 37 °C and shaking at 140 rpm. Ampicillin with a final concentration of 50 µg/ml and Chloramphenicol with a final concentration of 33 µg/ml. On the next day the main culture (200 ml LB medium as small scale or 2.5l as large-scale) was inoculated 1:100 with the preculture shaking at 120 rpm and 37 °C until an OD600 of 0,7 was reached. The expression of Cas9-6xHis was induced using a final concentration of 1 mM IPTG. During the expression for 18 hours over night, the temperature was lowered to 18 °C. On the next day the culture was split into several tubes and centrifuged for 1h at 3000 g. The supernatant was discarded and the pellets were resuspended in 8ml 1x Ni-NTA loading buffer and afterwards united. All further steps were done on ice. Sonification was at 50% power and Cycle = 30 for 5 times 30 s sonification and 30s cooling. The lysate was transferred into centrifuging tubes and centrifuged for 60 min at 20.000g. During the expression, 1ml samples were taken, centrifuged and the pellet was resuspended in 100 µl UREA buffer. The pellets were frozen at -20°C. To the supernatant, 1ml of equilibrated Ni-NTA beads (2ml bead suspension) was added (6 ml bead suspension for the large-scale production) and the incubation was for 1h at room temperature inverting. Afterwards, the lysate containing the Ni-NTA beads was loaded onto a drippling column. The beads were washed with 10cv 1x loading buffer and 10cv washing buffer (wash 1 and wash 2). The first elution was with 1ml elution buffer 1, second elution with 1 ml elution buffer 1 and third elution with 3ml elution buffer 2. For the large-scale production the elution was 2 x 3 mL elution buffer 1 and 1 x 3 ml elution buffer 2. The elution fractions were united.

During cultivation the following samples for analysis were taken:

t0 – prior to induction

t2h – after 1h of induction

t3h – after 2h of induction

pellet1 – the pellet after first centrifugation

lysate – after sonification

During the purification the following samples for analysis were taken:

load

flowthrough

wash1

wash2

elution 1

elution 2

elution 3

All samples were prepared for SDS-PAGE and western blot analysis.

The elution fractions were pooled and dialyzed over night at 4°C into 5l of Cas9 storage buffer. On the next day, concentrations were measured using the NanoDrop photo spectrometer by measuring the absorbance at 280nm or by using Bio-Rad Protein assay **(VIII.9)** according to the manufacturers protocol and using a 1mg/ml BSA standard.

VIII.12 In Vitro Transcription

The sgRNAs used in this study were transcribed in vitro using T7 RNAP. Therefore, the T7 promoter was attached to the 5'end of the sgRNA sequences. The dsDNA was used as a template within the transcription reaction. The reaction was incubated at 37 °C in a water bath for 18 hours.

Table 10: Composition of *in vitro* transcription reaction

Reagent	Stock concentration	Final concentration	Volume
Tris pH 7,9	200 mM	40 mM	20 µl
$MgCl_2$	1 M	25 mM	2.5 µl
NTPs	25 mM	2.5 mM	10 µl
DTT	100 mM	5 mM	5 µl
RNAsin	2 U/µl	0.5 U/µl	1.25 µl
dsDNA template		200 pmol	
T7 RNAP	50 U/µl	1 U/µl	2 µl
ddH_2O	ad 100 µl		

VIII.13 RNA purification

In order to purify the in vitro transcribed RNA, at first the DNA was precipitated by ethanol precipitation. A total of 10 µl of 3 M NaOAc (pH 5.4) and 300 µl 100% ethanol was added to 100 µl of transcription. The sample was shortly mixed and cooled at -80 °C for 20 min. Subsequently the precipitated RNA was centrifuged at 20.000 g for 45 min at 4 °C. The pellet was washed once with 70 % ethanol and centrifuged for further 15 min at 20.000 g at 4 °C. The precipitated RNA was then shortly dried placing the reaction tube for 10 min in a heating block at 60 °C with open lid, allowing residual ethanol to evaporate. Finally, the pellet was resuspended in 70 µl ddH_2O.

Following the precipitation, a DNAse digest was performed. To remove residual DNA from the sample, 8 µl DNAse buffer and 2 µl DNAse 1 (10 U/µl) were added. The reaction was performed for 30 min at 37 °C.

To remove salts and proteins from the reaction, the samples were run on a 10% Urea polyacrylamide gel.

Table 11: Composition of UREA polyacrylamide gel

Reagent	Stock concentration	Volume
Urea	8.3 M	35 ml
Urea in 10x TE	8.3 M	7 ml
ROTIPHORERSE Sequencing gel concentrate	25% (19:1)	28 ml
APS	10 %	560 µl
TEMED	-	28 µl

To each sample 27 µl 4 x PAA loading dye was added. The samples were heated for 3 min at 95 °C. First, the gel was pre warmed for 15 min at 375 V and 15 W per gel. As running buffer 1 x TBE was used. The gel pockets were rinsed with running buffer and the samples were loaded with the samples. The samples were run for 1.5 hours using the same power settings. The RNA was detected using a silica plate and 254 nm. The RNA was isolated from the gel using a sterile scalpel and directly used for electro elution.

For the isolation of RNA from the urea gel pieces, gel fragments were loaded into an electro elution chamber. Previously, the chamber was rinsed and filled with 1xTBE buffer. The precipitation channels were filled with 8 M ammonium acetate and the elution was running for 40 min at 150 V. After the reaction the precipitated RNA was recovered. Finally, the RNA was purified using ethanol precipitation as described above.

VIII.14 Cloning techniques

VIII.14.1 In-Fusion cloning

In-Fusion cloning can be used to fuse a DNA insert into a plasmid backbone by exploiting homolog regions at the end of the sequences. According to the In-Fusion EcoDry Cloning Kit manual[l], 150 ng linear plasmid and 70 ng insert, both sharing 20 - 30 nt homologous region at the 5' and 3' end are mixed in a total volume of 10 µl. The mixture was added to the freeze-dried reaction provided by the In-Fusion kit. The reaction is pipetted up and down a few times and then incubated for 15 min at 37 °C and subsequently for 15 min at 50 °C. The reaction can then be used for the transformation of bacteria.

VIII.14.2 AQUA cloning

AQUA cloning is a technique, where competent bacteria are transformed with linear DNA. The linear DNA, typically a plasmid backbone and an insert fragment have homologous ends (15 - 30 bp). The homologous ends are joined together by endogenous DNA repair mechanisms of the bacteria[189]. Typically, 80 ng of gel purified plasmid backbone were used and a 1:10 molar excess of gel purified insert. Plasmid backbones were generated by Inverse PCR. Plasmid backbone and insert were pre-mixed in a final volume of 5 µl in a tube and incubated for 1 h at RT. Then, the whole reaction was transformed into commercial Stellar competent *e. coli* cells. The transformation procedure was as described.

[l] takarabio.com/documents/User Manual/In/In-Fusion HD EcoDry Cloning Kit User Manual_080318.pdf, 23.01.2021

VIII.14.3 Gibson assembly

Gibson assembly can be used to fuse two or more DNA fragments together if they share homologous ends (15-30 bp)[247]. The Gibson assembly reaction was done according to the manufacturers protocol[m].

VIII.14.4 SOE-PCR

Splicing by overlapping extension (SOE) PCR is a technique, which can be used to splice DNA fragments within a PCR reaction. Phusion Flash High-Fidelity PCR was used. First the reaction was for 10 cycles and 10s at 98°C following 72 °C for 1 min without primers. In this phase the fragments, which share homologous ends are denatured and prime each other during the 72 °C step. During this step also the full-length product is spliced but not amplified by the DNA polymerase. After 10 cycles, a forward and a reverse primer were added to the reaction, which amplify the spliced full-length product. The PCR was restarted using the program below.

Table 12: SOE-PCR program

PCR step	Temperature	Time
Initial denaturation	98 °C	20 s
25 cycles:		
Denaturation	98 °C	10 s
Annealing	60 °C	10 s
Elongation	72 °C	60 s
Final elongation	72 °C	60 s
Hold	10 °C	∞

[m]neb.com/-/media/nebus/files/manuals/manuale2611.pdf?rev=9db62577a41b4cfda071e21864a6763e&hash=DEEE623067882EDC876C35BABCF9478B, 23.02.21

VIII.14.5 Inverse-PCR cloning

Inverse-PCR cloning was used to remove elements from plasmids or to introduce small tags into plasmids. When removing an element was desired, the primers were designed in a way, that the plasmid was amplified by inverse PCR but leaving out the desired region. In case of a simple excision, the purified fragment was phosphorylated using T4 PNK (NEB) and ligated using T4 ligase (NEB) according to the manufacturers protocol. Afterwards the plasmids were transformed into competent cells. In case of plasmid extension, primers were designed with overhangs, that attach nucleotides to the ends of the PCR products. If the PCR product was a full plasmid, the plasmid was extended by the primer-extended sequences. Similar to primer extension, primer mutagenesis was done. Primers were designed to have a nucleotide mutated at their 5' end. Within the inverse PCR reaction this leads to PCR products bearing the mutation. Usually, Phusion Flash High-Fidelity PCR was used to amplify the plasmid backbones. After the PCR, 1 µl of *DpnI* enzyme was added directly to the PCR reaction and the template plasmids were digested over night at 37 °C. On the next day the plasmids were run on agarose gels, cut out and purified.

VIII.15 Transformation

For transformation within AQUA cloning commercial Stellar competent *e. coli* were used. For plasmid propagation or after inverse PCR, homemade Stellar competent *e. coli* were used. Usually, 50 µl of competent e. coli were slowly thawed on ice. Then 5 µl of DNA mix was added to the bacteria and the reaction was mixed by snipping against the tube. The mixture was incubating on ice for 20 min. Afterwards, a heat shock was performed for 45 s at 42 °C in the water bath. Subsequently, the reaction was placed back on ice for another 5 min. Then, 450 µl of LB medium was added to the bacteria and the bacteria were incubated for 1 h at 37 °C and 800 rpm in a thermo block shaking for recovery. Finally, the bacteria were plated on appropriate agar plates containing antibiotics. The bacteria were grown over night at 37 °C.

VIII.16 Sanger sequencing

Isolated and purified plasmids were sent for sanger sequencing. For GATC Biotech, 20 µl of plasmid with a concentration of 50 - 200 ng/µl was prepared. For SeqLab, 12 µl of plasmid with a concentration of 50 – 200 ng/µl was prepared. Optionally, 3 µl of sequencing primer was directly added to 12 µl of sequencing reaction resulting in a final concentration of 2 µM sequencing primer. Sequences were aligned using EMBOSS Needle provided by the EMBL-EBI website.

VIII.17 Glycerol stocks

Glycerol stocks were made by mixing 800 µl of bacteria suspension from an over-night grown culture with 800 µl sterile filtered 50% (v/v) Glycerol in ddH_2O. Glycerol stocks were stored at -80 °C for long-term storage.

VIII.18 Cas9 cleavage Assay

In order to check the activity of homemade Cas9, the activity in homemade Cas9 buffer or ICR or to check the activity of modified sgRNAs, Cas9 cleavage assays were done. Usually, 150 ng of substrate was used within a cleavage assay so the cleavage products are visible on an agarose gel. The total reaction volume was 20 µl. The ratio between Cas9 enzyme, substrate and sgRNA was varying. If not stated different, the ratio between Cas9 and substrate was 8:1 and the ratio between Cas9 and sgRNA was 1:1. NEB Cas9 enzyme stock concentration was 20 µM and diluted 1:20 before usage. As substrate amplicons from the pmCherryC1 plasmid or amplicons from the pGL-8xsgRNA-eBFPplasmid were used.

Table 13: Composition of Cas9 cleavage assay

Reagent	Stock conc.	Final conc.	Final moles	Volume
Cas9 enzyme	1000 nM	66,84 nM	445.63 fmol	1.34 µl
sgRNA	250 nM	66,84 nM	445.63 fmol	5.35 µl
Cas9 reaction buffer	10 x	1 x		2.0 µl
Substrate	55,7 nM	8,36 nM	55.7 fmol	3.0 µl
ddH_2O			-	8.32 µl

First ddH_2O, Cas9 reaction buffer and sgRNA were added into a PCR reaction tube. Then, a sgRNA refolding program was started heating the mixture for 2 min at 95 °C and then cooling down with 6 °C / min until the sample temperature is at 20 °C. Then Cas9 enzyme is added and a 5 min incubation step allows for the complex formation of sgRNA and Cas9 enzyme. The substrate is added and the reaction is incubated for 35 min at 37 °C in a PCR cycler. When the reaction was finished, 1 µl Proteinase K was added and incubated for another 35 min at 55 °C to remove sgRNA:Cas9 complex from binding to the substrate. For analysis 4 µl 6 x DNA loading dye was added and the samples were run for 30 min on a 1% agarose gel at 130 V. The visualization was done using ethidium bromide and UV light. Variations of this protocol are highlighted in the results section.

For the cleavage quantification, ImageJ software was used. The images were first converted to 32 bit, then the minimum displayed value in the B&C menu was set to 200. The rectangle tool was used to specify the area of interest by clicking on Edit, Selection, Specify and then choosing a fixed rectangle size with a height of 50 pixels and a width of 60 pixels. Cleaved and non-cleaved fractions were analyzed. The software vertically integrates the pixel intensities which are then depicted as spikes. The baseline of the spikes was manually adjusted for each spike. Cleaved fractions were normalized against the total signal of cleaved and non-cleaved fraction.

VIII.19 Assays in mammalian cells

VIII.19.1 Transfection into mammalian cells

Transfection is a method to introduce foreign DNA into mammalian cells, which are cultivated in cell culture. One day before transfection, cells were plated in multi-well plates with a specific density. The transfection reaction was prepared by first mixing plasmid DNA with OptiMEM and Lipofectamine 2000 transfection reagent with OptiMEM in a separate reaction tube. Usually, 1.5 µl of Lipofectamine 2000 transfection reagent was used per well. For the transfection reagent a master mix was prepared. The reaction tube containing the plasmids and the reaction tube containing the transfection reagent were both shortly vortexed, centrifuged and incubated for 5 min at RT. Subsequently, plasmid mixes were added to transfection reagent master mixes, shortly vortexed, centrifuged and incubated for 30 min at RT. Usually, 50 µl of total transfection volume (DNA mix plus transfection reagent mix) were used per well for a 24-well plate. Right before transfection, the medium in the wells was exchanged to OptiMEM. Then, the transfection mix was added dropwise on top of the surface of the medium. Usually, cells were transfected for 4 hours. After transfection, the transfection reaction was removed by exchanging the medium to cultivation medium.

Polymer-based transfection was done using Xfect Transfection Reagent (TaKaRa)[n]. Xfect Transfection reagent was used to transfect the UniSAM plasmids. For UniSAM experiments, 1500 ng of UniSAM plasmid was used and 500 ng of pGL3-Basic-8xsgRNA_eGFP reporter plasmid. First reporter plasmid was diluted in 50 µl Xfect Reaction Buffer, then UniSAM plasmid was added. 1.5 µl Xfect polymer was added to each tube and the mix was shortly vortexed and incubated for 5 min at RT. Then, 50 µl transfection was added dropwise to each well. Transfection was incubated for 4 hours, then the medium was exchanged to cultivation medium.

[n] takarabio.com/documents/User Manual/Xfect Transfection Reagent Protocol/Xfect Transfection Reagent Protocol-At-A-Glance_103012.pdf, 23.01.2021

VIII.19.2 LSM analysis

For laser scanning microscopy (Zeiss LSM 710) analysis the cells were plated on 24-well glass-bottom µ-Plates, black walls and flat coverslip bottom (ibidi GmbH, Martinsried, Germany). After transfection, cells were kept in the dark until analyzed by the microscope. The fluorescence signal of mCherry was recorded at ex/em: 543/596-696 nm. The fluorescence signal of PAL in the dark conformation was recorded ex/em: 405/488 529 nm. Fluorescence of eBFP was recorded at ex/em: 383/448For switching analysis, first an image in the dark was recorded. Then, the cells were irradiated for 1 min with light of 465 nm to switch PAL into its light and another image was recorded. After 10 min of dark recovery another image was recorded to show the restored fluorescence intensity.

Photomask experiments were recorded as stitched images.

VIII.19.3 LPA experiments

For the light-intensity screening and spatio-temporal irradiation experiments, the Light Plate Apparatus (LPA, Tabor Lab) was used[314]. The LPA is a customizable 24-well plate, which allows for individual irradiation of the wells used in the study. The irradiation pattern can be programmed according to light-source and the modulation of the light intensity. For the experiments 4titude Vision Plates 24-well, glass-bottom and black walls were used.

For light-intensity screening experiments HeLa cells were seeded with 70.000 N/well. On the next day the cells were transfected. From the beginning of the transfection the cells were incubated in the LPA device, which was placed in the cell culture incubator to keep 37 °C, 5% CO_2 and humidity. Within the LPA, the samples were irradiated with various amounts of light, whereas the modulation of the light was kept pulsing at 30s on and 30s off. After 24 hours of incubation, the cells were

For spatio-temporal irradiation experiments, cells were seeded and transfected as described.

Three photomasks were designed for the experiment. The negative control shielded the cells from the light of the LPA. The positive control mask was open to irradiate all cells. The slide mask was constructed with a 5 mm wide and 2 cm high slide to irradiate only a small area of the cells. The incubation within the LPA was for 48 hours. Analysis was done using the LSM.

VIII.19.4 Spatio-temporal irradiation experiments evaluation

The images recorded from the LSM were processed to visualize the spatio-temporal activation. In the first step the images were resized to 1792 x 1792 pixels. Further, the images were converted to 8-bit grayscale. The white tonal correction was adjusted from 255 to 100 and gamma was corrected from 1,0 to 1,3. A python program was written to integrate the pixel values vertically for each pixel column within the image **(XI.6)**. The integrated and normalized grey values of the pixels were then plotted against the pixel position in x-direction.

VIII.19.5 Flow cytometry measurement of eBFP

Before flow cytometry measurement, the cells were detached from the wells. First, the medium was removed and the cells were washed once with 500 µl warm DPBS. The DPBS was discarded and 250 µl Accutase enzyme was added per well. The incubation for the detachment reaction was for 10 min at RT. The progress of detachment was followed by microscopy analysis. After detachment, 250 µl cultivation medium was added to the cells and to quench the reaction of the Accutase enzyme. The cells were transferred into FACS tubes and washed once with 1 ml DPBS. The cells were remaining in ~ 100 µl DPBS for flow cytometry measurement.

The BD FACS Canto II flow cytometer was set to FSC=50V, SSC=300V, Pacific-Blue=187V and FSC-Threshold to 5.000. The data was recorded in log-scale. From each sample 30.000 events were recorded.

VIII.19.6 Flow cytometry measurement of CXCR4

After incubation of the cells for CXCR4 activation experiments, the supernatant of the cells was removed and the cells were incubated for 10 min in 500 EDTA Buffer. Afterwards, the cells were transferred into FACS tubes and softly singularized by pipetting up and down using a blue pipette tip. The cells were washed 1x with 1 ml Staining Buffer. The supernatant was removed and 100 µl Blocking Buffer was added to the cells. The cells were shortly resuspended by vortexing for 2 seconds. Then, the cells were incubated for 15 min the cell culture incubator. For the antibody staining, 5 µl of 1:10 diluted human Alexafluor 405 conjugated antibodies were added to each tube. The antibody was diluted in Blocking Buffer. The cells were shortly resuspended by vortexing for 2 seconds and incubated for 30 min in the dark at RT. Then, 1 ml of Staining Buffer was added to the cells and the cells were washed 3 times using staining buffer before FACS analysis. The cells were remaining in ~ 100 µl DPBS for flow cytometry measurement.

The BD FACS Canto II flow cytometer was set to FSC=50V, SSC=300V, Pacific-Blue=187V and FSC-Threshold to 5.000. The data was recorded in log-scale. From each sample 30.000 events were recorded.

VIII.19.7 Flow cytometry data evaluation

Data retrieved from the flow cytometer was further processed before graphical display. First, all measurements were gated for single cells, to avoid high signals, which can be caused by duplets. This was achieved by applying the first gate within the SSC-A vs SSC-H plot. In the next step, the autofluorescence of the cells was subtracted from the Pacific-Blue-A signal. The autofluorescence of the cells was defined at 1% shift. Cells, that exceed 1% were considered as fluorescent cells. Usually, this step was set within a negative sample (untransfected or a dark sample where no activation was expected). First, the plot was set to SSC-A vs Pacific-Blue-A then, the quadrant gate was set to 1%. This step was applied to all samples. Finally, a data table was generated for all values from quadrant #2 **(Figure 96 A)**.

The generated data now represents crude shift values. These values were normalized using min-max-scaling. Regarding the maximum value, the data usually was normalized to sgRNA 2.0 and the SAM system. In experiments without gRNA 2.0 and SAM system, the data was normalized to the highest sample value if not stated differently. Regarding the lowest value, untransfected cells or a dark sample with the lowest signal was chosen. For each sample the feature-scaling method was applied **(Figure 96 B)**.

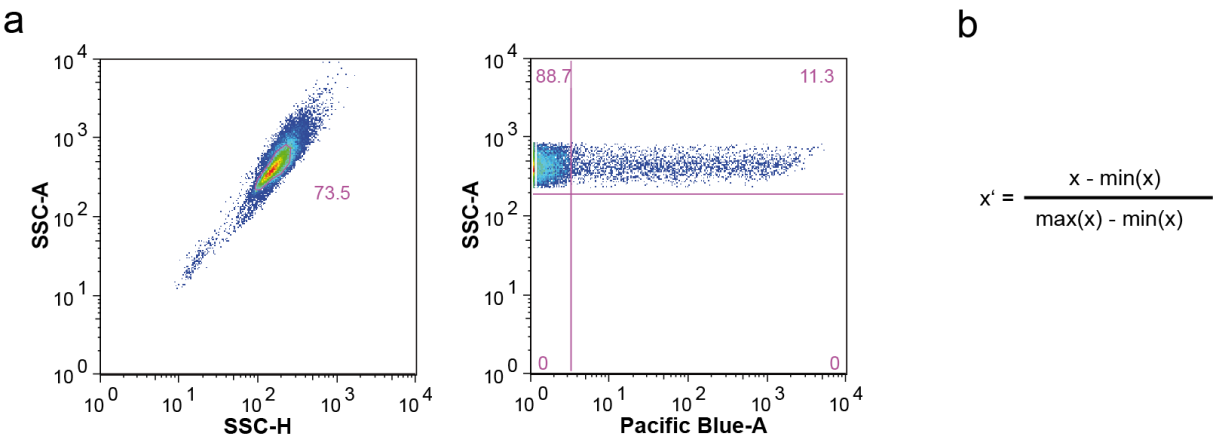


Figure 87: Gating strategy for the eBFP activation experiments.

VIII.19.8 qPCR sample preparation

For qPCR analysis of ASCL1 activated cells the cells-to-ct kit (ThermoFisher) was used[o]. After incubation of the cells in the cell culture incubator, the supernatant was removed. Then, the cells were washed with 500 µl warm DPBS. The cells were detached using 300 µl ice cold DPBS. The detaching was supported by scraping the cells from the surface of the cell culture wells. The cell suspension was transferred into 1,5 ml reaction tubes and centrifuged for 60 seconds at 1000g. 270 µl supernatant was removed. Then 10 µl of cell suspension was transferred into PCR stripes for further downstream processing. To each sample 49,5 µl of Lysis solution and 0,5 µl DNAse I

[o] assets.thermofisher.com/TFS-Assets/LSG/manuals/cms_056225.pdf, 23.02.2021

was added. A lysis master mix was prepared containing both reagents and 50 µl of the master mix was added. The reaction was mixed by pipetting up and down 5 times and incubated for 10 min at RT. Then 5 µl Stop Solution was added per sample and the reaction was mixed again by pipetting up and down 5 times. Incubation was for 2 min. For the reverse transcription, PCR stripes were prepared. 25 µl RT Buffer were mixed with 2,5 µl RT enzyme, 12,5 µl ddH_2O and 10 µl of RNA from each lysis reaction. Two minus-RT samples were considered using 2,5 µl of ddH_2O instead of RT enzyme.

Table 14: qPCR RT reaction

RT step	Temperature	Time
Reverse transcription	37 °C	30 min
RT inactivation	95 °C	5 min
Hold	4 °C	∞

For the qPCR reaction a BioRad Rad Hard-Shell® 96-Well PCR Plate was prepared. A master mix was prepared using 5 µl of 2xMasterMix, 0.5 µl of forward primer, 0.5 µl of reverse primer and 3 µl of ddH20. For each sample 1 µl of template from the RT reaction was used. Two NTCs were included.

Table 15: qPCR program

qPCR step	Temperature	Time
UNG incubation	50 °C	2 min
Enzyme activation	95 °C	20 s
40 cycles:		
Denature	95 °C	10 s
Anneal/Extend	60 °C	30 s

VIII.19.9 qPCR analysis

The data gained from the qPCR reaction was further processed. First the dct were calculated by subtracting the GAPDH values from the ASCL1 values and to normalize all samples to the housekeeping gene. In the next step, the samples were normalized to native sgRNA to remove the background signal and to make the independent assays comparable to each other. Resulting ddct values were exponentiated by calculating 2^ddct to gain the fold-change of ASCL1 mRNA.

VIII.19.10 mCherry Pulldown assay

The mCherry pulldown assay is used to test the binding of cellular expressed and mCherry-tagged PAL protein to immobilized aptamers. HeLa cells were seeded in double at 200.000 N/well in a 6-well plate and transfected with 2 µg of pmCherryPAL, pNLS-mCherryPAL or pmCherry plasmid. On the next day, the cells were harvested in 15 ml Falcon tubes and washed once with 10 ml PBS. The cells were centrifuged for 5 min at 300 x g. For the assay, the 1x ICB was modified (pulldown ICB). 8ml 1x pulldown ICB was mixed by using 7912 µl 1x ICB and adding 8 µl RNAsin (40u/µl) and 80 µl 100mM PMSF (final 1 mM). The supernatant was removed and the cell pellets were resuspended in 3 ml of 1x pulldown ICB. Then the cells were lysed on ice by sonication (1 min, cycle=50%, power=50). The lysate was then transferred into 2 ml reaction tubes and centrifuged for 1 h at 4°C and 20.000g.

The capture plates were prepared accordingly. Two plates with streptavidin coated wells were washed 3x with 190 µl 1x pulldown ICB and coated with 125 pmol of different aptamers and control sequences for 2 hours at RT. Afterwards, the wells were washed 3x with 190 µl 1x pulldown ICB and incubated with 100 µl cell lysate for 20 min in the presence and absence of light. Afterwards both plates were washed 3x with 190 µl 1x pulldown ICB and 100 µl 1x pulldown ICB was added. The mCherry-fluorescence (Ex/Em: 587/610 nm) was measure twice for both plates.

IX Material

IX.1Reagents and solutions

Table 16: Reagents and solutions

Chemical/Reagent	Manufacturer	Order No.
Agar Bacteriology grade	AppliChem, Darmstadt, Germany	A0949
Agarose LE	Genaxxon, Ulm, Germany	M3044
Ampicillin sodium salt	Carl Roth, Karlsruhe, Germany	HP62.1
Chloramphenicol		
Cutsmart Buffer	NEB, Ipswich, USA	B7204
Dimethyl sulfoxide	Merck, Darmstadt, Germany	472301
di-Sodium Hydrogen Phosphate Dihydrate ($Na_2HPO_4 \cdot 2\ H_2O$)	Carl Roth, Karlsruhe, Germany	4984
dNTP Bundle 100 mM (dATP, dCTP, dGTP, dTTP)	Jena Bioscience, Jena, Germany	NU-1005
Ethanol absolute	Merck, Darmstadt, Germany	32205
Ethidium bromide solution (1.00%)	Carl Roth, Karlsruhe, Germany	2218
FACS Clean Solution	Becton Dickinson, Franklin Lakes, USA	340345
FACS Flow Solution	Becton Dickinson, Franklin Lakes, USA	342003
FACS Shutdown Solution	Becton Dickinson, Franklin Lakes, USA	334224
GeneRuler 100 bp DNA Ladder	Thermo Fisher Scientific, Waltham, USA	SM0241
GeneRuler 1 kb DNA Ladder	Thermo Fisher Scientific, Waltham, USA	SM0313
Glyerol (86%)	Carl Roth, Karlsruhe, Germany	4043
Kanamycin sulfate	Merck, Darmstadt, Germany	K1637
LB Broth (Lennox)	Carl Roth, Karlsruhe, Germany	X964
Magnesium Chloride ($MgCl_2$) (25 mM)	Promega, Madison, USA	A351H
NEBuffer 3.1	NEB, Ipswich, USA	B7203
Potassium Dihydrogen Phosphate (KH_2PO_4)	Carl Roth, Karlsruhe, Germany	3904
Protino Ni-NTA Agarose	Macherey-Nagel, Düren, Germany	745400.100
Sodium Chloride (NaCl)	Thermo Fisher Scientific, Waltham, USA	J21618

IX.2Kits

Table 17: Kits

Kit	Manufacturer	Order No.
Bio-Rad Protein Assay	Bio-Rad, Düsseldorf, Germany	5000006
NucleoSpin Gel and PCR Clean-up	Macherey-Nagel, Düren, Germany	740609
NucleoSpin Plasmid	Macherey-Nagel, Düren, Germany	740588
CellTiter-Glo Luminescent Cell Viability Assay	Promega, Madison, USA	G7570
TaqMan Fast Advanced Cells-to-CT Kit	Thermo Fisher Scientific, Waltham, USA	A35374

IX.3Enzymes

Table 18: Enzymes

Enzyme	Manufacturer	Order No.
Antarctic Phosphatase	NEB, Ipswich, England	M0289S
*Afl*III	NEB, Ipswich, England	R0541S
*Avr*II	NEB, Ipswich, England	R0174L
*Bam*HI	NEB, Ipswich, England	R0136L
*Bgl*II	NEB, Ipswich, England	R0144S
Cas9	NEB, Ipswich, England	
*Dpn*I	NEB, Ipswich, England	R0176S
DNAse I	Roche, Basel, Schweiz	4716728001
*Mfe*I-HF	NEB, Ipswich, England	R3589S
*Nco*I	NEB, Ipswich, England	R0193S
*Nco*I-HF	NEB, Ipswich, England	R3193S
*Nhe*I-HF	NEB, Ipswich, England	R3131S
*Not*I	NEB, Ipswich, England	R0189S
Pfu Polymerase (2.5 U/µl)	Homemade	-
Phusion Flash High-Fidelity PCR Master Mix	Thermo Fisher Scientific, Waltham, USA	F548S
Proteinase K	Roth, Karlsruhe	7528.1
*Pst*I	NEB, Ipswich, England	R0140L
*Sal*I	NEB, Ipswich, England	R0138S
RNAsin	Promega, Mannheim	N2511
T4 DNA Ligase	NEB, Ipswich, England	M0202S
T4 PNK	NEB, Ipswich, England	M0201L

T7 RNA Polymerase	Homemade	-
*Xho*I	NEB, Ipswich, England	R0146S

IX.4Buffers

Table 19: Buffers

Buffer/solution	Composition	Storage
Ammonium Acetate (8M)	$NH_4CH_3CO_2$	RT
Cas9 Cleavage Buffer (5x)	500 mM KCl, 100 mM HEPES, 10 mM $MgCl_2$, pH 7,2 at RT	RT
DNA Loading Dye (6x)	0,03% (w/v) Bromphenol lue, 60 mM EDTA, 60% (w/v), Glycerol 10 mM Tris-HCl	RT
Intracellular Buffer ICB (4x)	10 % Glycerol, 135 mM KCl, 10 mM NaCl, 12 mM HEPES, 1 mM MgCl2, pH 7.0 at 37 ^oC	RT
LB medium	20 g LB broth, 1 l ddH_2O	RT
PAA Loading Dye (4x)	9 M Urea, 50 mM EDTA	
Pfu PCR buffer (10X)	200 mM Tris pH 8.8, 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 1% Triton X-100, 1 mg/ml BSA	-20 °C
Sodium Acetate (3M)	NaOAc	
TBE buffer (10X)	90 mM Tris pH 8.0, 90 mM Borate, 2 mM EDTA	RT
TBE buffer (1X)	1:10 dilution of TBE buffer (10X)	RT

IX.4.1 Cas9 expression buffers

Table 20: Cas9 experssion buffers

Buffer/solution	Composition	Storage
Cas9 storage buffer	100 mM KCl, 20 mM HEPES	RT
EDTA	100 mM EDTA	RT
IPTG	1M IPTG	-20°C
Laemmli Buffer (6x)	9% (v/v) β-mercaptoethanol, 48% (w/v) Glycerol, 6%	-20°C
Ni-NTA 6x Loading Buffer	300 mM NaH2PO4, 350 mM NaCl, 20 mM Imidazole, pH 8.0 at RT	RT
Ni-NTA Washing Buffer	50 mM NaH2PO4, 350 mM NaCl, 20 mM Imidazole, pH 8.0 at RT	RT
Ni-NTA Elution Buffer 1	50 mM NaH2PO4, 140 mM NaCl, 250 mM Imidazole, pH 8.0 at RT	RT
Ni-NTA Elution Buffer 2	50 mM NaH2PO4, 140 mM NaCl, 500 mM Imidazole, pH 8.0 at RT	RT

PBST	PBS, Tween20 0,05 % (v/v)	RT
SDS Running Buffer (5x)	960 mM Glycine, 0,5 % (w/v) SDS, 125 mM Tris-HCL	RT
UREA-Buffer	8M Urea, 200 mM Na2HPO4, 10 mM Tris, 10 mM Imidazol, pH 8.0 at RT	RT
Westernblot Anode Buffer 1	300 mM Tris-Cl, pH 10,4 at RT	RT
Westernblot Anode Buffer 2	25 mM Tris-Cl, pH 10,4 at RT	RT
Westernblot Cathode Buffer	25 mM Tris-Cl, 40 mM Glycine, pH 9,4 at RT	RT
Westernblot blocking solution	PBST, 1% BSA	4°C

IX.4.2 CXCR4 experiments buffers

Table 21: Buffers for experiments with CXCR4

Buffer/solution	Composition	Storage
EDTA Buffer	DPBS, 2 mM EDTA	RT
FACS Buffer	0,1% BSA, DPBS, 0,05 %(w/v) NaN_3	RT
Blocking Buffer	DPBS, 10% FCS (v/v)	RT

IX.5Antibodies

Table 22: Antibodies

Antibody name	Manufacturer	Order No.
Human CXCR4 alexa fluor 405 conjugated antibody 12g5r	RND Systems	Lot: 1505128 Fab170V-100ug

IX.6Equipment

Table 23: Equipment

Equipment	Manufacturer	Model
Autoclave	Systec, Linden, Germany	D-65
Balance	Mettler-Toledo, Columbus, USA	JB2002-G/FACT
	Sartorius, Göttingen, Germany	BP211D
Biological safety cabinets	Steril S.p.A, Bloomberg, USA	Antares 48
Centrifuges	Eppendorf, Hamburg, Germany	5417 C
	Eppendorf, Hamburg, Germany	5417 R
	Eppendorf, Hamburg, Germany	5424
	Eppendorf, Hamburg, Germany	5424 R
	Eppendorf, Hamburg, Germany	5810 R
	Eppendorf, Hamburg, Germany	MiniSpin
	VWR, Darmstadt, Germany	MiniStar silverline
Cell culture incubator	Heraeus, Hanau, Germany	Heracell 150
Contamination monitor	Berthold Technologies, Bad Wildbad, Germany	LB 147
Electrophoresis power supply	Topac, Cohasset, USA	Consort EV265
Flow cytometry system	Becton Dickinson, Franklin Lakes, USA	BD FACSCanto II
Freezers	New Brunswick Scientific, New Jersey, USA	-85^{o}C Ultra Low Freezer C99085
	Liebherr, Bulle, Switzerland	Liebherr Premium NoFrost -20^{o}C Freezer
Fridge	AEG, Frankfurt am Main	-4 ^{o}C AEG Electrolux
Gel documentation system	VWR, Darmstadt, Germany	GenoPlex
Incubator shaker	Thermo Fisher Scientific, Waltham, USA	Innova 4430
Laminar flow cabinet	Heraeus, Hanau, Germany	Herasafe HS 18
LED-Device 465nm	University Bonn, Bonn, Germany	Homemade
LED-Device 465nm LEDs	Kingbright Electronic Europe, Issum	Blue InGaN
Microwave	Bosch, Grelingen, Germany	HMT842K
Microscope	Carl Zeiss, Oberkochen, Germany	Axiovert25
	Carl Zeiss, Oberkochen, Germany	Observer D.1
	Carl Zeiss, Oberkochen, Germany	LSM710
Multimode plate Reader	PerkinElmer, Waltham, USA	EnSpire
Nanodrop spectrophotometer	Thermo Fisher Scientific, Waltham, USA	2000c
Orbital shaker	GFL, Burgwedel, Germany	3017
pH-meter	Mettler-Toledo, Columbus, USA	S20 – SevenEasy

Pipettes (2 µl, 10 µl, 100 µl 200 µl, 1000 µl)	Eppendorf, Hamburg, Germany	Research
	Eppendorf, Hamburg, Germany	Research plus
Pipette controllers	Brand, Wertheim, Germany	accu-jet
	Brand, Wertheim, Germany	accu-jet pro
	Starlab, Hamburg, Germany	StarPet Pro
Platereader	PerkinElmer, Waltham, USA	EnSpire Multimode
Thermal cycler	Eppendorf, Hamburg, Germany	Mastercycler personal
	Thermo Fisher Scientific, Waltham, USA	Veriti, 96 Well
Thermomixer	Eppendorf, Hamburg, Germany	comfort
	Eppendorf, Hamburg, Germany	C
Vortex mixer	Scientific industries, Bohemia, USA	Vortex-Genie 2
	VWR, Darmstadt, Germany	Analogue heavy duty
Water bath	GFL, Burgwedel, Germany	1004
	Grant Instruments, Shepreth, UK	SUB Aqua 26 Plus

IX.7Consumables

Table 24: Consumables

Consumable	**Manufacturer**
Cell culture flask (75 cm^2)	Sarstedt, Nümbrecht, Germany
Cryo tubes (2 ml)	Greiner Bio-One, Kremsmünster, Austria
Filter papers for Western Blot MN 440 B	Macherey-Nagel, Düren, Germany
Glass pasteur pipettes	Brand, Wertheim, Germany
Microcentrifuge tube (1.5 ml, 2ml)	Sarstedt, Nümbrecht, Germany
µ-Plate (24 well)	Ibidi, Planegg, Germany
Nitro cellulose membrane BA 85	GE, Solingen, Germany
PCR reaction tubes (0.2 ml)	Sarstedt, Nümbrecht, Germany
Pipette tips (200 µl)	Th. Geyer, Lohmar, Germany
Pipette tips (10 µl, 1000 µl)	Sarstedt, Nümbrecht, Germany
Plastic syringe (50 ml)	Henke-Sass Wolf, Tuttlingen, Germany
Polysterene petri dishes (100x15 mm)	Corning, Corning, USA
Polysterene round-bottom tubes (5 mL)	Corning, Corning, USA
Serological pipettes (5 mL, 10 mL, 25 mL, 50 mL)	Sarstedt, Nümbrecht, Germany
Sterile syringe Filter (0.2 µm)	VWR, Darmstadt, Germany
TC-plate (12 well, 24 well, 48 well)	Sarstedt, Nümbrecht, Germany
Tube with canonical bottom (15 mL, 50 mL)	Sarstedt, Nümbrecht, Germany
White polystyrene assay plate (96 well)	Corning, Corning, USA

IX.8Solutions

Table 25: Solutions

Solution	Manufacturer	Order No.
0.05% Trypsin-EDTA (1X)	Thermo Fisher Scientific, Waltham, USA	25300054
Accutase solution	Merck, Darmstadt, Germany	A6964
DMEM (1X) + GlutaMAX-I	Thermo Fisher Scientific, Waltham, USA	61965026
DPBS (1X)	Thermo Fisher Scientific, Waltham, USA	14190144
Fetal calf serum (FCS)	Sigma-Aldrich, Taufkirchen, Germany	F9665 (Charge:054M3399)
FuGENE HD Transfection Reagent	Promega, Fitchburg, USA	E2311
Lipofectamine 2000	Thermo Fisher Scientific, Waltham, USA	11668019
MEM NEAA (100X)	Thermo Fisher Scientific, Waltham, USA	11140035
Opti-MEM I (1X)	Thermo Fisher Scientific, Waltham, USA	31985062
Penicillin/Streptomycin (PenStrep) (100X)	PAA, Cambridge, UK	P11-010
Poly-L-lysine solution (0.01% in H_2O)	Merck, Darmstadt, Germany	A-005-C
Sodium Pyruvate 100 mM (100X)	Thermo Fisher Scientific, Waltham, USA	11360070
TrypLE Express	Thermo Fisher Scientific, Waltham, USA	12604013
Xfect Transfection Reagent	Takara Bio, Kusatsu, Japan	631317

IX.9 Organisms

Table 26: Organisms

Name	Manufacturer	Order No.
HST08 stellar cells (*E. coli*)	Takara Bio, Kusatsu, Japan	636763
E. coli BL21 DE3 RIL	Agilent Santa Clara, USA, Provided by CAESAR institute in Bonn	-
Name	**Description**	**Source**
HeLa	Human cervical cancer cell line	CLS Cell Lines
HEK293	Human embryonic kidney cell line	CLS Cell Lines
HEK293T	Human embryonic kidney cell line	Gift from H.

IX.10 Plasmids

Table 27: Plasmids

Plasmid	Antibiotic	Size (bp)
pAV-8xgRNA-U6	Ampicillin	6273
pAV-8xgRNA-U6+27-Tornado-F30-Broccoli	Ampicillin	6597
pAV-U6+27-Tornado-F30-Broccoli-empty	Ampicillin	6419
pdCas9	Ampicillin	13611
pdCas9-EGFP	Ampicillin	14328
pEGFP-N1	Ampicillin	4733
pENTR.hu6 empty	Kanamycin	2854
pENTR.hU6_gRNA_CMV_M2S9	Kanamycin	2717
pENTR.hU6_gRNA_CMV_M2S9M14	Kanamycin	2717
pGL3-Basic-8xgRNA-EBFP	Ampicillin	4188
pGL3-Basic-8xgRNA-EGFP	Ampicillin	4188
pGL3-Basic-8xgRNA-U6	Ampicillin	3375
pmCherry-C1	Kanamycin	4722
pmCherry-PAL	Kanamycin	5796
pNLS-65-HSF1-PAL-T2A-mCherry	Kanamycin	6884
pNLS-mCherry-PAL	Kanamycin	5817
pSEAP2-Basic	Ampicillin	4677

IX.11 Software

Table 28: Software

Software	Manufacturer	Version
ApE- A plasmid Editor	M. Wayne Davis	0.55
BD FACSDiva Software	Becton Dickinson, Franklin Lakes, USA	6.1.2
Emboss Needle	EMBL-EBI, Hinxton, UK	State: 12/19
FlowJo	Becton Dickinson, Franklin Lakes, USA	9.6.3
Gel Documentation Imager	GeneCapture (VWR, Leuven, Belgium)	
GraphPad Prism	GaphPad Software, San Diego, USA	5.00
Illustrations	Adobe Illustrator (California, USA)	CS5
	Adobe Photoshop (California, USA)	CS5
ImageJ	National Institutes of Health (USA)	1.50i
Microscopy image analysis	Zeiss ZEN (black/blue)	2012
Spectrophotometer measurement	NanoDrop Reader (Thermo Fisher)	
Reverse Compliment	Paul Stothard	State: 12/19

IX.12 Antibiotics

Table 29: Antibiotics

Kanamycin, 100mg/mL

Ampicilin: 50 mg/mL

Chloramphenicol : 33 mg/ml (RIL)

Penicillin Streptomycin (Gibco), PENSTREP: 1%

X Literature

1. The light fantastic. *Nat Chem Biol* **10**, 483 (2014).
2. Chen, A.H. & Silver, P.A. Designing biological compartmentalization. *Trends Cell Biol* **22**, 662-670 (2012).
3. Ferreira, R., Nilsson, J.R., Solano, C., Andreasson, J. & Grotli, M. Design, Synthesis and Inhibitory Activity of Photoswitchable RET Kinase Inhibitors. *Sci Rep* **5**, 9769 (2015).
4. Deisseroth, K. Optogenetics. *Nat Methods* **8**, 26-29 (2011).
5. Rockwell, N.C., Martin, S.S., Feoktistova, K. & Lagarias, J.C. Diverse two-cysteine photocycles in phytochromes and cyanobacteriochromes. *Proc Natl Acad Sci U S A* **108**, 11854-11859 (2011).
6. Fenno, L., Yizhar, O. & Deisseroth, K. The development and application of optogenetics. *Annu Rev Neurosci* **34**, 389-412 (2011).
7. Boyden, E.S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nat Neurosci* **8**, 1263-1268 (2005).
8. Aravanis, A.M. et al. An optical neural interface: in vivo control of rodent motor cortex with integrated fiberoptic and optogenetic technology. *J Neural Eng* **4**, S143-156 (2007).
9. Bacchus, W. & Fussenegger, M. The use of light for engineered control and reprogramming of cellular functions. *Curr Opin Biotechnol* **23**, 695-702 (2012).
10. Shcherbakova, D.M., Shemetov, A.A., Kaberniuk, A.A. & Verkhusha, V.V. Natural photoreceptors as a source of fluorescent proteins, biosensors, and optogenetic tools. *Annu Rev Biochem* **84**, 519-550 (2015).
11. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat Biotechnol* **33**, 755-760 (2015).
12. Li, F. et al. Optogenetic gene editing in regional skin. *Cell Res* **29**, 862-865 (2019).
13. Eleftheriou, C., Cesca, F., Maragliano, L., Benfenati, F. & Maya-Vetencourt, J.F. Optogenetic Modulation of Intracellular Signalling and Transcription: Focus on Neuronal Plasticity. *J Exp Neurosci* **11**, 1179069517703354 (2017).
14. Shao, J. et al. Synthetic far-red light-mediated CRISPR-dCas9 device for inducing functional neuronal differentiation. *Proc Natl Acad Sci U S A* **115**, E6722-E6730 (2018).
15. Polesskaya, O. et al. Optogenetic regulation of transcription. *BMC Neurosci* **19**, 12 (2018).
16. Strack, R. Controlling mRNA localization with light. *Nat Methods* **17**, 362 (2020).

17. Xue, G., Wang, K., Zhou, D., Zhong, H. & Pan, Z. Light-Induced Protein Degradation with Photocaged PROTACs. *J Am Chem Soc* **141**, 18370-18374 (2019).
18. Baaske, J. et al. Dual-controlled optogenetic system for the rapid down-regulation of protein levels in mammalian cells. *Sci Rep* **8**, 15024 (2018).
19. Vogt, N. Optogenetic inhibition. *Nat Methods* **17**, 29 (2020).
20. Levskaya, A., Weiner, O.D., Lim, W.A. & Voigt, C.A. Spatiotemporal control of cell signalling using a light-switchable protein interaction. *Nature* **461**, 997-1001 (2009).
21. Gautier, A. et al. How to control proteins with light in living systems. *Nat Chem Biol* **10**, 533-541 (2014).
22. van der Horst, M.A. & Hellingwerf, K.J. Photoreceptor proteins, "star actors of modern times": a review of the functional dynamics in the structure of representative members of six different photoreceptor families. *Acc Chem Res* **37**, 13-20 (2004).
23. Christie, J.M., Gawthorne, J., Young, G., Fraser, N.J. & Roe, A.J. LOV to BLUF: flavoprotein contributions to the optogenetic toolkit. *Mol Plant* **5**, 533-544 (2012).
24. Rockwell, N.C. & Lagarias, J.C. A brief history of phytochromes. *Chemphyschem* **11**, 1172-1180 (2010).
25. Losi, A. & Gartner, W. The evolution of flavin-binding photoreceptors: an ancient chromophore serving trendy blue-light sensors. *Annu Rev Plant Biol* **63**, 49-72 (2012).
26. Hartmann, D., Smith, J.M., Mazzotti, G., Chowdhry, R. & Booth, M.J. Controlling gene expression with light: a multidisciplinary endeavour. *Biochem Soc Trans* (2020).
27. Hartmann, D., Smith, J.M., Mazzotti, G., Chowdhry, R. & Booth, M.J. Controlling gene expression with light: a multidisciplinary endeavour. *Biochem Soc Trans* **48**, 1645-1659 (2020).
28. Devor, A. et al. Frontiers in optical imaging of cerebral blood flow and metabolism. *J Cereb Blood Flow Metab* **32**, 1259-1276 (2012).
29. Hughes, R.M. A compendium of chemical and genetic approaches to light-regulated gene transcription. *Crit Rev Biochem Mol Biol* **53**, 453-474 (2018).
30. Faal, T. et al. 4-Hydroxytamoxifen probes for light-dependent spatiotemporal control of Cre-ER mediated reporter gene expression. *Mol Biosyst* **11**, 783-790 (2015).
31. Wong, P.T. et al. Control of an Unusual Photo-Claisen Rearrangement in Coumarin Caged Tamoxifen through an Extended Spacer. *ACS Chem Biol* **12**, 1001-1010 (2017).
32. Gorka, A.P., Yamamoto, T., Zhu, J. & Schnermann, M.J. Cyanine Photocages Enable Spatial Control of Inducible Cre-Mediated Recombination. *Chembiochem* **19**, 1239-1243 (2018).
33. Zhang, W. et al. Control of Protein Activity and Gene Expression by Cyclofen-OH Uncaging. *Chembiochem* **19**, 1232-1238 (2018).
34. Fournier, L. et al. A blue-absorbing photolabile protecting group for in vivo chromatically orthogonal photoactivation. *ACS Chem Biol* **8**, 1528-1536 (2013).
35. Cambridge, S.B. et al. Doxycycline-dependent photoactivated gene expression in eukaryotic systems. *Nat Methods* **6**, 527-531 (2009).
36. Deiters, A. et al. Photocaged morpholino oligomers for the light-regulation of gene function in zebrafish and Xenopus embryos. *J Am Chem Soc* **132**, 15644-15650 (2010).
37. Tang, X., Swaminathan, J., Gewirtz, A.M. & Dmochowski, I.J. Regulating gene expression in human leukemia cells using light-activated oligodeoxynucleotides. *Nucleic Acids Res* **36**, 559-569 (2008).
38. Tang, X., Maegawa, S., Weinberg, E.S. & Dmochowski, I.J. Regulating gene expression in zebrafish embryos using light-activated, negatively charged peptide nucleic acids. *J Am Chem Soc* **129**, 11000-11001 (2007).
39. Ouyang, X. et al. Versatile synthesis and rational design of caged morpholinos. *J Am Chem Soc* **131**, 13255-13269 (2009).
40. Chou, C. & Deiters, A. Light-activated gene editing with a photocaged zinc-finger nuclease. *Angew Chem Int Ed Engl* **50**, 6839-6842 (2011).
41. Luo, J. et al. Genetically encoded optical activation of DNA recombination in human cells. *Chem Commun (Camb)* **52**, 8529-8532 (2016).

42. Brown, W. & Deiters, A. Light-activation of Cre recombinase in zebrafish embryos through genetic code expansion. *Methods Enzymol* **624**, 265-281 (2019).
43. Shimizu-Sato, S., Huq, E., Tepperman, J.M. & Quail, P.H. A light-switchable gene promoter system. *Nat Biotechnol* **20**, 1041-1044 (2002).
44. Muller, K. et al. A red/far-red light-responsive bi-stable toggle switch to control gene expression in mammalian cells. *Nucleic Acids Res* **41**, e77 (2013).
45. Uda, Y. et al. Efficient synthesis of phycocyanobilin in mammalian cells for optogenetic control of cell signaling. *Proc Natl Acad Sci U S A* **114**, 11962-11967 (2017).
46. Kyriakakis, P. et al. Biosynthesis of Orthogonal Molecules Using Ferredoxin and Ferredoxin-NADP(+) Reductase Systems Enables Genetically Encoded PhyB Optogenetics. *ACS Synth Biol* **7**, 706-717 (2018).
47. Kaberniuk, A.A., Shemetov, A.A. & Verkhusha, V.V. A bacterial phytochrome-based optogenetic system controllable with near-infrared light. *Nat Methods* **13**, 591-597 (2016).
48. Redchuk, T.A., Omelina, E.S., Chernov, K.G. & Verkhusha, V.V. Near-infrared optogenetic pair for protein regulation and spectral multiplexing. *Nat Chem Biol* **13**, 633-639 (2017).
49. Moglich, A., Ayers, R.A. & Moffat, K. Design and signaling mechanism of light-regulated histidine kinases. *J Mol Biol* **385**, 1433-1444 (2009).
50. Schwerdtfeger, C. & Linden, H. VIVID is a flavoprotein and serves as a fungal blue light photoreceptor for photoadaptation. *EMBO J* **22**, 4846-4855 (2003).
51. Wang, X., Chen, X. & Yang, Y. Spatiotemporal control of gene expression by a light-switchable transgene system. *Nat Methods* **9**, 266-269 (2012).
52. Konermann, S. et al. Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472-476 (2013).
53. Cao, J., Arha, M., Sudrik, C., Schaffer, D.V. & Kane, R.S. Bidirectional regulation of mRNA translation in mammalian cells by using PUF domains. *Angew Chem Int Ed Engl* **53**, 4900-4904 (2014).
54. Kim, N.Y. et al. Optogenetic control of mRNA localization and translation in live cells. *Nat Cell Biol* **22**, 341-352 (2020).
55. Crosson, S. & Moffat, K. Structure of a flavin-binding plant photoreceptor domain: insights into light-mediated signal transduction. *Proc Natl Acad Sci U S A* **98**, 2995-3000 (2001).
56. Nash, A.I. et al. Structural basis of photosensitivity in a bacterial light-oxygen-voltage/helix-turn-helix (LOV-HTH) DNA-binding protein. *Proc Natl Acad Sci U S A* **108**, 9449-9454 (2011).
57. Niopek, D., Wehler, P., Roensch, J., Eils, R. & Di Ventura, B. Optogenetic control of nuclear protein export. *Nat Commun* **7**, 10624 (2016).
58. Motta-Mena, L.B. et al. An optogenetic gene expression system with rapid activation and deactivation kinetics. *Nat Chem Biol* **10**, 196-202 (2014).
59. Reade, A. et al. TAEL: a zebrafish-optimized optogenetic gene expression system with fine spatial and temporal control. *Development* **144**, 345-355 (2017).
60. Strickland, D. et al. TULIPs: tunable, light-controlled interacting protein tags for cell biology. *Nat Methods* **9**, 379-384 (2012).
61. Hallett, R.A., Zimmerman, S.P., Yumerefendi, H., Bear, J.E. & Kuhlman, B. Correlating in Vitro and in Vivo Activities of Light-Inducible Dimers: A Cellular Optogenetics Guide. *ACS Synth Biol* **5**, 53-64 (2016).
62. Yamada, M., Suzuki, Y., Nagasaki, S.C., Okuno, H. & Imayoshi, I. Light Control of the Tet Gene Expression System in Mammalian Cells. *Cell Rep* **25**, 487-500 e486 (2018).
63. Kawano, F., Suzuki, H., Furuya, A. & Sato, M. Engineered pairs of distinct photoswitches for optogenetic control of cellular proteins. *Nat Commun* **6**, 6256 (2015).
64. Nihongaki, Y. et al. CRISPR-Cas9-based photoactivatable transcription systems to induce neuronal differentiation. *Nat Methods* **14**, 963-966 (2017).
65. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol* **169**, 5429-5433 (1987).

66. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S.D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551-2561 (2005).
67. Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**, 174-182 (2005).
68. Jansen, R., Embden, J.D., Gaastra, W. & Schouls, L.M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**, 1565-1575 (2002).
69. Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I. & Koonin, E.V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1**, 7 (2006).
70. Haft, D.H., Selengut, J., Mongodin, E.F. & Nelson, K.E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* **1**, e60 (2005).
71. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712 (2007).
72. Brouns, S.J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964 (2008).
73. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**, E2579-2586 (2012).
74. Westra, E.R. et al. CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell* **46**, 595-605 (2012).
75. Staals, R.H.J. et al. Structure and activity of the RNA-targeting Type III-B CRISPR-Cas complex of Thermus thermophilus. *Mol Cell* **52**, 135-145 (2013).
76. Barrangou, R. & Marraffini, L.A. CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol Cell* **54**, 234-244 (2014).
77. Plagens, A., Richter, H., Charpentier, E. & Randau, L. DNA and RNA interference mechanisms by CRISPR-Cas surveillance complexes. *FEMS Microbiol Rev* **39**, 442-463 (2015).
78. Rath, D., Amlinger, L., Rath, A. & Lundgren, M. The CRISPR-Cas immune system: biology, mechanisms and applications. *Biochimie* **117**, 119-128 (2015).
79. Makarova, K.S. et al. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9**, 467-477 (2011).
80. Vestergaard, G., Garrett, R.A. & Shah, S.A. CRISPR adaptive immune systems of Archaea. *RNA Biol* **11**, 156-167 (2014).
81. Heler, R., Marraffini, L.A. & Bikard, D. Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Mol Microbiol* **93**, 1-9 (2014).
82. Heler, R. et al. Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**, 199-202 (2015).
83. Datsenko, K.A. et al. Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun* **3**, 945 (2012).
84. Nunez, J.K. et al. Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol* **21**, 528-534 (2014).
85. van der Oost, J., Westra, E.R., Jackson, R.N. & Wiedenheft, B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol* **12**, 479-492 (2014).
86. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).
87. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013).
88. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).

89. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949 (2014).
90. Mojica, F.J.M., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733-740 (2009).
91. Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607 (2011).
92. Hu, J.H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57-63 (2018).
93. Kleinstiver, B.P. et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481-485 (2015).
94. Carlson-Stevermer, J. et al. Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. *Nat Commun* **8**, 1711 (2017).
95. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771 (2015).
96. Campa, C.C., Weisbach, N.R., Santinha, A.J., Incarnato, D. & Platt, R.J. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. *Nat Methods* **16**, 887-893 (2019).
97. Zetsche, B. et al. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat Biotechnol* **35**, 31-34 (2017).
98. Cox, D.B.T. et al. RNA editing with CRISPR-Cas13. *Science* **358**, 1019-1027 (2017).
99. Abudayyeh, O.O. et al. RNA targeting with CRISPR-Cas13. *Nature* **550**, 280-284 (2017).
100. Abudayyeh, O.O. et al. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**, aaf5573 (2016).
101. Yang, H. & Patel, D.J. CasX: a new and small CRISPR gene-editing protein. *Cell Res* **29**, 345-346 (2019).
102. Pausch, P. et al. CRISPR-CasPhi from huge phages is a hypercompact genome editor. *Science* **369**, 333-337 (2020).
103. Wang, S.R. et al. Conditional control of RNA-guided nucleic acid cleavage and gene editing. *Nat Commun* **11**, 91 (2020).
104. Chen, H. et al. Aptazyme-mediated direct modulation of post-transcriptional sgRNA level for conditional genome editing and gene expression. *J Biotechnol* **288**, 23-29 (2018).
105. Lin, B. et al. Control of CRISPR-Cas9 with small molecule-activated allosteric aptamer regulating sgRNAs. *Chem Commun (Camb)* **55**, 12223-12226 (2019).
106. Richter, F. et al. Engineering of temperature- and light-switchable Cas9 variants. *Nucleic Acids Res* **44**, 10003-10014 (2016).
107. Aschenbrenner, S. et al. Coupling Cas9 to artificial inhibitory domains enhances CRISPR-Cas9 target specificity. *Sci Adv* **6**, eaay0187 (2020).
108. Bubeck, F. et al. Engineered anti-CRISPR proteins for optogenetic control of CRISPR-Cas9. *Nat Methods* **15**, 924-927 (2018).
109. Hoffmann, M.D. et al. Optogenetic control of Neisseria meningitidis Cas9 genome editing using an engineered, light-switchable anti-CRISPR protein. *Nucleic Acids Res* (2020).
110. Ando, R., Mizuno, H. & Miyawaki, A. Regulated fast nucleocytoplasmic shuttling observed by reversible protein highlighting. *Science* **306**, 1370-1373 (2004).
111. Zhou, X.X. et al. A Single-Chain Photoswitchable CRISPR-Cas9 Architecture for Light-Inducible Gene Editing and Transcription. *ACS Chem Biol* **13**, 443-448 (2018).
112. Shao, J. et al. Smartphone-controlled optogenetically engineered cells enable semiautomatic glucose homeostasis in diabetic mice. *Sci Transl Med* **9** (2017).
113. Yu, Y. et al. Engineering a far-red light-activated split-Cas9 system for remote-controlled genome editing of internal organs and tumors. *Sci Adv* **6**, eabb1777 (2020).
114. Qi, L.S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183 (2013).

115. Maeder, M.L. et al. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* **10**, 977-979 (2013).
116. Perez-Pinera, P. et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods* **10**, 973-976 (2013).
117. Mali, P. et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31**, 833-838 (2013).
118. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583-588 (2015).
119. Hilton, I.B. et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* **33**, 510-517 (2015).
120. Thakore, P.I. et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods* **12**, 1143-1149 (2015).
121. Vojta, A. et al. Repurposing the CRISPR-Cas9 system for targeted DNA methylation. *Nucleic Acids Res* **44**, 5615-5628 (2016).
122. Josipovic, G., Zoldos, V. & Vojta, A. Active fusions of Cas9 orthologs. *J Biotechnol* **301**, 18-23 (2019).
123. Hannon, G.J. RNA interference. *Nature* **418**, 244-251 (2002).
124. Beerli, R.R. & Barbas, C.F., 3rd Engineering polydactyl zinc-finger transcription factors. *Nat Biotechnol* **20**, 135-141 (2002).
125. Zhang, F. et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* **29**, 149-153 (2011).
126. Klug, A. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu Rev Biochem* **79**, 213-231 (2010).
127. Gilbert, L.A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451 (2013).
128. Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S. & Vale, R.D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635-646 (2014).
129. Zhou, H. et al. In vivo simultaneous transcriptional activation of multiple genes in the brain using CRISPR-dCas9-activator transgenic mice. *Nat Neurosci* **21**, 440-446 (2018).
130. Chavez, A. et al. Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* **12**, 326-328 (2015).
131. Liao, H.K. et al. In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans-epigenetic Modulation. *Cell* **171**, 1495-1507 e1415 (2017).
132. Dahlman, J.E. et al. Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat Biotechnol* **33**, 1159-1161 (2015).
133. Kiani, S. et al. Cas9 gRNA engineering for genome editing, activation and repression. *Nat Methods* **12**, 1051-1054 (2015).
134. Zalatan, J.G. et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339-350 (2015).
135. Nissim, L., Perli, S.D., Fridkin, A., Perez-Pinera, P. & Lu, T.K. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol Cell* **54**, 698-710 (2014).
136. Kiani, S. et al. CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat Methods* **11**, 723-726 (2014).
137. Liu, Y. et al. Synthesizing AND gate genetic circuits based on CRISPR-Cas9 for identification of bladder cancer cells. *Nat Commun* **5**, 5393 (2014).
138. Gander, M.W., Vrana, J.D., Voje, W.E., Carothers, J.M. & Klavins, E. Digital logic circuits in yeast with CRISPR-dCas9 NOR gates. *Nat Commun* **8**, 15459 (2017).
139. Famulok, M. & Mayer, G. Aptamers and SELEX in Chemistry & Biology. *Chem Biol* **21**, 1055-1058 (2014).
140. Stoltenburg, R., Reinemann, C. & Strehlitz, B. SELEX--a (r)evolutionary method to generate high-affinity nucleic acid ligands. *Biomol Eng* **24**, 381-403 (2007).

141. Witherell, G.W., Gott, J.M. & Uhlenbeck, O.C. Specific interaction between RNA phage coat proteins and RNA. *Prog Nucleic Acid Res Mol Biol* **40**, 185-220 (1991).
142. Lim, F. & Peabody, D.S. RNA recognition site of PP7 coat protein. *Nucleic Acids Res* **30**, 4138-4144 (2002).
143. Ellington, A.D. & Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822 (1990).
144. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510 (1990).
145. Kaur, H., Bruno, J.G., Kumar, A. & Sharma, T.K. Aptamers in the Therapeutics and Diagnostics Pipelines. *Theranostics* **8**, 4016-4032 (2018).
146. Pfeiffer, F. & Mayer, G. Selection and Biosensor Application of Aptamers for Small Molecules. *Front Chem* **4**, 25 (2016).
147. Lee, J.W., Kim, H.J. & Heo, K. Therapeutic aptamers: developmental potential as anticancer drugs. *BMB Rep* **48**, 234-237 (2015).
148. Sundaram, P., Kurniawan, H., Byrne, M.E. & Wower, J. Therapeutic RNA aptamers in clinical trials. *Eur J Pharm Sci* **48**, 259-271 (2013).
149. Famulok, M., Blind, M. & Mayer, G. Intramers as promising new tools in functional proteomics. *Chem Biol* **8**, 931-939 (2001).
150. Ishizaki, J., Nevins, J.R. & Sullenger, B.A. Inhibition of cell proliferation by an RNA ligand that selectively blocks E2F function. *Nat Med* **2**, 1386-1389 (1996).
151. Martell, R.E., Nevins, J.R. & Sullenger, B.A. Optimizing aptamer activity for gene therapy applications using expression cassette SELEX. *Mol Ther* **6**, 30-34 (2002).
152. Shi, H., Hoffman, B.E. & Lis, J.T. RNA aptamers as effective protein antagonists in a multicellular organism. *Proc Natl Acad Sci U S A* **96**, 10033-10038 (1999).
153. Lebruska, L.L. & Maher, L.J., 3rd Selection and characterization of an RNA decoy for transcription factor NF-kappa B. *Biochemistry* **38**, 3168-3174 (1999).
154. Blind, M., Kolanus, W. & Famulok, M. Cytoplasmic RNA modulators of an inside-out signal-transduction cascade. *Proc Natl Acad Sci U S A* **96**, 3606-3610 (1999).
155. Mayer, G. et al. Controlling small guanine-nucleotide-exchange factor function through cytoplasmic RNA intramers. *Proc Natl Acad Sci U S A* **98**, 4961-4965 (2001).
156. Lennarz, S., Heider, E., Blind, M. & Mayer, G. An aptamer to the MAP kinase insert region. *ACS Chem Biol* **10**, 320-327 (2015).
157. Urak, K.T. et al. RNA inhibitors of nuclear proteins responsible for multiple organ dysfunction syndrome. *Nat Commun* **10**, 116 (2019).
158. Pan, L. et al. Aptamer-based regulation of transcription circuits. *Chem Commun (Camb)* **55**, 7378-7381 (2019).
159. Trachman, R.J., 3rd, Truong, L. & Ferre-D'Amare, A.R. Structural Principles of Fluorescent RNA Aptamers. *Trends Pharmacol Sci* **38**, 928-939 (2017).
160. Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H. & Sato, M. CRISPR-Cas9-based photoactivatable transcription system. *Chem Biol* **22**, 169-174 (2015).
161. Polstein, L.R. & Gersbach, C.A. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nat Chem Biol* **11**, 198-200 (2015).
162. Putri, R.R. & Chen, L. Spatiotemporal control of zebrafish (Danio rerio) gene expression using a light-activated CRISPR activation system. *Gene* **677**, 273-279 (2018).
163. Zetsche, B., Volz, S.E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol* **33**, 139-142 (2015).
164. Zhou, X.X., Fan, L.Z., Li, P., Shen, K. & Lin, M.Z. Optical control of cell signaling by single-chain photoswitchable kinases. *Science* **355**, 836-842 (2017).
165. Kim, J.H. et al. LADL: light-activated dynamic looping for endogenous gene expression control. *Nat Methods* **16**, 633-639 (2019).
166. Weber, A.M. et al. A blue light receptor that mediates RNA binding and translational regulation. *Nat Chem Biol* **15**, 1085-1092 (2019).

167. Moglich, A., Ayers, R.A. & Moffat, K. Structure and signaling mechanism of Per-ARNT-Sim domains. *Structure* **17**, 1282-1294 (2009).
168. Lin, J.T. & Stewart, V. Nitrate and nitrite-mediated transcription antitermination control of nasF (nitrate assimilation) operon expression in Klebsiella pheumoniae M5al. *J Mol Biol* **256**, 423-435 (1996).
169. Stewart, V. & van Tilbeurgh, H. Found: the elusive ANTAR transcription antiterminator. *PLoS Genet* **8**, e1002773 (2012).
170. Taylor, B.L. & Zhulin, I.B. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol Mol Biol Rev* **63**, 479-506 (1999).
171. Banerjee, R. & Batschauer, A. Plant blue-light receptors. *Planta* **220**, 498-502 (2005).
172. Christie, J.M., Salomon, M., Nozue, K., Wada, M. & Briggs, W.R. LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (nph1): binding sites for the chromophore flavin mononucleotide. *Proc Natl Acad Sci U S A* **96**, 8779-8783 (1999).
173. Salomon, M. et al. An optomechanical transducer in the blue light receptor phototropin from Avena sativa. *Proc Natl Acad Sci U S A* **98**, 12357-12361 (2001).
174. Buckley, A.M., Petersen, J., Roe, A.J., Douce, G.R. & Christie, J.M. LOV-based reporters for fluorescence imaging. *Curr Opin Chem Biol* **27**, 39-45 (2015).
175. Zayner, J.P. & Sosnick, T.R. Factors that control the chemistry of the LOV domain photocycle. *PLoS One* **9**, e87074 (2014).
176. Kaur Grewal, R., Mitra, D. & Roy, S. Mapping networks of light-dark transition in LOV photoreceptors. *Bioinformatics* **31**, 3608-3616 (2015).
177. Yee, E.F. et al. Signal transduction in light-oxygen-voltage receptors lacking the adduct-forming cysteine residue. *Nat Commun* **6**, 10079 (2015).
178. Glantz, S.T. et al. Functional and topological diversity of LOV domain photoreceptors. *Proc Natl Acad Sci U S A* **113**, E1442-1451 (2016).
179. Kasahara, M. et al. Photochemical properties of the flavin mononucleotide-binding domains of the phototropins from Arabidopsis, rice, and Chlamydomonas reinhardtii. *Plant Physiol* **129**, 762-773 (2002).
180. Losi, A. & Gartner, W. Old chromophores, new photoactivation paradigms, trendy applications: flavins in blue light-sensing photoreceptors. *Photochem Photobiol* **87**, 491-510 (2011).
181. Huala, E. et al. Arabidopsis NPH1: a protein kinase with a putative redox-sensing domain. *Science* **278**, 2120-2123 (1997).
182. Losi, A., Gardner, K.H. & Moglich, A. Blue-Light Receptors for Optogenetics. *Chem Rev* **118**, 10659-10709 (2018).
183. Wright, A.V. et al. Rational design of a split-Cas9 enzyme complex. *Proc Natl Acad Sci U S A* **112**, 2984-2989 (2015).
184. Shaner, N.C. et al. Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol* **22**, 1567-1572 (2004).
185. Labun, K., Montague, T.G., Gagnon, J.A., Thyme, S.B. & Valen, E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res* **44**, W272-276 (2016).
186. Anders, C. & Jinek, M. In vitro enzymology of Cas9. *Methods Enzymol* **546**, 1-20 (2014).
187. Sundaresan, R., Parameshwaran, H.P., Yogesha, S.D., Keilbarth, M.W. & Rajan, R. RNA-Independent DNA Cleavage Activities of Cas9 and Cas12a. *Cell Rep* **21**, 3728-3739 (2017).
188. Liang, C. et al. Tumor cell-targeted delivery of CRISPR/Cas9 by aptamer-functionalized lipopolymer for therapeutic genome editing of VEGFA in osteosarcoma. *Biomaterials* **147**, 68-85 (2017).
189. Beyer, H.M. et al. AQUA Cloning: A Versatile and Simple Enzyme-Free Cloning Approach. *PLoS One* **10**, e0137652 (2015).
190. Garcia-Nafria, J., Watson, J.F. & Greger, I.H. IVA cloning: A single-tube universal cloning system exploiting bacterial In Vivo Assembly. *Sci Rep* **6**, 27459 (2016).
191. Egorova, A. et al. Chemokine-derived peptides as carriers for gene delivery to CXCR4 expressing cells. *J Gene Med* **11**, 772-781 (2009).

192. Uhlen, M. et al. Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
193. Fidanza, A. et al. An all-in-one UniSam vector system for efficient gene activation. *Sci Rep* **7**, 6394 (2017).
194. Weis, D. & Di Ventura, B. Optogenetic Control of Nucleocytoplasmic Protein Transport. *Methods Mol Biol* **2173**, 127-136 (2020).
195. Kosugi, S., Hasebe, M., Tomita, M. & Yanagawa, H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. *Proc Natl Acad Sci U S A* **106**, 10171-10176 (2009).
196. Pilsl, S., Morgan, C., Choukeife, M., Moglich, A. & Mayer, G. Optoribogenetic control of regulatory RNA molecules. *Nat Commun* **11**, 4825 (2020).
197. Lighting the way. *Nat Chem Biol* **12**, 381 (2016).
198. Chandrasegaran, S. & Carroll, D. Origins of Programmable Nucleases for Genome Engineering. *J Mol Biol* **428**, 963-989 (2016).
199. Liang, X. et al. Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *J Biotechnol* **208**, 44-53 (2015).
200. Zhang, X.H., Tee, L.Y., Wang, X.G., Huang, Q.S. & Yang, S.H. Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther Nucleic Acids* **4**, e264 (2015).
201. Nunez, J.K., Harrington, L.B. & Doudna, J.A. Chemical and Biophysical Modulation of Cas9 for Tunable Genome Engineering. *ACS Chem Biol* **11**, 681-688 (2016).
202. Richter, F. et al. Switchable Cas9. *Curr Opin Biotechnol* **48**, 119-126 (2017).
203. Davis, K.M., Pattanayak, V., Thompson, D.B., Zuris, J.A. & Liu, D.R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat Chem Biol* **11**, 316-318 (2015).
204. Liu, K.I. et al. A chemical-inducible CRISPR-Cas9 system for rapid control of genome editing. *Nat Chem Biol* **12**, 980-987 (2016).
205. Maji, B. et al. Multidimensional chemical control of CRISPR-Cas9. *Nat Chem Biol* **13**, 9-11 (2017).
206. Oakes, B.L. et al. Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat Biotechnol* **34**, 646-651 (2016).
207. Gao, Y. et al. Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nat Methods* **13**, 1043-1049 (2016).
208. Nguyen, D.P. et al. Ligand-binding domains of nuclear receptors facilitate tight control of split CRISPR activity. *Nat Commun* **7**, 12009 (2016).
209. Bao, Z., Jain, S., Jaroenpuntaruk, V. & Zhao, H. Orthogonal Genetic Regulation in Human Cells Using Chemically Induced CRISPR/Cas9 Activators. *ACS Synth Biol* **6**, 686-693 (2017).
210. Hemphill, J., Borchardt, E.K., Brown, K., Asokan, A. & Deiters, A. Optical Control of CRISPR/Cas9 Gene Editing. *J Am Chem Soc* **137**, 5642-5645 (2015).
211. Liu, Y. et al. Directing cellular information flow via CRISPR signal conductors. *Nat Methods* **13**, 938-944 (2016).
212. Tang, W., Hu, J.H. & Liu, D.R. Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nat Commun* **8**, 15939 (2017).
213. Ferry, Q.R., Lyutova, R. & Fulga, T.A. Rational design of inducible CRISPR guide RNAs for de novo assembly of transcriptional programs. *Nat Commun* **8**, 14633 (2017).
214. Lee, Y.J., Hoynes-O'Connor, A., Leong, M.C. & Moon, T.S. Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system. *Nucleic Acids Res* **44**, 2462-2473 (2016).
215. Robichon, C., Luo, J., Causey, T.B., Benner, J.S. & Samuelson, J.C. Engineering Escherichia coli BL21(DE3) derivative strains to minimize E. coli protein contamination after purification by immobilized metal affinity chromatography. *Appl Environ Microbiol* **77**, 4634-4646 (2011).
216. Carmignotto, G.P. & Azzoni, A.R. On the expression of recombinant Cas9 protein in E. coli BL21(DE3) and BL21(DE3) Rosetta strains. *J Biotechnol* **306**, 62-70 (2019).

217. Maruyama, F., Watanabe, T. & Nakagawa, I. in Streptococcus pyogenes : Basic Biology to Clinical Manifestations. (eds. J.J. Ferretti, D.L. Stevens & V.A. Fischetti) (Oklahoma City (OK); 2016).
218. Peters, J.M. et al. Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi. *Nat Microbiol* **4**, 244-250 (2019).
219. Brady, P.N. & Macnaughtan, M.A. Evaluation of colorimetric assays for analyzing reductively methylated proteins: Biases and mechanistic insights. *Anal Biochem* **491**, 43-51 (2015).
220. Knight, M.I. & Chambers, P.J. Problems associated with determining protein concentration: a comparison of techniques for protein estimations. *Mol Biotechnol* **23**, 19-28 (2003).
221. Labun, K. et al. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. *Nucleic Acids Res* **47**, W171-W174 (2019).
222. Wu, X., Kriz, A.J. & Sharp, P.A. Target specificity of the CRISPR-Cas9 system. *Quant Biol* **2**, 59-70 (2014).
223. Kuzmine, I., Gottlieb, P.A. & Martin, C.T. Binding of the priming nucleotide in the initiation of transcription by T7 RNA polymerase. *J Biol Chem* **278**, 2819-2823 (2003).
224. Wang, S., Su, J.H., Zhang, F. & Zhuang, X. An RNA-aptamer-based two-color CRISPR labeling system. *Sci Rep* **6**, 26857 (2016).
225. Nowak, C.M., Lawson, S., Zerez, M. & Bleris, L. Guide RNA engineering for versatile Cas9 functionality. *Nucleic Acids Res* **44**, 9555-9564 (2016).
226. Larson, M.H. et al. CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc* **8**, 2180-2196 (2013).
227. Lai, D., Proctor, J.R. & Meyer, I.M. On the importance of cotranscriptional RNA structure formation. *RNA* **19**, 1461-1473 (2013).
228. Mekler, V., Minakhin, L., Semenova, E., Kuznedelov, K. & Severinov, K. Kinetics of the CRISPR-Cas9 effector complex assembly and the role of 3'-terminal segment of guide RNA. *Nucleic Acids Res* **44**, 2837-2845 (2016).
229. Kundert, K. et al. Controlling CRISPR-Cas9 with ligand-activated and ligand-deactivated sgRNAs. *Nat Commun* **10**, 2127 (2019).
230. Chaiyarit, S. & Thongboonkerd, V. Comparative analyses of cell disruption methods for mitochondrial isolation in high-throughput proteomics study. *Anal Biochem* **394**, 249-258 (2009).
231. Romanienko, P.J. et al. A Vector with a Single Promoter for In Vitro Transcription and Mammalian Cell Expression of CRISPR gRNAs. *PLoS One* **11**, e0148362 (2016).
232. Ran, F.A. et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380-1389 (2013).
233. Zeng, B. et al. Expansion of CRISPR targeting sites in Bombyx mori. *Insect Biochem Mol Biol* **72**, 31-40 (2016).
234. Ding, Q. et al. Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* **12**, 393-394 (2013).
235. Ma, H. et al. Pol III Promoters to Express Small RNAs: Delineation of Transcription Initiation. *Mol Ther Nucleic Acids* **3**, e161 (2014).
236. Ranganathan, V., Wahlin, K., Maruotti, J. & Zack, D.J. Expansion of the CRISPR-Cas9 genome targeting space through the use of H1 promoter-expressed guide RNAs. *Nat Commun* **5**, 4516 (2014).
237. Zhou, H.X. & Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem Rev* **118**, 1691-1741 (2018).
238. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415 (2003).
239. Waugh, A. et al. RNAML: a standard syntax for exchanging RNA information. *RNA* **8**, 707-717 (2002).
240. Zuker, M. & Jacobson, A.B. Using reliability information to annotate RNA secondary structures. *RNA* **4**, 669-679 (1998).
241. Lee, J.C. & Gutell, R.R. Helix capping in RNA structure. *PLoS One* **9**, e93664 (2014).

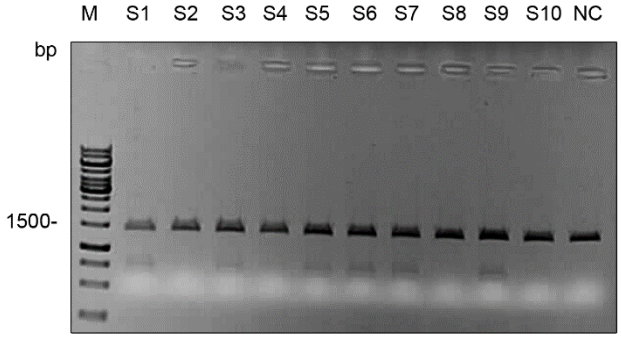
242. Kruth, H.S. Flow cytometry: rapid biochemical analysis of single cells. *Anal Biochem* **125**, 225-242 (1982).
243. Gilbert, L.A. et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**, 647-661 (2014).
244. Yang, J. et al. Genome-Scale CRISPRa Screen Identifies Novel Factors for Cellular Reprogramming. *Stem Cell Reports* **12**, 757-771 (2019).
245. Szymczak-Workman, A.L., Vignali, K.M. & Vignali, D.A. Design and construction of 2A peptide-linked multicistronic vectors. *Cold Spring Harb Protoc* **2012**, 199-204 (2012).
246. Chen, X., Zaro, J.L. & Shen, W.C. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev* **65**, 1357-1369 (2013).
247. Gibson, D.G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345 (2009).
248. Li, L., Jiang, W. & Lu, Y. A Modified Gibson Assembly Method for Cloning Large DNA Fragments with High GC Contents. *Methods Mol Biol* **1671**, 203-209 (2018).
249. Higuchi, R., Krummel, B. & Saiki, R.K. A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Res* **16**, 7351-7367 (1988).
250. Oftebro, R. & Wolf, I. Mitosis of bi- and multinucleate HeLa cells. *Exp Cell Res* **48**, 39-52 (1967).
251. Kapuscinski, J. DAPI: a DNA-specific fluorescent probe. *Biotech Histochem* **70**, 220-233 (1995).
252. Latt, S.A., Stetten, G., Juergens, L.A., Willard, H.F. & Scher, C.D. Recent developments in the detection of deoxyribonucleic acid synthesis by 33258 Hoechst fluorescence. *J Histochem Cytochem* **23**, 493-505 (1975).
253. Peabody, D.S. The RNA binding site of bacteriophage MS2 coat protein. *EMBO J* **12**, 595-600 (1993).
254. Valegard, K. et al. The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J Mol Biol* **270**, 724-738 (1997).
255. Sigrist, C.J. et al. New and continuing developments at PROSITE. *Nucleic Acids Res* **41**, D344-347 (2013).
256. Ramesh, A. et al. The mechanism for RNA recognition by ANTAR regulators of gene expression. *PLoS Genet* **8**, e1002666 (2012).
257. Jankowsky, E. & Harris, M.E. Specificity and nonspecificity in RNA-protein interactions. *Nat Rev Mol Cell Biol* **16**, 533-544 (2015).
258. Reddy Chichili, V.P., Kumar, V. & Sivaraman, J. Linkers in the structural biology of protein-protein interactions. *Protein Sci* **22**, 153-167 (2013).
259. Biggs, R. & Macmillan, R.L. The Errors of Some Haematological Methods as They Are Used in a Routine Laboratory. *J Clin Pathol* **1**, 269-287 (1948).
260. Davis, J.D. The evolution of the progressive-era hemocytometer. *Caduceus* **11**, 164-183 (1995).
261. Shapiro, H.M. "Cellular astronomy"--a foreseeable future in cytometry. *Cytometry A* **60**, 115-124 (2004).
262. Weber, M., Moller, K., Welzeck, M. & Schorr, J. Short technical reports. Effects of lipopolysaccharide on transfection efficiency in eukaryotic cells. *Biotechniques* **19**, 930-940 (1995).
263. Ehlert, F., Bierbaum, P. & Schorr, J. Importance of DNA quality for transfection efficiency. *Biotechniques* **14**, 546 (1993).
264. Hacker, G. The morphology of apoptosis. *Cell Tissue Res* **301**, 5-17 (2000).
265. Jazurek, M., Ciesiolka, A., Starega-Roslan, J., Bilinska, K. & Krzyzosiak, W.J. Identifying proteins that bind to specific RNAs - focus on simple repeat expansion diseases. *Nucleic Acids Res* **44**, 9050-9070 (2016).
266. Kalderon, D., Richardson, W.D., Markham, A.F. & Smith, A.E. Sequence requirements for nuclear location of simian virus 40 large-T antigen. *Nature* **311**, 33-38 (1984).

267. Hodel, M.R., Corbett, A.H. & Hodel, A.E. Dissection of a nuclear localization signal. *J Biol Chem* **276**, 1317-1325 (2001).
268. Nakashima, Y., Ohta, S. & Wolf, A.M. Blue light-induced oxidative stress in live skin. *Free Radic Biol Med* **108**, 300-310 (2017).
269. Wan, F. & Lenardo, M.J. Specification of DNA binding activity of NF-kappaB proteins. *Cold Spring Harb Perspect Biol* **1**, a000067 (2009).
270. Lecoq, L. et al. Structural characterization of interactions between transactivation domain 1 of the p65 subunit of NF-kappaB and transcription regulatory factors. *Nucleic Acids Res* **45**, 5564-5576 (2017).
271. Paal, K., Baeuerle, P.A. & Schmitz, M.L. Basal transcription factors TBP and TFIIB and the viral coactivator E1A 13S bind with distinct affinities and kinetics to the transactivation domain of NF-kappaB p65. *Nucleic Acids Res* **25**, 1050-1055 (1997).
272. Chatterjee, S. & Struhl, K. Connecting a promoter-bound protein to TBP bypasses the need for a transcriptional activation domain. *Nature* **374**, 820-822 (1995).
273. Klages, N. & Strubin, M. Stimulation of RNA polymerase II transcription initiation by recruitment of TBP in vivo. *Nature* **374**, 822-823 (1995).
274. Eastmond, D.L. & Nelson, H.C. Genome-wide analysis reveals new roles for the activation domains of the Saccharomyces cerevisiae heat shock transcription factor (Hsf1) during the transient heat shock response. *J Biol Chem* **281**, 32909-32921 (2006).
275. Sakurai, H., Hashikawa, N., Imazu, H. & Fukasawa, T. Carboxy-terminal region of the yeast heat shock factor contains two domains that make transcription independent of the TFIIH protein kinase. *Genes Cells* **8**, 951-961 (2003).
276. Chavez, A. et al. Comparison of Cas9 activators in multiple species. *Nat Methods* **13**, 563-567 (2016).
277. Sun, Y. et al. CXCL12-CXCR4 axis promotes the natural selection of breast cancer cell metastasis. *Tumour Biol* **35**, 7765-7773 (2014).
278. Murdoch, C., Monk, P.N. & Finn, A. Functional expression of chemokine receptor CXCR4 on human epithelial cells. *Immunology* **98**, 36-41 (1999).
279. Nelms, B.L. & Labosky, P.A. A predicted hairpin cluster correlates with barriers to PCR, sequencing and possibly BAC recombineering. *Sci Rep* **1**, 106 (2011).
280. Weaver, D.T. & DePamphilis, M.L. Specific sequences in native DNA that arrest synthesis by DNA polymerase alpha. *J Biol Chem* **257**, 2075-2086 (1982).
281. Hawkins, J.S., Wong, S., Peters, J.M., Almeida, R. & Qi, L.S. Targeted Transcriptional Repression in Bacteria Using CRISPR Interference (CRISPRi). *Methods Mol Biol* **1311**, 349-362 (2015).
282. Kallimasioti-Pazi, E.M. et al. Heterochromatin delays CRISPR-Cas9 mutagenesis but does not influence the outcome of mutagenic DNA repair. *PLoS Biol* **16**, e2005595 (2018).
283. Campenhout, C.V. et al. Guidelines for optimized gene knockout using CRISPR/Cas9. *Biotechniques* **66**, 295-302 (2019).
284. Brown, A. et al. Multiplexed and tunable transcriptional activation by promoter insertion using nuclease-assisted vector integration. *Nucleic Acids Res* **47**, e67 (2019).
285. Liu, Y. et al. CRISPR Activation Screens Systematically Identify Factors that Drive Neuronal Fate and Reprogramming. *Cell Stem Cell* **23**, 758-771 e758 (2018).
286. Martella, A. et al. Systematic Evaluation of CRISPRa and CRISPRi Modalities Enables Development of a Multiplexed, Orthogonal Gene Activation and Repression System. *ACS Synth Biol* **8**, 1998-2006 (2019).
287. Axelson, H. The Notch signaling cascade in neuroblastoma: role of the basic helix-loop-helix proteins HASH-1 and HES-1. *Cancer Lett* **204**, 171-178 (2004).
288. Svec, D., Tichopad, A., Novosadova, V., Pfaffl, M.W. & Kubista, M. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomol Detect Quantif* **3**, 9-16 (2015).
289. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol* **19**, 621-637 (2018).

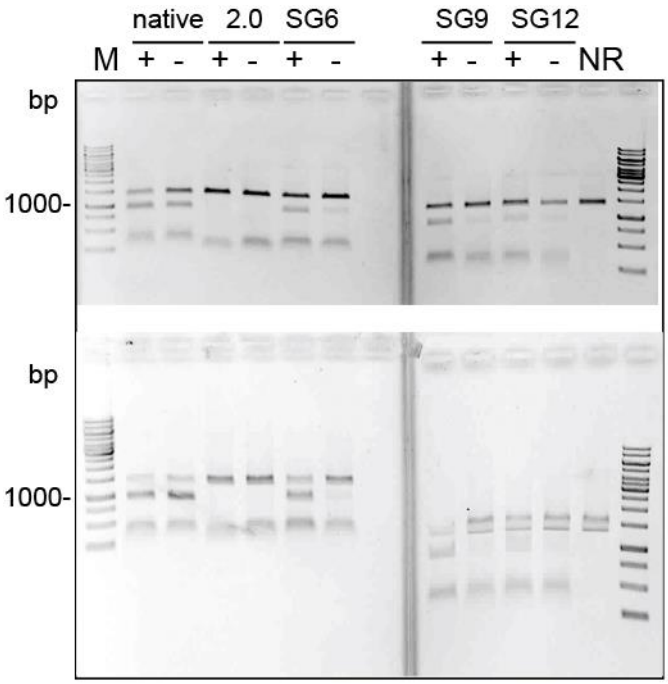
290. Yella, V.R., Kumar, A. & Bansal, M. Identification of putative promoters in 48 eukaryotic genomes on the basis of DNA free energy. *Sci Rep* **8**, 4520 (2018).
291. Wang, G. et al. Multiplexed activation of endogenous genes by CRISPRa elicits potent antitumor immunity. *Nat Immunol* **20**, 1494-1505 (2019).
292. Golmohammadi, R., Valegard, K., Fridborg, K. & Liljas, L. The refined structure of bacteriophage MS2 at 2.8 A resolution. *J Mol Biol* **234**, 620-639 (1993).
293. Erickson, H.P. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. *Biol Proced Online* **11**, 32-51 (2009).
294. Gossett, A.J. & Lieb, J.D. DNA Immunoprecipitation (DIP) for the Determination of DNA-Binding Specificity. *CSH Protoc* **2008**, pdb prot4972 (2008).
295. Wang, Y., Wang, F., Wang, R., Zhao, P. & Xia, Q. 2A self-cleaving peptide-based multi-gene expression system in the silkworm Bombyx mori. *Sci Rep* **5**, 16273 (2015).
296. Liu, Z. et al. Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci Rep* **7**, 2193 (2017).
297. Gelinas, A.D., Davies, D.R. & Janjic, N. Embracing proteins: structural themes in aptamer-protein complexes. *Curr Opin Struct Biol* **36**, 122-132 (2016).
298. Suter, D.M. Transcription Factors and DNA Play Hide and Seek. *Trends Cell Biol* **30**, 491-500 (2020).
299. Marfori, M. et al. Molecular basis for specificity of nuclear import and prediction of nuclear localization. *Biochim Biophys Acta* **1813**, 1562-1577 (2011).
300. Chen, P. et al. Retinal Neuron Is More Sensitive to Blue Light-Induced Damage than Glia Cell Due to DNA Double-Strand Breaks. *Cells* **8** (2019).
301. Ratnayake, K., Payton, J.L., Lakmal, O.H. & Karunarathne, A. Blue light excited retinal intercepts cellular signaling. *Sci Rep* **8**, 10207 (2018).
302. Niopek, D. et al. Engineering light-inducible nuclear localization signals for precise spatiotemporal control of protein dynamics in living cells. *Nat Commun* **5**, 4404 (2014).
303. Naso, M.F., Tomkowicz, B., Perry, W.L., 3rd & Strohl, W.R. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. *BioDrugs* **31**, 317-334 (2017).
304. Grieger, J.C. & Samulski, R.J. Packaging capacity of adeno-associated virus serotypes: impact of larger genomes on infectivity and postentry steps. *J Virol* **79**, 9933-9944 (2005).
305. Bohm, S. et al. A gene therapy for inherited blindness using dCas9-VPR-mediated transcriptional activation. *Sci Adv* **6**, eaba5614 (2020).
306. Wilbie, D., Walther, J. & Mastrobattista, E. Delivery Aspects of CRISPR/Cas for in Vivo Genome Editing. *Acc Chem Res* **52**, 1555-1564 (2019).
307. Wang, D. et al. Adenovirus-Mediated Somatic Genome Editing of Pten by CRISPR/Cas9 in Mouse Liver in Spite of Cas9-Specific Immune Responses. *Hum Gene Ther* **26**, 432-442 (2015).
308. Lattanzi, A. et al. Optimization of CRISPR/Cas9 Delivery to Human Hematopoietic Stem and Progenitor Cells for Therapeutic Genomic Rearrangements. *Mol Ther* **27**, 137-150 (2019).
309. Tan, Y. et al. Rationally engineered Staphylococcus aureus Cas9 nucleases with high genome-wide specificity. *Proc Natl Acad Sci U S A* **116**, 20969-20976 (2019).
310. Liu, J.J. et al. CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* **566**, 218-223 (2019).
311. Wagner, D.L. et al. High prevalence of Streptococcus pyogenes Cas9-reactive T cells within the adult human population. *Nat Med* **25**, 242-248 (2019).
312. Parsi, K.M., Hennessy, E., Kearns, N. & Maehr, R. Using an Inducible CRISPR-dCas9-KRAB Effector System to Dissect Transcriptional Regulation in Human Embryonic Stem Cells. *Methods Mol Biol* **1507**, 221-233 (2017).
313. Xu, X. & Qi, L.S. A CRISPR-dCas Toolbox for Genetic Engineering and Synthetic Biology. *J Mol Biol* **431**, 34-47 (2019).
314. Gerhardt, K.P. et al. An open-hardware platform for optogenetics and photobiology. *Sci Rep* **6**, 35363 (2016).

XI Appendix

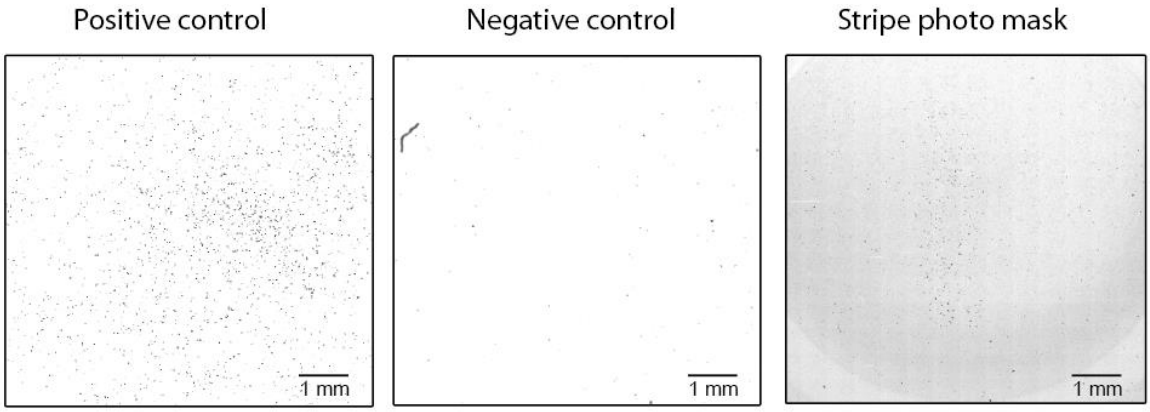
XI.1Supplemental Figures



Supporting Figure 1: 1% agarose gel with products from the Cas9 cleavage assay using various sgRNAs modified with PAL aptamers (N=2). M: Thermoscientific GeneRuler 1kb DNA Ladder S1: Cas9 cleavage reaction with native sgRNA targeting mCherry CDS (positive control). S2: Cas9 cleavage reaction with non-binding sgRNA. S3: Cas9 cleavage reaction with sg09.19di7 sgRNA. S4: Cas9 cleavage reaction with sg09.19dmu sgRNA. S5: Cas9 cleavage reaction with sg09.19tri sgRNA. S6: Cas9 cleavage reaction with sg09.23di sgRNA. S7: Cas9 cleavage reaction with sg09.23tri sgRNA. S8: Cas9 cleavage reaction with sg09.23dmu sgRNA. S9: Cas9 cleavage reaction with sg09.23 sgRNA. S10: Cas9 cleavage reaction with sg09.23t7 sgRNA. NC: substrate only. The gel was running for 30 min at 130V and stained with ethidium bromide.



Supporting Figure 2: Raw 1% agarose gel pictures of Cas9 cleavage assay. Top gel: N=1, Bottom gel: N=2. The gel was running for 30 min at 130V and stained with ethidium bromide.



Supporting Figure 3: Raw stitched images from LSM analysis of the photo mask experiment. Positive control: no mask, Negative control: fully masked (dark), Stripe photo mask: covert with a stripe to let the light through. The image was converted to black/white and a tone value correction was applied to the white parameter by setting it from 255 to 50. Contrast and brightness were adjusted and the colours were inverted for better visibility.

XI.2Sequences

XI.2.1 Primers

Table 30: Primers

Name	DNA sequence (5’ -> 3’)
gRNA_T7_fwd	GGGGGAATTCTAATACGACTCACTATA
sgRNA_native_rev	AAAAGCACCGACTCGGTG
mCherrySubstrate_fwd	GGGTCATTAGTTCATAGCCC
mCherrySubstrate_rev	TGGTATGGCTGATTATGATCAG
eBFP_fwd	GTACTGTTGGTAAAGCCACcATGGTGAGCAAGGGCGA
eBFP_rev	GGAG
pGL3-Basic-8x-gRNA_rev	GCTTAAGTTTAAACGCTAGCTCACTTGTACAGCTCGTC
pGL3-Basic-8x-gRNA_fwd	CATGCC
pEGFP-N1_fwd	gGTGGCTTTACCAACAGTACCGGAATG
pEGFP-N1_rev	
MS2-P1	GCTAGCGTTTAAACTTAAGCTTGGTACCGAG
MS2-P2	Catggtggcgaccggtg
PAL-P1	taaagcggccgcgactctagatc
PAL-P2	cgggcccgggatccaccggtcgccaccatgGCTTCAAACTTTACTCAGTTCGTG
PAL-P3	tgattatgatctagagtcgcggccgctttaCTTGTACAGCTCGTCCATGC
PAL-P4	cgggcccgggatccaccggtcgccaccatgCCTAAGAAAAAGAGGAAGGTGGC
PAL-P5	acccgaaccaccacccgaaccaccacccgaaccaccGGAGACAGTGGGGTCCTTG
PAL-P7	
PAL-P8	
T2A_mCherry_fwd	ggtggttcgggtggtggttcgggtggtggttcgggtAAGGTGAACCGGCCC

NHPTM-fwd	ACCGCATGTTAGCAGACTTCCTCTGCCCTC-CGACGCCAGCTGCTC
NHPTM-rev	GAGGGCAGAGGAAGTCTGCTA
p65-in-NHPTM-fwd	cgggcccgggat
	tgattatgatctagagtcgcggc
p65-in-NHPTM-rev	GAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTC GAGGAGAATCCTGGCCCAATGGTGAGCAAGGGCGAG
NHPTM-BB-fwd	AGTGGGCAGGGAGGAGGTGGAAGCGGCTTCAGCGT GGAC
NHPTM-BB-rev	caccttcctctttttcttaggcatggtggcgaccggtgg
NHPPTM_mut_corr_fwd	
NHPPTM_mut_corr_rev	CAAAGCCAAGGACCCCACTGTCTCCgcggccgctggatccC
NPHPTM_NLS_corr_fwd	CTTCAGGGCAG
NPHPTM_NLS_corr_rev	
dCas9_-VP64_fwd	
dCas9_-VP64_rev	CACCTTacccgaaccaccacccgaaccaccacccgaaccaccAGAG
	GAAATCTGTGACAGCAGGGCACTAAAGTCCATATCAG
eBFP_sgRNA_in_pENTR_fwd	CGATGC
	ggtggttcgggtggtggttcgggtggtggttcgggtAAGGTGAACCGGC
eBFP_sgRNA_in_pENTR_rev	CCGCGGAGC
	GGAGACAGTGGGGTCCTTGGCTTTGGGAGGCTCC
pENTR.hU6-F_BB	gAGTGGGCAGGGAGGAGGTGGAAGCGG
	accttcctctttttcttaggcatggtggcgaccggtg
pENTR.hU6-R_BB	agaggaaggtggcggccgctggatcc
	ttttcttaggcatggtggcgaccggtggatccc
	ATTAACGCTAGCGGCAGTGGAGAGGGCAGAGGAAG
M2S12_mut_fwd	GGCCGCCACCTTCCTCTTTTTCTTAGGTCCGCTACC
M2S12_mut_rev	ATCTTGTGGAAAGGACGAAACACCGGGAAAGGTCGA GAAACTGCAAA
CXCR4_sgRNA_in_pENTR_fwd	TACAAGAAAGCTGGGTCTAGAAAAAGCACCGACTCGG TGC

sgRNA_in_pENTR_rev	CTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCATTGCTTATCAATTTGTTGCAACGAACAGG
CXCR4_2.0_mut_fwd	GGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTCC
CXCR4_2.0_mut_rev	GCAAAGTTTTAGAGCTAGgcCAGGCCGCCGC
CXCR4_M2S9_mut_fwd	AGTTTCTCGACCTTTCCCGGTGTtTCGTCCTTTCCACAAGATA
CXCR4_M2S9_mut_rev	
CXCR4_M2S9_new_fwd	ATCTTGTGGAAAGGACGAAACACCGCAGACGCGAGGAAGGAGG
CXCR4_M2S9_new_rev	TACAAGAAAGCTGGGTCTAGAAAAAGCACCGACTCGGTGC
gRNA_insert_universal_rev	AAGGAGGGCGCGTTTTAGAGCTAGGCCAACATGAGG CCTCGCGTCTGCGGTGTTTCGTCCTTTCCACA GGCCAGGCCGCCGCGTACAGC
gRNA_M2S9_sgCXCR4.1_fwd	TAGCTCTAAAACGCGCCCTCCTTCCTCGCGT
gRNA_M2S9_sgCXCR4.2_fwd	GCAGACGCGAGGAAGGAGGGCGCGTTTTAGAGCTAGGCCAGGCCGCCGCG GGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGCATATG
gRNA_M2S9_sgCXCR4.3_fwd	CGGTGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAATCGAAATACTTTCAAGTTACGGTAAGC
gRNA_M2S9_sgCXCR4.4_fwd	GGCTAGGAACGCGTCTCTCTGGTTTTAGAGCTAGGCCAGGCCGCCGCGTAC
gRNA_M2S9_sgCXCR4.5_fwd	GCCTGAAGACAGGTGGGAAGCGCGTTTTAGAGCTAGGCCAGGCCGCCGCGTAC
gRNA_M2S9_sgCXCR4.7_fwd	GAGCCGGACAGGACCTCCCAGGTTTTAGAGCTAGGCCAGGCCGCCGCGTAC GCGGGTGGTCGGTAGTGAGTCGTTTTAGAGCTAGGCCAGGCCGCCGCGTAC

gRNA_M2S9_sgCXCR4.8_fwd	GGACCCTGCTGTTTGCGGGTGGTGTTTTAGAGCTAGGCCAGGCCGCCGCGTAC
gRNA_M2S9_sgCXCR4.9_fwd	GCAAGTCACTCCCCTTCCCTGTTTTAGAGCTAGGCCAGGCCGCCGCGTAC
gRNA_M2S9_sgCXCR4.10_fwd	GAATTCCATCCACTTTAGCAAGGAGTTTTAGAGCTAGGCCAGGCCGCCGCGTAC
sgASCL1-2_2.0_fwd	GCCCGCGCTTCCCACCTGTCTTCGTTTTAGAGCTAGGCCAGGCCGCCGCGTAC
	GCCTCTGGGAGGTCCTGTCCGGCTCGTTTTAGAGCTAGGCCAGGCCGCCGCGTAC
sgASCL1-1_M2S9_fwd	gcagccgctcgctgcagcagGTTTTAGAGCTAGGCCAACATGAGGATCACCCATGTCTGCAGG
sgASCL1-2_M2S9_fwd	ggctgggtgtcccattgaaaGTTTTAGAGCTAGGCCAGGCCGCCGCGTAC
sgASCL1-3_M2S9_fwd	gcagccgctcgctgcagcagGTTTTAGAGCTAGGCCAGGCCGCCGCGTAC
sgASCL1-4_M2S9_fwd	gtgtttattcagccgggagtcGTTTTAGAGCTAGGCCAGGCCGCCGCGTAC
sgASCL1-5_M2S9_fwd	gatggagagtttgcaaggagcGTTTTAGAGCTAGGCCAGGCCGCCGCGTAC
sgASCL1-6_M2S9_fwd	gcaggagggacaaatggctctGTTTTAGAGCTAGGCCAGGCCGCCGCGTAC
PB-UniSAM-Fwd	gagtgaggagagagagaaaacGTTTTAGAGCTAGGCCAGGCCGCCGCGTAC
PB-UniSAM-Rev	TAGTGATAAAAGCGGCCGCAATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAACGCCAGCAACG
NPHPTM-fwd	CATTCCGCTTCCAGGTCCAGGGTTCTCCTCCACG
NPHPTM-rev	GAACCCTGGACCTGGAAGCGGAATGcctaagaaaaagaggaaggtggcg

eBFP_sgRNA_in_U6_fwd	TTTATTGCGGCCGCTTTTATCACTACTTGTACAGCTCG TCCATGCC
eBFP_sgRNA_in_U6_rev	TATATCTTGTGGAAAGGACGAAACAGGGAAAGGTCGA GAAACTGCAAA
PB-UniSAM_U6_fwd	GTTGGCTAGCTTAATTAAAGAATTCAAAAAGCACCGAC TCGGTGC
PB-UniSAM_U6_rev	GAATTCTTTAATTAAGCTAGCCAACAAGCTCGTCATCG CTTTGCAGAAGAGCAGAGAGG
PAL_L192K_mut_fwd	TGTTTCGTCCTTTCCACAAGATATATAAAGCCAAGAAA
PAL_L192K_mut_rev	TCGAAATACTTTCAAGTTACGGTAAGCATATGATAGTC C
	TTGCGGCGGCGGCTGTCCACC
	TGGCAACGCGGAGTTGGCCGCTGC

XI.2.2 Sequencing Primer

Table 31: Sequencing primers

Name	DNA sequence (5’ -> 3’)
CMV-Fwd	CGCAAATGGGCGGTAGGCGTG
pEGFP_C2-RP	TTTAAAGCAAGTAAAACCTC
dCas9_rev_seq	TCACTGGGCCAGGATTCTC
NPHPTM_seq_1_fwd	CATTACATCGGCTACCAGC
NPHPTM_seq_2_fwd	GTCCGGAGATGAAGACTTCTC
NPHPTM_seq_1_rev	GAGGCACCATAGCACTAGAG
NPHPTM_seq_2_rev	GTTCACCTTacccgaaccac
hU6_seq_fwd	CTTTTGCTCAGCTAGCGAGG
SV40-pA-rev	CCTCTACAAATGTGGTATGG

pGL3-RV3	CTAGCAAAATAGGCTGTCC
KAN2-RP	GCAATGTAACATCAGAGATTTTGAG
pENTattL2-rev	ACATCAGAGATTTTGAGACACG
pEGFP_C2-RP	TTTAAAGCAAGTAAAACCTC

XI.2.3 RNA sequences

Biotinylated RNA for mCherryPAL pulldown assay

Table 32: RNA sequences

Name	RNA sequence (5’ -> 3’)
04.17	GGUUGAAGCAGACGACC
04.17 M2 53.19	GGUUGAUGCAGACGACC
53.19 M27	CGGUACAGCAGCGAUGCCG
	CGGUACUCCUCCGAUGCCG

XI.3 CXCR4 promoter sequence and sgRNAs

CXCR4

>gi|568815596|ref|NC_000002.12|:136117550-136124642 Homo sapiens chromosome 2, GRCh38.p13 Primary Assembly
AGGCGCTTAACCCCGAAGTTAAAGCGGGCGCAATCTCCTCCTGGGAACTCAGCCCAGGCACGCCGCCCTC
CGCCTCTAAATTCAGACAATGTAACTCGCTCCAAGACATCCCCGCTTCCCCAAGGAAGAGACCGGTGGTC
TGAGTCCCGAGGCAGCGCGCACGCCTTCTCTGCACTTGTGCACAGAATGTTCTTACGTTTGCAAACAGCG
TGCAAGCCGCCGCGCGCGGCGGGACTCAAGGGGGAGACACATGCAGCCACTGGAACGCTCTTTCCAGTCG
TTTCTCCTCGACTCACAGAGAAAAAGATTCCAATCCTGCTCCCCCCCCACCCACCCGCACTATATAGGCA
TGGTCAAGAAAACTCCTTTCGGTGACCCTTTTTTGGAGTACGGGTACCTCCAATGTCCTGGCCGCTTCTG
CCCGCTCGGAGAGGGGCTGCGCTCTAAGTTCAAACGTTTGTACATTTATGACAAAGCAGGTTGAAACTGG
ACTTACACTGATCCCCTCCATGGTAACCGCTGGTTCTCCAGATGCGGTGGCTACTGGAGCACTCAGGCCC
TCGGCGTCACTTTGCTACCTGCTGCCGCAGCCAACAAACTGAAGTTTCTGGCCGCGGCCGGACTTTTATA
AAAACACGCTCCGAGCGCGGCGCATGCGCCGC[4]TGGGGCGGGGAGGGAGAAGGC[1]GGGGTGGGGGAGAGGG
G
CAGACGCGAGGAAGGAGGGCGCAGGGGCG[3]GGGCGGAGAGGTGCGCGGCTT[5]GGGAAGCCCAGGGGACCCT
G
CTGTTTGCGGGTGGTCGGTAGTGAGTC[4]CGGAGCCGGACAGGACCTCCCAG[3]AGGCATTTCCTAAGTTTGA
G
GGAAGCGGGATGCGCCTGAAGACAGGTGGGAAGCGC[2]GGGCTAGGAACGCGTCTCTCTG[1]TGGCGCGTCCA
T

CCTTGCTAAAGTGGATGGAATTCCAGACCTGGGAATGCTACAGTAGAGACACTGAGGCCCAGGGAAGGGG
AGTGACTTGCTGGAGTAACCCAACCAGTAGGCGGCAATGCTGGAAGTAGGGCCAAGCTTCCAACGCCATT
CCCCGCGCTTTCAGCAGACGCAGCCCATTCAGGAGGTAAAAGGAGTAAAGATCCTTCCTTCGGAGGATGT
AGCTATACTTGCACATTCACAGCATCACTAGGGTCAGGTGCAGATAATTAGGAGATACTTTCAGAAGAAG
GTAGGATTTGGCATAATTCTGGAGACTGATCAGATGTAGATATGTATTGCTCAGTAATTTCATAATAGCG
GTAACCAATTCGCGAATAGTGCTTTATTATCCTAAAAGTATGCTTTATTATAAGAGTAATTTAGCGACCA
GTACTTTTTGTAACAGATCACTAGGAACTTGCACAGAATCACTCATTCATTCTCACAACACCGTGTGGGT
ATTACCATCACAGAGATGCATTTAGTCCGATATTTATAGAAGGACGAACTGCGTGCCACGCACTGTTCTA
GGCGTTGGGGATCCAGCGTGAACCAGACAAGGCCCGCATTCGGAGAAACCAGCAAATAAGCCCGGAGAGA
TGAAATGCCAGGGATGGGCAGGGATTCCAAGCCGGGCTTGAACCCAGGTCTTCGGTCCCCATACTCGGCG
TCTTCCACGATTTTGAGCAGAGGACAAGCTTTGGCACGAATCAGAAATTAGGTTGAAGAACTGGGCGGGG
CGGAGTGGGGGTGGGGGTCAGACGTTTAAAGCGATCTCAGACTGGCTGAGGTTTGCAGTGGGCGCGCAGA
CAGTGCTTTCCCCCTCGCGCCACTTCCCTCCGGCGCGGAGACTGGGTGGGGCGGGGGAGCGAGTGTGGGG
CGGGGCCGAGAAACTGCGTCCGCAGGCGGGGTGGCCGCTGCGGAGATAGCTGGGACCGAGGCGCGTGCGT
CGCGACACGGACCCCAGAGGCCCCCACCCCGAAATGGTCCCGAGGAGAGGGATTGCCCAGACTGCACTCT
ACACACGATCCCTCTTTTCTATATCAAAAGAGGAAACATTTCAGTCTGGAGTGTTCCAGAGGTAGTTCGG
AGGCAGGGAACCTTTTCTCTACAGAGCATGACTTTCCCCCTTTGTCTGGGACCCCTCCCTAATGTTGGTG
GTTAAAGGTTCTCATCTACTCTGCACTTTCACTGACACAGGAAAGATGATATCCGGGCCGGGAGATGGGA
CTCTCAGAAACGAACGAGTTAGGACACACACTAATGGCCAGATAAAATCAATTCTAATATTACATTTGTA
GAATTAAAAAAAAAAAAAGAAACTCCAGGTTCTTGGAGGGTGGTGGCTGGAAGAGGTTTCGGAAGCTCCA
CGCCCCTTCTCACAGGCCTTGCATTACGTATCTCTTCATCTGTATCCTTTATAATATCCTTTATAATAAA
CGATAACGGTAAGAAATAAAAATGAAAATAAAACCGTAAAAACACTTTTACTTTTTTCGAATCTTTAATA
AAACTTAAGGATCTTCATATTTTTATAGACACGGTAGATCCTATCTACTGAAGGTCTAGAATTTCAGAGC
GTTTTCATGACTTGCTTTTCACAAAGCCTGGTAGTGGCAGAATCAGTATGTTTTATCCTGGACCGTAAGC
TGCACCCTACCTCAAAACTCCTCCCTGCACGATGTCTAAAACCTGAAAGGTGAGATCTCCCAAATCTGTC
CAAAAAGGGAAAGAGGAGAAAACTCGGAGCAAAATGGTTTGCATAATAAATATTTGTTGATTGTAGACAG
ACAAAAGAAAGCTCCTTCCTTTATCTGAAGACTCCTTTCAGATCTTTTTAGGGATCTCTACTAAAAGGAA
GCAGATAATTAAAAGAACAATGAGATGCTATTTGGTGACTACCAAATTAGGCAAAAATCCATTTTTTTTC
AAAGATAATATTTGGTCTCCGGATTTGAAAAGACCAGCTGTCTCTCGCTTGTCACTGATGGGAATGTGGC
TTGGGACAGCCTCACTGGAAAGCATTTTCTATTCAGCAGGAACTTCAGAACAGTAACTCCACTCATAAAA
ACTTATGCCACAAAGGTATTTGTGGAAACATTTTTTTTTAAATAGTGAAATGTCTGAACATCCCTAAATA
ACCACTGTTAGTGGAATAGTTGACAAGATTTTGGTTCATCCACATCTGAAATTCCATATGGACATCGCGA
ATGTTATTTGTGAAGAATTTTTAGTGACATGGGGAAATACTTATGAAACATTAAGTGCAAAAGGCAGGAT
TCAGAATTGTGTATATAGTGTGATCTCAGTTATGTTAAAATATGCATTTGGAAAATATAGGGAGGAAAAC
CAAAATGTTTATAGGCGTTTTCCTTGGTGCAGAATTATATTTGATTTTTGCTTTCTTCCTGGGGTTTTTC
TATGCTTTCCAGATTTTCTGTAATAGACATGAATCATTTTCATATTCATAAAAAAGCCAACAAATTGTAT
GACAAAAAGGGAATCAGAAGGCCTCCTTCATGTATGTTTCCACTTCTAAAGTATCTACTAAGATGTCAGA
CTTTTTTCACAATAGTCCTTTCCCCTTTCAAGGAAGGCTTGTTTTCCAATGGGCTAATCTTAAACTTCTA
TTTCCACTGGGCTACCTAAACCTTATTCTCAACTATTCTATGATGTTCTTTAAAAAACAAACAAAACTCA
AGCATTTACTGAGCGCTTGTTATGTGCAACTGACCTCATATAATCCCCTCAATAACTCTACAAGGCAGGT
GCCATTACTTTTATTACATGGTAAAGAAATTAGGGTCTGGAAAGATGAGGAACTGGCCTGTGACCCCGGA
ATGAGGATCTGCACCCAGTCTGTCTGGCTCTGAAGCCCAGGTCATTTGGCATTTACGTACATACTGCTTT
GAGTGGCATCCCTTGTGATGTTACTCAACTTTTAAATGAGTATATGTCATACTCCTATCCCTCCAATATT
TTAGGTTCCTTAAAGTGAATGCCCAGATATCCTCTATCCAGCTAGCCCATTGCTGAACAAATACTGGGCT
CCTTAATCTTTTACTGAACCTTTCTTAAAAGAATGAATGACTTTCCTGTGTGGCTCTGGCAACAGAGTCA
GGCTTAGGGGCTGTCAGGGATTTGGCAGGGTCCCCAACTTCTCAGGTTGGCTTTAGCCCAGGGAATAACT
AGCCTTGGCCACAGTTTGGAGGGGATGCTTACAAATCCAGTGACACAGTGGCAAGCATCCTGCCACCTCC
TGTTGCTCGCTCTGCAGCTGGTTGATTTGTGTGTGAGTGTGTGTATCTGTGTGTGTGTGATGAAAAGCTG
TACTCTGTCCCTAAATTGGCTTCATGTTAGCTAGGCAGAAAGCTTAGGAACAGACCTTTTCCCAGTGGTG
GATGGCAGGAGGTTAGAGAAAGGAGGAGGTGCCAGGAGGAGGTGGATGGCAGGAGGCATTAAAGGCCTTC
AGAATACACAGGGTCTTCTGGTTCCTTCGCTGGCCCCTCTCCTGTCTCTAAAAAAATTGATCGCTTTTGT
TTCTGAAAGAGGAAAAAAGTGCAAACTTCCTTGACAGTAGTAAGGGATACTTAGAGAAAGTTAATGTTTT
CCTAAGTGATTAAAAAAAAAAAACAAAAACGGAGACTATTTCTCAAGGCGCAAGTAATTCTAACACAGGG
TTACAAATCTCAAAGAGACAATGGTATCAAATGTCAACTATTTCCAAGCTGGTGTTAGGTGGAACAAATT
GTCCCTTGCCTAACCACACCCCGCCCTCTCCTCCTTCCCTAAATCCTTCTGTCTCCAACACCTACAGATG
AAAAGTGACTTTGAAAACAAACCCAAACCTTCCATCCTACAGTTCTTTACCAATTTAGGAAATTATTACT
GTTTTTTGGGGGAGAGCCATTTTTACTCTAGATTTAAGTTGTATCTTGAATGTTGAAAAGGTCTGCCTTC
ACTAAAATGGGATGATTAGTGATTTCCATCAATGCTTTATAGCAGTCAAGACTAGTCAACAAATTTCCTA
TACAAAGAGAATAAACTAGACTAGTAAACAAATATAGGCAATTTAGAGCCATCCTGGTAAGTCATTTAAA
CAAAACAGCCACTTAATTAACAGTGCAATGATCATTACTCTACTAGTTCCACTGCCAGGAGGATGACTGT
TCACAAAGTTTGCTGTGCTCCTGTGGCAGGTGCTTGTGTGGCTTTATGGGTGTAAAGAACAGCAGCAAAC
AGGCTCGGTGCTATTGCCTCAAAAATCATGGTGTCAAACTGGGAAGGAGGGTACTGCTTTAGGAAATAAT

TGAGTGAGTGAACCCTGTCCTCCATGGAATAGGCTTGTCATTGTTTTCCTTAAACCAGTAATTTATTGGT
GGTGACCTCAGACAGCTATATAAAAAGAGGAAGCGCCTAGAGCCCTGTGGACTGTCAAGTTTTATAAAGA
GAACTGGGCATCCCATGGCAGAGCCTTGGTGCTCCCAGAAATGTGGTGATCTGATCGCCCATACAGATGG
GCTTTGAGATATTTCTGAGGTCATTCTTCCAAGAATATTATATTAGCAAAAGAACAGGGTGGGAATTACT
TGTGTGTGTGTGTGTGCGTGTGTGTGTATGTGTGTGTGTAGGGGAATCCATACAATGTTTTGCTTGAGTT
TTCTAAATCTAAAGCTATCTGTCAGCTATCAAAGCTGTGTCTTTGAGATTTGTAACCTTCTGTTAGAATT
ACATGCCCCTTGAGAAATAGTCTCCTTAAAAAAAAATAACTTGGGAATGATCTTAGAGCTACTTGTAGGT
CTAAGACACCAGTGACAACATCTAAAAACCTGCAGGTATTGTAAATGAGTGAGGCAATAGGGACTGGTGG
GGACTGTGGAGAACCAAAGGGTGCATGCCCTGTCCTAAAGCAGCAACCACTACCCTCCATTGCCCTTTGA
GGGAATTTGGGCCATATTATTCACTTTTTTTGTTTTTTTTTTTTTTTGAGACGGAGTCTTGCTCTGTTGC
CCAGGCTGGAGTGCAGTGGCACAATCTCGGCTCATTGCAAGCTCCGCCTCCTGGGTCCACCCCATTCTCC
TGCCTCAGCCTCCTGTGTAGCTGGGACTACAGGCACCCACCACCACGCCCAGCTAATTCTTTTTGTATTT
TTAGTAGAGACGGGGTTTCACTGTGTTAGCCAGGATGGTCTCGATCTCCTGACCTTGTAATCCGCCCATC
TCGACCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACTGCGCCCAGCCTATTCAACTTTTTTTAAAGA
GAAAGTGAAAGTCCTGATTTTTTAATAAGAAATCTCTCAACTTAAAAATGTTGGTTCAAAACAAACAAAC
ACAAAAACTCTACTAGCTTCCAGTCTTCCAGTTTACACTGACCTAAGGTGTTGATGATAGCCATCAAAAG
AAAGTATGAGACTAGGTATTGTGGTTCACACCTGTAATCCCAGCACTTTGGGAGGCTGAGGAAGGCCAAT
CAGATCACTTGAGCTCAGGAGTTTGAGACCAGCATGGGCAACATGGTGAAACCCTGTCTCTACCAAAAAA
GTACAAAAAAATTAGCCAAGTGTGGTGGCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCAAGGAGATCAC
TTGAGGTTGGGAGGCAGAGGCTGCAGTGAGTCGAGATCGCATCTCTGCACTGCAGCCTGCATGCCAGAGT
GTCACTCTGTTTCAACAAAGAAAAAAAGAAAAAAAAGAAAGGATAGGGTTAGGCAGTCAGGAGACTTGGG
AGGATGGAACAGACACACATGGACAAAAACTGGCTGCAGAATCCTTTTGGGGTGGTCTCTATGTCATAAT
TTGTCTCTCTGAGACTGAAATTAGTTACTTGGGATTTCCAGGTCAGTCCTCCCTTGCGGCACCTAACCGC
ACTCAGGTTAATGTCAGGGACCCTGAAGAGAATTGACAAGCTTTGTGTGATCCTTACATTCCTCAACCCT
TTTCCCATCCCCAGGCCCTATGCCTTCTTCACTGTCTTGATGCATGAGCACTGGGCCTGTCTTTTCTTTC
AGTTATTATAGGACAACGTCCTTCAGGGAGAAGAACATTCTTCCTTCTGCTCTTTAAACAGTGGGAACAA
CAGCAGAAAGGAGGATTGCAAGC

Yellow: database request http://sam.genome-engineering.org

Blue: sgRNAs used in this study

Red: TSS

Supporting Figure 4: CXCR4 sgRNA binding sites

XI.4 Raw data of ASCL1 qPCR analysis

XI.4.1 Calibration

Table 33: qPCR calibration raw data N=1

Target	Type	Sample	Cq
ASCL1	Unkn	L_UT1	
ASCL1	Unkn	L_UT2	
ASCL1	Unkn	L_MS2_1	33,47997473
ASCL1	Unkn	L_MS2_2	33,92553213
ASCL1	Unkn	4	35,5110908
ASCL1	Unkn	8	
ASCL1	Unkn	16	39,68645702

Target	Type	Sample	Cq
GAPDH	Unkn	L_UT1	19,29174992
GAPDH	Unkn	L_UT2	20,31549786
GAPDH	Unkn	L_MS2_1	18,5405237
GAPDH	Unkn	L_MS2_2	18,99906018
GAPDH	Unkn	4	21,48477016
GAPDH	Unkn	8	21,71238646
GAPDH	Unkn	16	22,71315414

ASCL1	Unkn	32		GAPDH	Unkn	32	23,5216074
ASCL1	Unkn	64	38,87617791	GAPDH	Unkn	64	24,55115004
ASCL1	Unkn	128		GAPDH	Unkn	128	25,43435543
ASCL1	Unkn	D_UT1		GAPDH	Unkn	D_UT1	19,19849862
ASCL1	Unkn	D_UT2	38,49202611	GAPDH	Unkn	D_UT2	18,72656329
ASCL1	Unkn	D_MS2_1	34,27303949	GAPDH	Unkn	D_MS2_1	19,13497005
ASCL1	Unkn	D_MS2_2	34,52301263	GAPDH	Unkn	D_MS2_2	19,1537277
ASCL1	Unkn	4	38,45087173	GAPDH	Unkn	4	21,71271607
ASCL1	Unkn	8	37,21794062	GAPDH	Unkn	8	21,86492152
ASCL1	Unkn	16		GAPDH	Unkn	16	23,00239899
ASCL1	Unkn	32	39,44788751	GAPDH	Unkn	32	23,33293104
ASCL1	Unkn	64	37,38152987	GAPDH	Unkn	64	24,63816313
ASCL1	Unkn	128		GAPDH	Unkn	128	25,85874782
ASCL1	NRT			GAPDH	NRT		
ASCL1	NTC			GAPDH	NTC		
ASCL1	NRT			GAPDH	NRT		
ASCL1	NTC			GAPDH	NTC		

Table 34: qPCR calibration raw data N=2

Target	Type	Sample	Cq	Target	Type	Sample	Cq
ASCL1	Unkn	L_UT1		GAPDH	Unkn	L_UT1	17,99905249
ASCL1	Unkn	L_UT2	38,96892315	GAPDH	Unkn	L_UT2	18,070772
ASCL1	Unkn	L_MS2_1	31,39742707	GAPDH	Unkn	L_MS2_1	17,8839814
ASCL1	Unkn	L_MS2_2	31,291108	GAPDH	Unkn	L_MS2_2	17,61234144
ASCL1	Unkn	4	34,20619488	GAPDH	Unkn	4	20,08026573
ASCL1	Unkn	8	35,22066791	GAPDH	Unkn	8	20,71967759
ASCL1	Unkn	16	35,29650119	GAPDH	Unkn	16	21,49155725
ASCL1	Unkn	32	36,87564155	GAPDH	Unkn	32	22,42274434
ASCL1	Unkn	64	38,02918668	GAPDH	Unkn	64	23,51187419
ASCL1	Unkn	128	37,53985984	GAPDH	Unkn	128	24,57126226
ASCL1	Unkn	D_UT1		GAPDH	Unkn	D_UT1	18,131079
ASCL1	Unkn	D_UT2		GAPDH	Unkn	D_UT2	17,98726636
ASCL1	Unkn	D_MS2_1	31,98235251	GAPDH	Unkn	D_MS2_1	18,00936911
ASCL1	Unkn	D_MS2_2	31,95610672	GAPDH	Unkn	D_MS2_2	17,90275762
ASCL1	Unkn	4	34,29833629	GAPDH	Unkn	4	19,50312521
ASCL1	Unkn	8	35,04703798	GAPDH	Unkn	8	20,38978174
ASCL1	Unkn	16	35,90683743	GAPDH	Unkn	16	21,36869121
ASCL1	Unkn	32	38,7924068	GAPDH	Unkn	32	22,93826882
ASCL1	Unkn	64	37,79047977	GAPDH	Unkn	64	23,47151307
ASCL1	Unkn	128		GAPDH	Unkn	128	24,43505097
ASCL1	NRT			GAPDH	NRT		35,30296971
ASCL1	NTC			GAPDH	NTC		
ASCL1	NRT			GAPDH	NRT		36,26211532
ASCL1	NTC			GAPDH	NTC		

XI.4.2 Position screening

Table 35: qPCR position screening raw data N=1

Target	Type	Sample	Cq	Target	Type	Sample	Cq
ASCL1	Unkn	L_pos1	36,40941676	GAPDH	Unkn	L_pos1	18,41857023
ASCL1	Unkn	L_pos1		GAPDH	Unkn	L_pos1	18,82251707
ASCL1	Unkn	D_pos1	38,75011484	GAPDH	Unkn	D_pos1	19,20794846
ASCL1	Unkn	D_pos1	38,17887885	GAPDH	Unkn	D_pos1	19,20994746
ASCL1	Unkn	4		GAPDH	Unkn	4	20,47677873
ASCL1	Unkn	8		GAPDH	Unkn	8	21,39049773
ASCL1	Unkn	16	39,3822463	GAPDH	Unkn	16	22,90168619
ASCL1	Unkn	32		GAPDH	Unkn	32	24,07042679
ASCL1	Unkn	64		GAPDH	Unkn	64	25,26518592
ASCL1	Unkn	128		GAPDH	Unkn	128	26,91789761
ASCL1	Unkn	L_pos2	35,15392315	GAPDH	Unkn	L_pos2	19,08605255
ASCL1	Unkn	L_pos2	35,69815359	GAPDH	Unkn	L_pos2	18,94054837
ASCL1	Unkn	D_pos2	37,75782192	GAPDH	Unkn	D_pos2	18,88626742
ASCL1	Unkn	D_pos2	38,6294618	GAPDH	Unkn	D_pos2	18,41682765
ASCL1	Unkn	4		GAPDH	Unkn	4	20,54391703
ASCL1	Unkn	8		GAPDH	Unkn	8	21,60142075
ASCL1	Unkn	16	39,11983657	GAPDH	Unkn	16	22,57373106
ASCL1	Unkn	32		GAPDH	Unkn	32	24,10426914
ASCL1	Unkn	64		GAPDH	Unkn	64	25,09690318
ASCL1	Unkn	128		GAPDH	Unkn	128	26,75182942
ASCL1	Unkn	L_pos3	35,8629183	GAPDH	Unkn	L_pos3	18,83262454
ASCL1	Unkn	L_pos3	36,10393088	GAPDH	Unkn	L_pos3	18,5631472
ASCL1	Unkn	D_pos3	38,09983554	GAPDH	Unkn	D_pos3	18,92383391
ASCL1	Unkn	D_pos3		GAPDH	Unkn	D_pos3	17,88969897
ASCL1	NRT			GAPDH	NRT		
ASCL1	NTC			GAPDH	NTC		
ASCL1	Unkn	L_pos4	35,20058844	GAPDH	Unkn	L_pos4	18,66594855
ASCL1	Unkn	L_pos4	34,2342821	GAPDH	Unkn	L_pos4	18,62980565
ASCL1	Unkn	D_pos4	38,49815341	GAPDH	Unkn	D_pos4	19,51238562
ASCL1	Unkn	D_pos4		GAPDH	Unkn	D_pos4	18,78583033
ASCL1	NRT			GAPDH	NRT		
ASCL1	NTC			GAPDH	NTC		
ASCL1	Unkn	L_pos5	35,84229679	GAPDH	Unkn	L_pos5	18,58635566
ASCL1	Unkn	L_pos5	35,57828382	GAPDH	Unkn	L_pos5	18,35196391
ASCL1	Unkn	D_pos5	38,2541441	GAPDH	Unkn	D_pos5	18,43708099
ASCL1	Unkn	D_pos5	37,51816345	GAPDH	Unkn	D_pos5	18,54761413
ASCL1	Unkn	L_pos6	36,18746641	GAPDH	Unkn	L_pos6	18,17328861
ASCL1	Unkn	L_pos6	36,0096046	GAPDH	Unkn	L_pos6	18,16187569
ASCL1	Unkn	D_pos6		GAPDH	Unkn	D_pos6	18,44139144
ASCL1	Unkn	D_pos6	38,33600256	GAPDH	Unkn	D_pos6	18,46809302
ASCL1	Unkn	L_nat	38,37598116	GAPDH	Unkn	L_nat	18,54529622
ASCL1	Unkn	L_nat		GAPDH	Unkn	L_nat	18,46250345
ASCL1	Unkn	D_nat	36,87071812	GAPDH	Unkn	D_nat	18,30442347
ASCL1	Unkn	D_nat	38,50831243	GAPDH	Unkn	D_nat	18,400045

Table 36: qPCR position screening raw data N=2

Target	Type	Sample	Cq	Target	Type	Sample	Cq
ASCL1	Unkn	L_pos1	38,2484717	GAPDH	Unkn	L_pos1	18,77412539
ASCL1	Unkn	L_pos1		GAPDH	Unkn	L_pos1	18,75006619

ASCL1	Unkn	D_pos1		GAPDH	Unkn	D_pos1	18,59899096
ASCL1	Unkn	D_pos1	37,70824349	GAPDH	Unkn	D_pos1	18,58958107
ASCL1	Unkn	4	37,41284583	GAPDH	Unkn	4	20,76250228
ASCL1	Unkn	8	38,34177918	GAPDH	Unkn	8	21,50208362
ASCL1	Unkn	16		GAPDH	Unkn	16	22,28369072
ASCL1	Unkn	32		GAPDH	Unkn	32	23,76427777
ASCL1	Unkn	64		GAPDH	Unkn	64	24,68395914
ASCL1	Unkn	128		GAPDH	Unkn	128	26,11538508
ASCL1	Unkn	L_pos2	34,70986329	GAPDH	Unkn	L_pos2	19,09864569
ASCL1	Unkn	L_pos2	34,39519609	GAPDH	Unkn	L_pos2	18,12548456
ASCL1	Unkn	D_pos2		GAPDH	Unkn	D_pos2	18,92150879
ASCL1	Unkn	D_pos2		GAPDH	Unkn	D_pos2	18,66088348
ASCL1	Unkn	4	37,00721046	GAPDH	Unkn	4	20,59857077
ASCL1	Unkn	8	37,2831842	GAPDH	Unkn	8	21,77896106
ASCL1	Unkn	16	38,49509056	GAPDH	Unkn	16	22,97227523
ASCL1	Unkn	32		GAPDH	Unkn	32	23,40543932
ASCL1	Unkn	64		GAPDH	Unkn	64	24,87409852
ASCL1	Unkn	128		GAPDH	Unkn	128	25,85222978
ASCL1	Unkn	L_pos3	35,69540448	GAPDH	Unkn	L_pos3	18,5058198
ASCL1	Unkn	L_pos3	35,60161827	GAPDH	Unkn	L_pos3	18,58442123
ASCL1	Unkn	D_pos3		GAPDH	Unkn	D_pos3	18,77779944
ASCL1	Unkn	D_pos3		GAPDH	Unkn	D_pos3	18,72096985
ASCL1	NRT			GAPDH	NRT		
ASCL1	NTC			GAPDH	NTC		
ASCL1	Unkn	L_pos4	34,63994136	GAPDH	Unkn	L_pos4	18,35808235
ASCL1	Unkn	L_pos4	34,82794349	GAPDH	Unkn	L_pos4	19,01512001
ASCL1	Unkn	D_pos4		GAPDH	Unkn	D_pos4	18,76412692
ASCL1	Unkn	D_pos4		GAPDH	Unkn	D_pos4	18,43211596
ASCL1	NRT			GAPDH	NRT		
ASCL1	NTC			GAPDH	NTC		36,40485279
ASCL1	Unkn	L_pos5	37,3919958	GAPDH	Unkn	L_pos5	19,22379584
ASCL1	Unkn	L_pos5	37,37316338	GAPDH	Unkn	L_pos5	18,87710195
ASCL1	Unkn	D_pos5		GAPDH	Unkn	D_pos5	18,74228197
ASCL1	Unkn	D_pos5	38,23203862	GAPDH	Unkn	D_pos5	18,93845638
ASCL1	Unkn	L_pos6	37,67334291	GAPDH	Unkn	L_pos6	18,26357216
ASCL1	Unkn	L_pos6	36,4011147	GAPDH	Unkn	L_pos6	19,1173576
ASCL1	Unkn	D_pos6		GAPDH	Unkn	D_pos6	18,89763841
ASCL1	Unkn	D_pos6	38,07412631	GAPDH	Unkn	D_pos6	18,86118513
ASCL1	Unkn	L_nat	38,94355929	GAPDH	Unkn	L_nat	18,54901841
ASCL1	Unkn	L_nat	37,61162075	GAPDH	Unkn	L_nat	17,68774596
ASCL1	Unkn	D_nat	38,18594791	GAPDH	Unkn	D_nat	18,63729389
ASCL1	Unkn	D_nat		GAPDH	Unkn	D_nat	18,88908281

XI.4.3 sgRNA #2 and sgRNA #4

Table 37: qPCR ASCL1 activation N=1

Target	Type	Sample	Cq	Target	Type	Sample	Cq
ASCL1	Unkn	L_nat1		GAPDH	Unkn	L_nat1	18,79207784
ASCL1	Unkn	L_nat2		GAPDH	Unkn	L_nat2	19,01613445
ASCL1	Unkn	D_nat1		GAPDH	Unkn	D_nat1	19,05409289
ASCL1	Unkn	D_nat2		GAPDH	Unkn	D_nat2	18,62679802
ASCL1	Unkn	L_MS2_1	33,05290628	GAPDH	Unkn	L_MS2_1	18,58325153
ASCL1	Unkn	L_MS2_2	32,77182929	GAPDH	Unkn	L_MS2_2	18,48172842
ASCL1	Unkn	D_MS2_1	33,92116126	GAPDH	Unkn	D_MS2_1	18,96412537
ASCL1	Unkn	D_MS2_2	33,26094668	GAPDH	Unkn	D_MS2_2	18,72808157
ASCL1	Unkn	L_pos2_1	35,16329355	GAPDH	Unkn	L_pos2_1	18,45388039
ASCL1	Unkn	L_pos2_2	35,08059524	GAPDH	Unkn	L_pos2_2	18,5903271

ASCL1	Unkn	D_pos2_1	38,25615348	GAPDH	Unkn	D_pos2_1	18,99100822
ASCL1	Unkn	D_pos2_2		GAPDH	Unkn	D_pos2_2	
ASCL1	Unkn	L_pos4_1	35,66530212	GAPDH	Unkn	L_pos4_1	18,80636273
ASCL1	Unkn	L_pos4_2	36,21751621	GAPDH	Unkn	L_pos4_2	18,66877916
ASCL1	Unkn	D_pos4_1	38,33417724	GAPDH	Unkn	D_pos4_1	18,57889879
ASCL1	Unkn	D_pos4_2	38,47990766	GAPDH	Unkn	D_pos4_2	17,37645249
ASCL1	Unkn	L_pos24_1	30,30784034	GAPDH	Unkn	L_pos24_1	18,93479652
ASCL1	Unkn	L_pos24_2	30,09102474	GAPDH	Unkn	L_pos24_2	18,30362187
ASCL1	Unkn	D_pos24_1		GAPDH	Unkn	D_pos24_1	19,05724183
ASCL1	Unkn	D_pos24_2	37,63989058	GAPDH	Unkn	D_pos24_2	18,41751081
ASCL1	NRT			GAPDH	NRT		
ASCL1	NRT			GAPDH	NRT		
ASCL1	NTC			GAPDH	NTC		
ASCL1	NTC			GAPDH	NTC		

Table 38: qPCR ASCL1 activation N=2

Target	Type	Sample	Cq	Target	Type	Sample	Cq
ASCL1	Unkn	L_nat1	38,33186723	GAPDH	Unkn	L_nat1	18,66833388
ASCL1	Unkn	L_nat2	36,92081283	GAPDH	Unkn	L_nat2	18,53708375
ASCL1	Unkn	D_nat1		GAPDH	Unkn	D_nat1	19,2825752
ASCL1	Unkn	D_nat2		GAPDH	Unkn	D_nat2	18,40271711
ASCL1	Unkn	L_MS2_1	32,29008457	GAPDH	Unkn	L_MS2_1	18,51572421
ASCL1	Unkn	L_MS2_2	32,44913002	GAPDH	Unkn	L_MS2_2	18,38559096
ASCL1	Unkn	D_MS2_1	32,7073168	GAPDH	Unkn	D_MS2_1	18,6606889
ASCL1	Unkn	D_MS2_2	32,09783746	GAPDH	Unkn	D_MS2_2	18,40165599
ASCL1	Unkn	L_pos2_1	37,20511387	GAPDH	Unkn	L_pos2_1	19,04154791
ASCL1	Unkn	L_pos2_2	35,49650212	GAPDH	Unkn	L_pos2_2	18,65788977
ASCL1	Unkn	D_pos2_1		GAPDH	Unkn	D_pos2_1	19,12475303
ASCL1	Unkn	D_pos2_2	38,11380035	GAPDH	Unkn	D_pos2_2	18,69301486
ASCL1	Unkn	L_pos4_1	35,03552215	GAPDH	Unkn	L_pos4_1	18,61513254
ASCL1	Unkn	L_pos4_2	35,32371567	GAPDH	Unkn	L_pos4_2	18,45786234
ASCL1	Unkn	D_pos4_1	36,45825483	GAPDH	Unkn	D_pos4_1	18,96524858
ASCL1	Unkn	D_pos4_2	38,6409214	GAPDH	Unkn	D_pos4_2	18,56295561
ASCL1	Unkn	L_pos24_1	30,81789652	GAPDH	Unkn	L_pos24_1	19,13621869
ASCL1	Unkn	L_pos24_2	30,38661069	GAPDH	Unkn	L_pos24_2	18,49392079
ASCL1	Unkn	D_pos24_1		GAPDH	Unkn	D_pos24_1	19,27446293
ASCL1	Unkn	D_pos24_2	37,91314245	GAPDH	Unkn	D_pos24_2	18,34632594
ASCL1	NRT			GAPDH	NRT		
ASCL1	NRT			GAPDH	NRT		
ASCL1	NTC			GAPDH	NTC		
ASCL1	NTC			GAPDH	NTC		

Table 39: qPCR ASCL1 activation N=3

Target	Type	Sample	Cq	Target	Type	Sample	Cq
ASCL1	Unkn	L_nat1	38,69167848	GAPDH	Unkn	L_nat1	19,11768009
ASCL1	Unkn	L_nat2		GAPDH	Unkn	L_nat2	19,01317237
ASCL1	Unkn	D_nat1	38,47749373	GAPDH	Unkn	D_nat1	18,74633897
ASCL1	Unkn	D_nat2		GAPDH	Unkn	D_nat2	18,65195194
ASCL1	Unkn	L_MS2_1	32,13451577	GAPDH	Unkn	L_MS2_1	18,91788378

ASCL1	Unkn	L_MS2_2	32,03095471	GAPDH	Unkn	L_MS2_2	18,41399827
ASCL1	Unkn	D_MS2_1	32,58066762	GAPDH	Unkn	D_MS2_1	18,59070977
ASCL1	Unkn	D_MS2_2	32,18567271	GAPDH	Unkn	D_MS2_2	18,40282695
ASCL1	Unkn	L_pos2_1	34,99939989	GAPDH	Unkn	L_pos2_1	18,87588261
ASCL1	Unkn	L_pos2_2	35,13596505	GAPDH	Unkn	L_pos2_2	18,56112326
ASCL1	Unkn	D_pos2_1		GAPDH	Unkn	D_pos2_1	19,03058279
ASCL1	Unkn	D_pos2_2	39,09912347	GAPDH	Unkn	D_pos2_2	18,70922085
ASCL1	Unkn	L_pos4_1	35,0136585	GAPDH	Unkn	L_pos4_1	18,77174948
ASCL1	Unkn	L_pos4_2	36,51920616	GAPDH	Unkn	L_pos4_2	18,90131448
ASCL1	Unkn	D_pos4_1		GAPDH	Unkn	D_pos4_1	18,83370999
ASCL1	Unkn	D_pos4_2	38,89537741	GAPDH	Unkn	D_pos4_2	18,48094404
ASCL1	Unkn	L_pos24_1		GAPDH	Unkn	L_pos24_1	
ASCL1	Unkn	L_pos24_2	29,96962253	GAPDH	Unkn	L_pos24_2	18,58934317
ASCL1	Unkn	D_pos24_1	39,10079784	GAPDH	Unkn	D_pos24_1	19,29193973
ASCL1	Unkn	D_pos24_2	38,76999474	GAPDH	Unkn	D_pos24_2	18,67764262
ASCL1	NRT		30,54283142	GAPDH	NRT		19,31622451
ASCL1	NRT			GAPDH	NRT		
ASCL1	NTC			GAPDH	NTC		
ASCL1	NTC			GAPDH	NTC		

XI.5 Sequencing Data

```
SEQ_ID: 001
Name: pGL3-Basic-8xsgRNA-eBFP
Clone: 4
Service: GATC
Primer: pGL3-RV3
Direction: Forward


eGFP_ref        --------ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCG
eBFP_ref        --GCTAGCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCG
seq             AAGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCG
                        **************************************************

eGFP_ref        AGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATG
eBFP_ref        AGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGAGGGGCGAGGGCGAGGGCGATG
seq             AGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGAGGGGCGAGGGCGAGGGCGATG
                ************************************   ******************

eGFP_ref        CCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCT
eBFP_ref        CCACCAACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCT
seq             CCACCAACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCT
                ***** **************************************************

eGFP_ref        GGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACC
eBFP_ref        GGCCCACCCTCGTGACCACCCTGAGCCACGGCGTGCAGTGCTTCGCCCGCTACCCCGACC
seq             GGCCCACCCTCGTGACCACCCTGAGCCACGGCGTGCAGTGCTTCGCCCGCTACCCCGACC
                ********************** * ****************  **************

eGFP_ref        ACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA
eBFP_ref        ACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA
seq             ACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCA
                ************************************************************
```

```
eGFP_ref        CCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG
eBFP_ref        CCATCTTCTTCAAGGACGACGGCACCTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG
seq             CCATCTTCTTCAAGGACGACGGCACCTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCG
                ********************** **********************************

eGFP_ref        ACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCC
eBFP_ref        ACACCCTGGTGAACCGCATCGAGCTGAAGGGCGTCGACTTCAAGGAGGACGGCAACATCC
seq             ACACCCTGGTGAACCGCATCGAGCTGAAGGGCGTCGACTTCAAGGAGGACGGCAACATCC
                ****************************** *************************

eGFP_ref        TGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGC
eBFP_ref        TGGGGCACAAGCTGGAGTACAACTTCAACAGCCACAACATCTATATCATGGCCGTCAAGC
seq             TGGGGCACAAGCTGGAGTACAACTTCAACAGCCACAACATCTATATCATGGCCGTCAAGC
                ********************** ************* *************** *****

eGFP_ref        AGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGC
eBFP_ref        AGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACGTGGAGGACGGCAGCGTGC
seq             AGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACGTGGAGGACGGCAGCGTGC
                ************************************* * ****************

eGFP_ref        AGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
eBFP_ref        AGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
seq             AGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCG
                ************************************************************

eGFP_ref        ACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATC
eBFP_ref        ACAGCCACTACCTGAGCACCCAGTCCGTGCTGAGCAAAGACCCCAACGAGAAGCGCGATC
seq             ACAGCCACTACCTGAGCACCCAGTCCGTGCTGAGCAAAGACCCCAACGAGAAGCGCGATC
                *** **********************  *******************************

eGFP_ref        ACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGT
eBFP_ref        ACATGGTCCTGCTGGAGTTCCGCACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGT
seq             ACATGGTCCTGCTGGAGTTCCGCACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGT
                ********************   ************************************

eGFP_ref        ACAAGTAA----------------------
eBFP_ref        ACAAGTGA----------------------
seq             ACAAGTGAGCTAGCGTTTAAACTTAAGCTT
                ****** *
```

SEQ_ID: 002
Name: pMS2-p65-HSF1-T2A-eGFP
Clone: B2
Service: GATC
Primer: CMV-Fwd
Direction: Forward

```
ref               1 -----accggtcgccaccatgGCTTCAAACTTTACTCAGTTCGTGCTCGT     45
                       ||||||||||||||||||||||||||||||||||||||||||||||||
seq             101 GATCCACCGGTCGCCACCATGGCTTCAAACTTTACTCAGTTCGTGCTCGT    150

ref              46 GGACAATGGTGGGACAGGGGATGTGACAGTGGCTCCTTCTAATTTCGCTA     95
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq             151 GGACAATGGTGGGACAGGGGATGTGACAGTGGCTCCTTCTAATTTCGCTA    200

ref              96 ATGGGGTGGCAGAGTGGATCAGCTCCAACTCACGGAGCCAGGCCTACAAG    145
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq             201 ATGGGGTGGCAGAGTGGATCAGCTCCAACTCACGGAGCCAGGCCTACAAG    250

ref             146 GTGACATGCAGCGTCAGGCAGTCTAGTGCCCAGAAgAGAAAGTATACCAT    195
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq             251 GTGACATGCAGCGTCAGGCAGTCTAGTGCCCAGAAGAGAAAGTATACCAT    300

ref             196 CAAGGTGGAGGTCCCCAAAGTGGCTACCCAGACAGTGGGCGGAGTCGAAC    245
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq             301 CAAGGTGGAGGTCCCCAAAGTGGCTACCCAGACAGTGGGCGGAGTCGAAC    350

ref             246 TGCCTGTCGCCGCTTGGAGGTCCTACCTGAACATGGAGCTCACTATCCCA    295
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq             351 TGCCTGTCGCCGCTTGGAGGTCCTACCTGAACATGGAGCTCACTATCCCA    400

ref             296 ATTTTCGCTACCAATTCTGACTGTGAACTCATCGTGAAGGCAATGCAGGG    345
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq             401 ATTTTCGCTACCAATTCTGACTGTGAACTCATCGTGAAGGCAATGCAGGG    450
```

```
ref              346 GCTCCTCAAAGACGGTAATCCTATCCCTTCCGCCATCGCCGCTAACTCAG    395
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              451 GCTCCTCAAAGACGGTAATCCTATCCCTTCCGCCATCGCCGCTAACTCAG    500

ref              396 GTATCTACagcgctGGAGGAGGTGGAAGCGGAGGAGGAGGAAGCGGAGGA    445
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              501 GTATCTACAGCGCTGGAGGAGGTGGAAGCGGAGGAGGAGGAAGCGGAGGA    550

ref              446 GGAGGTAGCggacctaagaaaaagaggaaggtggcggccgctggatccCC    495
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              551 GGAGGTAGCGGACCTAAGAAAAAGAGGAAGGTGGCGGCCGCTGGATCCCC    600

ref              496 TTCAGGGCAGATCAGCAACCAGGCCCTGGCTCTGGCCCCTAGCTCCGCTC    545
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              601 TTCAGGGCAGATCAGCAACCAGGCCCTGGCTCTGGCCCCTAGCTCCGCTC    650

ref              546 CAGTGCTGGCCCAGACTATGGTGCCCTCTAGTGCTATGGTGCCTCTGGCC    595
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              651 CAGTGCTGGCCCAGACTATGGTGCCCTCTAGTGCTATGGTGCCTCTGGCC    700

ref              596 CAGCCACCTGCTCCAGCCCCTGTGCTGACCCCAGGACCACCCCAGTCACT    645
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              701 CAGCCACCTGCTCCAGCCCCTGTGCTGACCCCAGGACCACCCCAGTCACT    750

ref              646 GAGCGCTCCAGTGCCCAAGTCTACACAGGCCGGCGAGGGGACTCTGAGTG    695
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              751 GAGCGCTcCAGTGcCCaAGtCTACACAGGcCGGcgAGGGGAcTCTGAgTG    800

ref              696 AAGCTCTGCTGCACCTGCAGTTCGACGCTGATGAGGACCTGGGAGCTCTG    745
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              801 AAGCTCTGCTGCACCTGcAGTTCGACGCTGATGAgGacCTGGGAgCTCTG    850

ref              746 CTGGGGAACAGCACCGATCCCGGAGTGTTCACAGATCTGGCC-TCCGTGG    794
                     ||||||||||||||||||||||||.||||||||||||||||| .|||||||
seq              851 CTGGGgAAcAgcaCCgaTcCCGGanTGttCaCAgATcTGGccnncCGtGg    900

ref              795 ACAACTCTGAGTTTCAGCAGCTGCTGAATCAGGGCG-TGTCCATGTCTCA    843
                     ..|||||||||||||||||||||||||||||||||| |||||
seq              901 nnaAcTCTGAGTTTCaGCAGCTGCTGAATCaGGgcgntGtcC--------    942

SEQ_ID: 003
Name: pMS2-p65-HSF1-T2A-eGFP
Clone: B2
Service: GATC
Primer: pEGFP_C2-RP
Direction: Reverse


ref       2051 TGCTGACAGGCTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCCgct  2100
                         |||||||.|||||.
seq          1 ------------------------------------cCaCtgTnTCcGcn    14

ref       2101 agcGGCAGTGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGA  2150
               |||||||||||||||.|||||||||||||||||||||||||||||||||||
seq         15 AGCGGCAGTGGAGAGGnCAGAgGAAgTcTGCTAaCATGCGGtGACGTCGA    64

ref       2151 GGAGAAT-CCTGGCCCAGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTG  2199
               ||||..| |.|||||||||||||||||||||||||||||||||||||||||
seq         65 GGAgnntncntggCCCAGTGAGCAAGGGCGAGGAGctgTTCACCGGGGTG   114

ref       2200 GTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAG  2249
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq       115 GTGCCCATCCTGGTCGAGCtGGACGGCGACGTAAACGGCCACAAGTTCAG   164

ref       2250 CGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGA  2299
               |||||||||||||||||||||||||||||||||||.||.|||||||
seq       165 CGTgTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAgntGnCCCTGA   214

ref       2300 AGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTG  2349
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq       215 AGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTG   264
```

```
ref       2350 ACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACAT  2399
               ||||||||||||||||||||||||||||||||||||||||||||||||||
seq        265 ACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGCTACCCCGACCACAT   314

ref       2400 GAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGG  2449
               ||||||||||||||||||||||||||||||||||||||||||||||||||
seq        315 GAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGG   364

ref       2450 AGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG  2499
               ||||||||||||||||||||||||||||||||||||||||||||||||||
seq        365 AGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAG   414

ref       2500 GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCAT  2549
               ||||||||||||||||||||||||||||||||||||||||||||||||||
seq        415 GTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCAT   464

ref       2550 CGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACT  2599
               ||||||||||||||||||||||||||||||||||||||||||||||||||
seq        465 CGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACT   514

ref       2600 ACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATC  2649
               ||||||||||||||||||||||||||||||||||||||||||||||||||
seq        515 ACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATC   564

ref       2650 AAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCT  2699
               ||||||||||||||||||||||||||||||||||||||||||||||||||
seq        565 AAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCT   614

ref       2700 CGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGC  2749
               ||||||||||||||||||||||||||||||||||||||||||||||||||
seq        615 CGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGC   664

ref       2750 TGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCC  2799
               ||||||||||||||||||||||||||||||||||||||||||||||||||
seq        665 TGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCC   714

ref       2800 AACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGG  2849
               ||||||||||||||||||||||||||||||||||||||||||||||||||
seq        715 AACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTCGTGACCGCCGCCGG   764

ref       2850 GATCACTCTCGGCATGGACGAGCTGTACAAGtaaagcggccgcgactcta  2899
               ||||||||||||||||||||||||||||||||.||||||||||||||||||
seq        765 GATCACTCTCGGCATGGACGAGCtGTACAAGtnAAGCGGCCGCGACTCta   814

ref       2900 gatcataatcagccataccacatttgtagaggttttacttgctttaaaaa  2949
               ||||.|
seq        815 GatCnt----------------------------------------  820
```

```
SEQ_ID: 004
Name: pNLS-p65-HSF1-PAL-T2A-mCherry
Clone: 3
Service: GATC
Primer: CMV-Fwd
Direction: Forward
```

```
ref                1 --------cgccaccatgcctaagaaaaagaggaaggtggcggccgctgg     42
                         ||||||||||||||||||||.||||||||||||||||||||||||
seq              101 CCACCGGTCGCCACCATGCCTAAGAAAAGGAGGAAGGTGGCGGCCGCTGG    150

ref               43 atccCCTTCAGGGCAGATCAGCAACCAGGCCCTGGCTCTGGCCCCTAGCT     92
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              151 ATCCCCTTCAGGGCAGATCAGCAACCAGGCCCTGGCTCTGGCCCCTAGCT    200

ref               93 CCGCTCCAGTGCTGGCCCAGACTATGGTGCCCTCTAGTGCTATGGTGCCT    142
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              201 CCGCTCCAGTGCTGGCCCAGACTATGGTGCCCTCTAGTGCTATGGTGCCT    250

ref              143 CTGGCCCAGCCACCTGCTCCAGCCCCTGTGCTGACCCCAGGACCACCCCA    192
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              251 CTGGCCCAGCCACCTGCTCCAGCCCCTGTGCTGACCCCAGGACCACCCCA    300
```

```
ref              193 GTCACTGAGCGCTCCAGTGCCCAAGTCTACACAGGCCGGCGAGGGGACTC    242
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              301 GTCACTGAGCGCTCCAGTGCCCAAGTCTACACAGGCCGGCGAGGGGACTC    350

ref              243 TGAGTGAAGCTCTGCTGCACCTGCAGTTCGACGCTGATGAGGACCTGGGA    292
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              351 TGAGTGAAGCTCTGCTGCACCTGCAGTTCGACGCTGATGAGGACCTGGGA    400

ref              293 GCTCTGCTGGGGAACAGCACCGATCCCGGAGTGTTCACAGATCTGGCCTC    342
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              401 GCTCTGCTGGGGAACAGCACCGATCCCGGAGTGTTCACAGATCTGGCCTC    450

ref              343 CGTGGACAACTCTGAGTTTCAGCAGCTGCTGAATCAGGGCGTGTCCATGT    392
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              451 CGTGGACAACTCTGAGTTTCAGCAGCTGCTGAATCAGGGCGTGTCCATGT    500

ref              393 CTCATAGTACAGCCGAACCAATGCTGATGGAGTACCCCGAAGCCATTACC    442
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              501 CTCATAGTACAGCCGAACCAATGCTGATGGAGTACCCCGAAGCCATTACC    550

ref              443 CGGCTGGTGACCGGCAGCCAGCGGCCCCCCGACCCCGCTCCAACTCCCCT    492
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              551 CGGCTGGTGACCGGCAGCCAGCGGCCCCCCGACCCCGCTCCAACTCCCCT    600

ref              493 GGGAACCAGCGGCCTGCCTAATGGGCTGTCCGGAGATGAAGACTTCTCAA    542
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              601 GGGAACCAGCGGCCTGCCTAATGGGCTGTCCGGAGATGAAGACTTCTCAA    650

ref              543 GCATCGCTGATATGGACTTTAGTGCCCTGCTGTCACAGATTTCCTCTAGT    592
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              651 GCATCGCTGATATGGACTTTAGTGCCCTGCTGTCACAGATTTCCTCTAGT    700

ref              593 GGGCAGGGAGGAGGTGGAAGCGGCTTCAGCGTGGACACCAGTGCCCTGCT    642
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              701 GGGCAGGGAGGAGGTGGAAGCGGCTTCAGCGTGGACACCAGTGCCCTGCT    750

ref              643 GGACCTGTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACC    692
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              751 GGACCTGTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACC    800

ref              693 TTGACAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCC    742
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              801 TTGACAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCC    850

ref              743 CCCAGGCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCT    792
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              851 CCCAGGCCTCCCGAGGCAGAGAACAGCAGCccGGATTCAGGGAAGCAGCT    900

ref              793 GGTGCACTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGG    842
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              901 GGTGCACTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCggCTCCGTGG    950

ref              843 ACACC-GGGAGCAACGACCTGCC-GGTGCTGTTTGAG-CTGGGAGAGGGC    889
                     ||||| |||||||||||||||| ||||||||||||. |||||..|.|||
seq              951 ACACCGGggAGCAACGACCTgccnggtgCTGTTTganccTGGggaanggc   1000

ref              890 -TCCTACTT-CTCCGAA--GGGGACGG-CTTCGCCGAGGACCCCACCATC    934
                      .|||||||  |||||||  |||.|||| |||| ||..|.|
seq             1001 nnCcTACTtncTCCGAAAGGGgAaCGgccTtc-cCnnGaa----------   1039

SEQ_ID: 005
Name: pNLS-p65-HSF1-PAL-T2A-mCherry
Clone: 3
Service: GATC
Primer: pEGFP_C2-RP
Direction: Reverse


ref             1851 AGGTGCTGGGCC--GCAA--CTGCCGC----TTCCT--GCAGG--CCGAG   1888
                       .||  |||||  ||||  .|||| |    |||||  .||||  |.|||
aqua3              1 --tTg--GGGcCCngcAAACttgcc-CgGttTTCCtTgccAgGGCCngaG     45

ref             1889 T---CCGGCG--ATCCG-CAC-GAG-CGGT--CGGCGA-TTCG--CTCGG   1925
```

```
                    |   |.||||  .|||| ||| ||| ||||  .||||| ||||  .||||
aqua3                 46 TTCCCnggcggattcCgccacgGagccgGTTCnGGcgaTTTCggcntcgG     95

ref                 1926 -CGA-TCGCC--AAC-GGCGACGCGGTCA-CCACCC-TGA-TCCGCAA--   1965
                         ||| |||||  ||| ||.|||||||||| |||||| ||| |||||||
aqua3                 96 ccGATTCGCCCnAACggGnGACGcgGtcACCCACCCTTGAttcCGCAAAc    145

ref                 1966 CTTCC-GCCA-GGAC-GGGCACG-CCTTCTGG-AACGAGTT-CCACCTGT   2009
                         .|||| |||| |||| ||||||| |||||||| |||||||| ||.|||||
aqua3                146 nTTCCngcCAngGaCggGGcACGncCtTCTGGaAACGAGtTcccncCTGT    195

ref                 2010 -CCCCCGTGCGC-AACGGGGCCGG-CCGGGTCA-CCCATTACATC-GGCT   2054
                          |||||||||||| |||||||||||| |||||||| |||||||||||| ||||
aqua3                196 CCCCCCGTGCGCaaacgggGCCgGcccgggtCAccCcatTACATCggGCT    245

ref                 2055 A-CCAGCTCGACGTCA-CCGAGC-GGGTCGAGCGCGATCAGCAGTTGGAG   2101
                         | ||||||||||||||| |||||| ||||||||||||||||||||||||||
aqua3                246 anccAGCtcgaCGTCACCCGAGCgggGTCGAGCGCGATCAGCAGTTGGAG    295

ref                 2102 CAGCTGGCGTCGGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGA   2151
                         ||||||||||||||||||||||||||||||||||||||||||||||||||||
aqua3                296 CAGCTGGCGTCGGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGA    345

ref                 2152 GGAGAATCCTGGCCCAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCA   2201
                         ||||||||||||||||||||||||||||||||||||||||||||||||||||
aqua3                346 GGAGAATCCTGGCCCAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCA    395

ref                 2202 TCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAAC   2251
                         ||||||||||||||||||||||||||||||||||||||||||||||||||||
aqua3                396 TCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAAC    445

ref                 2252 GGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGG   2301
                         ||||||||||||||||||||||||||||||||||||||||||||||||||||
aqua3                446 GGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGG    495

ref                 2302 CACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCG   2351
                         ||||||||||||||||||||||||||||||||||||||||||||||||||||
aqua3                496 CACCCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCG    545

ref                 2352 CCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTG   2401
                         ||||||||||||||||||||||||||||||||||||||||||||||||||||
aqua3                546 CCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTG    595

ref                 2402 AAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGG   2451
                         ||||||||||||||||||||||||||||||||||||||||||||||||||||
aqua3                596 AAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGG    645

ref                 2452 CTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCG   2501
                         ||||||||||||||||||||||||||||||||||||||||||||||||||||
aqua3                646 CTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCG    695

ref                 2502 TGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAG   2551
                         ||||||||||||||||||||||||||||||||||||||||||||||||||||
aqua3                696 TGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAG    745

ref                 2552 CTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGAC   2601
                         ||||||||||||||||||||||||||||||||||||||||||||||||||||
aqua3                746 CTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGAC    795

ref                 2602 CATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCC   2651
                         ||||||||||||||||||||||||||||||||||||||||||||||||||||
aqua3                796 CATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCC    845

ref                 2652 TGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTAC   2701
                         ||||||||||||||||||||||||||||||||||||||||||||||||||||
aqua3                846 TGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTAC    895

ref                 2702 GACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCC   2751
                         ||||||||||||||||||||||||||||||||||||||||||||||||||||
aqua3                896 GACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCC    945

ref                 2752 CGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGG   2801
                         ||||||||||||||||||||||||||||||||||||||||||||||||||||
aqua3                946 CGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGG    995
```

```
ref            2802 ACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACC   2851
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
aqua3           996 ACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACC   1045

ref            2852 GGCGGCATGGACGAGCTGTACAAGtaaagcggccgcgactctagatcata   2901
                    |||||||||||||||||||||||||||||||||||||||||||
aqua3          1046 GGCGGCATGGACGAGCTGTACAAgtaAAGCGGCCGCGACtCt--------   1087
```

SEQ_ID: 006
Name: pNLS-HSF1-PAL-T2A-mCherry_mutated
Clone: 2
Service: SeqLab
Primer: CMV-Fwd
Direction: Forward

```
nhptm_ref         1 ----cgccaccatgcctaagaaaaagaggaaggtgAGTGGGCAGGGAGGA     46
                        |||||||||||||||||||||||||||||| ||||||||||||||||
c2              101 CGGTCGCCACCATGCCTAAGAAAAAGAGGAAGGT-AGTGGGCAGGGAGGA    149

nhptm_ref        47 GGTGGAAGCGGCTTCAGCGTGGACACCAGTGCCCTGCTGGACCTGTTCAG     96
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c2              150 GGTGGAAGCGGCTTCAGCGTGGACACCAGTGCCCTGCTGGACCTGTTCAG    199

nhptm_ref        97 CCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGACAGCAGCC    146
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c2              200 CCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGACAGCAGCC    249

nhptm_ref       147 TGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCAGGCCTCCC    196
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c2              250 TGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCAGGCCTCCC    299

nhptm_ref       197 GAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCACTACAC    246
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c2              300 GAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCACTACAC    349

nhptm_ref       247 AGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCGGGAGCA    296
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c2              350 AGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCGGGAGCA    399

nhptm_ref       297 ACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTCTCCGAA    346
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c2              400 ACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTCTCCGAA    449

nhptm_ref       347 GGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGGCTCGGA    396
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c2              450 GGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGGCTCGGA    499

nhptm_ref       397 GCCTCCCAAAGCCAAGGACCCCACTGTCTCC------------ggtggtt    434
                    |||||||||||||||||||||||||||||||            |||||||
c2              500 GCCTCCCAAAGCCAAGGACCCCACTGTCTCCGGTGGTTCGGGTGGTGGTT    549

nhptm_ref       435 cgggtggtggttcgggtggtggttcgggtAAGGTGAACCGGCCCGCGGAG    484
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c2              550 CGGGTGGTGGTTCGGGTGGTGGTTCGGGTAAGGTGAACCGGCCCGCGGAG    599

nhptm_ref       485 CGCGCCTCGTTCGGTTCGTTCGTGCTCGACGCCGGGTCCGCCCGGTTCGT    534
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c2              600 CGCGCCTCGTTCGGTTCGTTCGTGCTCGACGCCGGGTCCGCCCGGTTCGT    649

nhptm_ref       535 CGGCTCCGACGAGTTGGCGCTGGTGCTTGGCTTCGCCCCGGGTGATGTGG    584
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c2              650 CGGCTCCGACGAGTTGGCGCTGGTGCTTGGCTTCGCCCCGGGTGATGTGG    699

nhptm_ref       585 TGCTGACGCCGGCCGTGGTGCTGGCCCACCTGCATCCGGACGATCGGCTG    634
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c2              700 TGCTGACGCCGGCCGTGGTGCTGGCCCACCTGCATCCGGACGATCGGCTG    749

nhptm_ref       635 GAATGGCAGGCCGG----------------------------------    648
                    ||||||||||||||
c2              750 GAATGGCAGGCCGGTCTGCAGCGATGCCTGGCCACCGGGCGACCCGTGGT    799
```

```
SEQ_ID: 007
Name: pNLS-HSF1-p65-PAL-T2A-mCherry_mutated
Clone: 1
Service: SeqLab
Primer: CMV-Fwd
Direction: Forward


NHPPTM_ref         1 --------------------atgcctaagaaaaagaggaaggtgAGTGGG     30
                                         |||||||||||||||||||||||| ||||||
clone1           101 GGGATCCACCGGTCGCCACCATGCCTAAGAAAAAGAGGAAGGT-AGTGGG    149

NHPPTM_ref        31 CAGGGAGGAGGTGGAAGCGGCTTCAGCGTGGACACCAGTGCCCTGCTGGA     80
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           150 CAGGGAGGAGGTGGAAGCGGCTTCAGCGTGGACACCAGTGCCCTGCTGGA    199

NHPPTM_ref        81 CCTGTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTG    130
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           200 CCTGTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTG    249

NHPPTM_ref       131 ACAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCC    180
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           250 ACAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCC    299

NHPPTM_ref       181 AGGCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGT    230
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           300 AGGCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGT    349

NHPPTM_ref       231 GCACTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACA    280
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           350 GCACTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACA    399

NHPPTM_ref       281 CCGGGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTAC    330
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           400 CCGGGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTAC    449

NHPPTM_ref       331 TTCTCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGAC    380
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           450 TTCTCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGAC    499

NHPPTM_ref       381 AGGCTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCCgcggccgctg    430
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           500 AGGCTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCCGCGGCCGCTG    549

NHPPTM_ref       431 gatccCCTTCAGGGCAGATCAGCAACCAGGCCCTGGCTCTGGCCCCTAGC    480
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           550 GATCCCCTTCAGGGCAGATCAGCAACCAGGCCCTGGCTCTGGCCCCTAGC    599

NHPPTM_ref       481 TCCGCTCCAGTGCTGGCCCAGACTATGGTGCCCTCTAGTGCTATGGTGCC    530
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           600 TCCGCTCCAGTGCTGGCCCAGACTATGGTGCCCTCTAGTGCTATGGTGCC    649

NHPPTM_ref       531 TCTGGCCCAGCCACCTGCTCCAGCCCCTGTGCTGACCCCAGGACCACCCC    580
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           650 TCTGGCCCAGCCACCTGCTCCAGCCCCTGTGCTGACCCCAGGACCACCCC    699

NHPPTM_ref       581 AGTCACTGAGCGCTCCAGTGCCCAAGTCTACACAGGCCGGCGAGGGGACT    630
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           700 AGTCACTGAGCGCTCCAGTGCCCAAGTCTACACAGGCCGGCGAGGGGACT    749

NHPPTM_ref       631 CTGAGTGAAGCTCTGCTGCACCTGCAGTTCGACGCTGATGAGGACCTGGG    680
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           750 CTGAGTGAAGCTCTGCTGCACCTGCAGTTCGACGCTGATGAGGACCTGGG    799

NHPPTM_ref       681 AGCTCTGCTGGGGAACAGCACCGATCCCGGAGTGTTCACAGATCTGGCCT    730
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           800 AGCTCTGCTGGGGAACAGCACCGATCCCGGAGTGTTCACAGATCTGGCCT    849

NHPPTM_ref       731 CCGTGGACAACTCTGAGTTTCAGCAGCTGCTGAATCAGGGCGTGTCCATG    780
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           850 CCGTGGACAACTCTGAGTTTCAGCAGCTGCTGAATCAGGGCGTGTCCATG    899

NHPPTM_ref       781 TCTCATAGTACAGCCGAACCAATGCTGATGGAGTACCCCGAAGCCATTAC    830
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           900 TCTCATAGTACAGCCGAACCAATGCTGATGGAGTACCCCGAAGCCATTAC    949
```

```
NHPPTM_ref       831 CCGGCTGGTGACCGGCAGCCAGCGGCCCCCCGACCCCGCTCCAACTCCCC    880
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           950 CCGGCTGGTGACCGGCAGCCAGCGGCCCCCCGACCCCGCTCCAACTCCCC    999

NHPPTM_ref       881 TGGGAACCAGCGGCCTGCCTAATGGGCTGTCCGGAGATGAAGACTTCTCA    930
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1          1000 TGGGAACCAGCGGCCTGCCTAATGGGCTGTCCGGAGATGAAGACTTCTCA   1049

NHPPTM_ref       931 AGCATCGCTGATATGGACTTTAGTGCCCTGCTGTCACAGATTTCCTCTgg    980
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1          1050 AGCATCGCTGATATGGACTTTAGTGCCCTGCTGTCACAGATTTCCTCTGG   1099

NHPPTM_ref       981 tggttcgggtggtggttcgggtggtggttcgggtAAGGTGAA--------   1022
                     ||||||||||||||||||||||||||||||||||  ||||..
clone1          1100 TGGTTCGGGTGGTGGTTCGGGTGGTGGTTCGGGT--GGTGGTTCGGGTGG   1147
```

SEQ_ID: 008
Name: pNLS-HSF1-p65-PAL-T2A-mCherry
Clone: 1
Service: SeqLab E. coli NightSeq
Primer: CMV-Fwd
Direction: Forward

```
ref                 1 tatataagcagagctggtttagtgaaccgtcagatccgctagcgctaccg     50
                            ||||||||||||||||||||||||||||||||||||||||||||||||
clone1              1 -------GCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTACCG     43

ref                51 gactcagatctcgagctcaagcttcgaattctgcagtcgacggtaccgcg    100
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1             44 GACTCAGATCTCGAGCTCAAGCTTCGAATTCTGCAGTCGACGGTACCGCG     93

ref               101 ggcccgggatccaccggtcgccaccatgcctaagaaaaagaggaaggtgA    150
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1             94 GGCCCGGGATCCACCGGTCGCCACCATGCCTAAGAAAAAGAGGAAGGTGA    143

ref               151 GTGGGCAGGGAGGAGGTGGAAGCGGCTTCAGCGTGGACACCAGTGCCCTG    200
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1            144 GTGGGCAGGGAGGAGGTGGAAGCGGCTTCAGCGTGGACACCAGTGCCCTG    193

ref               201 CTGGACCTGTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGA    250
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1            194 CTGGACCTGTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGA    243

ref               251 CCTTGACAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGC    300
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1            244 CCTTGACAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGC    293

ref               301 CCCCCAGGCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAG    350
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1            294 CCCCCAGGCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAG    343

ref               351 CTGGTGCACTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGT    400
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1            344 CTGGTGCACTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGT    393

ref               401 GGACACCGGGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCT    450
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1            394 GGACACCGGGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCT    443

ref               451 CCTACTTCTCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTG    500
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1            444 CCTACTTCTCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTG    493

ref               501 CTGACAGGCTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCCgcggc    550
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1            494 CTGACAGGCTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCCGCGGC    543

ref               551 cgctggatccCCTTCAGGGCAGATCAGCAACCAGGCCCTGGCTCTGGCCC    600
                       ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1            544 CGCTGGATCCCCTTCAGGGCAGATCAGCAACCAGGCCCTGGCTCTGGCCC    593

ref               601 CTAGCTCCGCTCCAGTGCTGGCCCAGACTATGGTGCCCTCTAGTGCTATG    650
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
clone1          594 CTAGCTCCGCTCCAGTGCTGGCCCAGACTATGGTGCCCTCTAGTGCTATG    643

ref             651 GTGCCTCTGGCCCAGCCACCTGCTCCAGCCCCTGTGCTGACCCCAGGACC    700
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1          644 GTGCCTCTGGCCCAGCCACCTGCTCCAGCCCCTGTGCTGACCCCAGGACC    693

ref             701 ACCCCAGTCACTGAGCGCTCCAGTGCCCAAGTCTACACAGGCCGGCGAGG    750
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1          694 ACCCCAGTCACTGAGCGCTCCAGTGCCCAAGTCTACACAGGCCGGCGAGG    743

ref             751 GGACTCTGAGTGAAGCTCTGCTGCACCTGCAGTTCGACGCTGATGAGGAC    800
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1          744 GGACTCTGAGTGAAGCTCTGCTGCACCTGCAGTTCGACGCTGATGAGGAC    793

ref             801 CTGGGAGCTCTGCTGGGGAACAGCACCGATCCCGGAGTGTTCACAGATCT    850
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1          794 CTGGGAGCTCTGCTGGGGAACAGCACCGATCCCGGAGTGTTCACAGATCT    843

ref             851 GGCCTCCGTGGACAACTCTGAGTTTCAGCAGCTGCTGAATCAGGGCGTGT    900
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1          844 GGCCTCCGTGGACAACTCTGAGTTTCAGCAGCTGCTGAATCAGGGCGTGT    893

ref             901 CCATGTCTCATAGTACAGCCGAACCAATGCTGATGGAGTACCCCGAAGCC    950
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1          894 CCATGTCTCATAGTACAGCCGAACCAATGCTGATGGAGTACCCCGAAGCC    943

ref             951 ATTACCCGGCTGGTGACCGGCAGCCAGCGGCCCCCCGACCCCGCTCCAAC   1000
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1          944 ATTACCCGGCTGGTGACCGGCAGCCAGCGGCCCCCCGACCCCGCTCCAAC    993

ref            1001 TCCCCTGGGAACCAGCGGCCTGCCTAATGGGCTGTCCGGAGATGAAGACT   1050
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1          994 TCCCCTGGGAACCAGCGGCCTGCCTAATGGGCTGTCCGGAGATGAAGACT   1043

ref            1051 TCTCAAGCATCGCTGATATGGACTTTAGTGCCCTGCTGTCACAGATTTCC   1100
                    ||||||||||||
clone1         1044 TCTCAAGCATCG------------------------------------   1055
```

SEQ_ID: 009
Name: pNLS-p65-HSF1-PAL-T2A-mCherry_cor_1
Clone: 1
Service: SeqLab
Primer: CMV-Fwd
Direction: Forward

```
NPHPTM_ref        1 ----cgccaccatgcctaagaaaaagaggaaggtggcggccgctggatcc     46
                        ||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1         101 CGGTCGCCACCATGCCTAAGAAAAAGAGGAAGGTGGCGGCCGCTGGATCC    150

NPHPTM_ref       47 CCTTCAGGGCAGATCAGCAACCAGGCCCTGGCTCTGGCCCCTAGCTCCGC     96
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1         151 CCTTCAGGGCAGATCAGCAACCAGGCCCTGGCTCTGGCCCCTAGCTCCGC    200

NPHPTM_ref       97 TCCAGTGCTGGCCCAGACTATGGTGCCCTCTAGTGCTATGGTGCCTCTGG    146
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1         201 TCCAGTGCTGGCCCAGACTATGGTGCCCTCTAGTGCTATGGTGCCTCTGG    250

NPHPTM_ref      147 CCCAGCCACCTGCTCCAGCCCCTGTGCTGACCCCAGGACCACCCCAGTCA    196
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1         251 CCCAGCCACCTGCTCCAGCCCCTGTGCTGACCCCAGGACCACCCCAGTCA    300

NPHPTM_ref      197 CTGAGCGCTCCAGTGCCCAAGTCTACACAGGCCGGCGAGGGGACTCTGAG    246
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1         301 CTGAGCGCTCCAGTGCCCAAGTCTACACAGGCCGGCGAGGGGACTCTGAG    350

NPHPTM_ref      247 TGAAGCTCTGCTGCACCTGCAGTTCGACGCTGATGAGGACCTGGGAGCTC    296
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
clone_1          351 TGAAGCTCTGCTGCACCTGCAGTTCGACGCTGATGAGGACCTGGGAGCTC    400

NPHPTM_ref       297 TGCTGGGGAACAGCACCGATCCCGGAGTGTTCACAGATCTGGCCTCCGTG    346
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1          401 TGCTGGGGAACAGCACCGATCCCGGAGTGTTCACAGATCTGGCCTCCGTG    450

NPHPTM_ref       347 GACAACTCTGAGTTTCAGCAGCTGCTGAATCAGGGCGTGTCCATGTCTCA    396
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1          451 GACAACTCTGAGTTTCAGCAGCTGCTGAATCAGGGCGTGTCCATGTCTCA    500

NPHPTM_ref       397 TAGTACAGCCGAACCAATGCTGATGGAGTACCCCGAAGCCATTACCCGGC    446
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1          501 TAGTACAGCCGAACCAATGCTGATGGAGTACCCCGAAGCCATTACCCGGC    550

NPHPTM_ref       447 TGGTGACCGGCAGCCAGCGGCCCCCCGACCCCGCTCCAACTCCCCTGGGA    496
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1          551 TGGTGACCGGCAGCCAGCGGCCCCCCGACCCCGCTCCAACTCCCCTGGGA    600

NPHPTM_ref       497 ACCAGCGGCCTGCCTAATGGGCTGTCCGGAGATGAAGACTTCTCAAGCAT    546
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1          601 ACCAGCGGCCTGCCTAATGGGCTGTCCGGAGATGAAGACTTCTCAAGCAT    650

NPHPTM_ref       547 CGCTGATATGGACTTTAGTGCCCTGCTGTCACAGATTTCCTCTAGTGGGC    596
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1          651 CGCTGATATGGACTTTAGTGCCCTGCTGTCACAGATTTCCTCTAGTGGGC    700

NPHPTM_ref       597 AGGGAGGAGGTGGAAGCGGCTTCAGCGTGGACACCAGTGCCCTGCTGGAC    646
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1          701 AGGGAGGAGGTGGAAGCGGCTTCAGCGTGGACACCAGTGCCCTGCTGGAC    750

NPHPTM_ref       647 CTGTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGA    696
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1          751 CTGTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGA    800

NPHPTM_ref       697 CAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCA    746
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1          801 CAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCA    850

NPHPTM_ref       747 GGCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTG    796
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1          851 GGCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTG    900

NPHPTM_ref       797 CACTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACAC    846
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1          901 CACTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACAC    950

NPHPTM_ref       847 CGGGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACT    896
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1          951 CGGGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACT   1000

NPHPTM_ref       897 TCTCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACA    946
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_1         1001 TCTCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACA   1050

NPHPTM_ref       947 GGCTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCC-----------    985
                     |||||||||||||||||||||||||| ||||||||||||
clone_1         1051 GGCTCGGAGCCTCCCAAAGCCAAGGA-CCCACTGTCTCCGGTGTTCGGTG   1099


SEQ_ID: 010
Name: pNLS-p65-HSF1-PAL-T2A-mCherry_cor_2
Clone: 2
Service: SeqLab
Primer: CMV-Fwd
Direction: Forward


NPHPTM_ref         1 ------------------------------------------------cgc      3
                                                                     |||
clone_2           51 CTTCGAATTCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGTCGC    100

NPHPTM_ref         4 caccatgcctaagaaaaagaggaaggtggcggccgctggatccCCTTCAG     53
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          101 CACCATGCCTAAGAAAAAGAGGAAGGTGGCGGCCGCTGGATCCCCTTCAG    150

NPHPTM_ref        54 GGCAGATCAGCAACCAGGCCCTGGCTCTGGCCCCTAGCTCCGCTCCAGTG    103
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          151 GGCAGATCAGCAACCAGGCCCTGGCTCTGGCCCCTAGCTCCGCTCCAGTG    200

NPHPTM_ref       104 CTGGCCCAGACTATGGTGCCCTCTAGTGCTATGGTGCCTCTGGCCCAGCC    153
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          201 CTGGCCCAGACTATGGTGCCCTCTAGTGCTATGGTGCCTCTGGCCCAGCC    250

NPHPTM_ref       154 ACCTGCTCCAGCCCCTGTGCTGACCCCAGGACCACCCCAGTCACTGAGCG    203
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          251 ACCTGCTCCAGCCCCTGTGCTGACCCCAGGACCACCCCAGTCACTGAGCG    300

NPHPTM_ref       204 CTCCAGTGCCCAAGTCTACACAGGCCGGCGAGGGGACTCTGAGTGAAGCT    253
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          301 CTCCAGTGCCCAAGTCTACACAGGCCGGCGAGGGGACTCTGAGTGAAGCT    350

NPHPTM_ref       254 CTGCTGCACCTGCAGTTCGACGCTGATGAGGACCTGGGAGCTCTGCTGGG    303
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          351 CTGCTGCACCTGCAGTTCGACGCTGATGAGGACCTGGGAGCTCTGCTGGG    400

NPHPTM_ref       304 GAACAGCACCGATCCCGGAGTGTTCACAGATCTGGCCTCCGTGGACAACT    353
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          401 GAACAGCACCGATCCCGGAGTGTTCACAGATCTGGCCTCCGTGGACAACT    450

NPHPTM_ref       354 CTGAGTTTCAGCAGCTGCTGAATCAGGGCGTGTCCATGTCTCATAGTACA    403
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          451 CTGAGTTTCAGCAGCTGCTGAATCAGGGCGTGTCCATGTCTCATAGTACA    500

NPHPTM_ref       404 GCCGAACCAATGCTGATGGAGTACCCCGAAGCCATTACCCGGCTGGTGAC    453
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          501 GCCGAACCAATGCTGATGGAGTACCCCGAAGCCATTACCCGGCTGGTGAC    550

NPHPTM_ref       454 CGGCAGCCAGCGGCCCCCCGACCCCGCTCCAACTCCCCTGGGAACCAGCG    503
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          551 CGGCAGCCAGCGGCCCCCCGACCCCGCTCCAACTCCCCTGGGAACCAGCG    600

NPHPTM_ref       504 GCCTGCCTAATGGGCTGTCCGGAGATGAAGACTTCTCAAGCATCGCTGAT    553
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          601 GCCTGCCTAATGGGCTGTCCGGAGATGAAGACTTCTCAAGCATCGCTGAT    650

NPHPTM_ref       554 ATGGACTTTAGTGCCCTGCTGTCACAGATTTCCTCTAGTGGGCAGGGAGG    603
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          651 ATGGACTTTAGTGCCCTGCTGTCACAGATTTCCTCTAGTGGGCAGGGAGG    700

NPHPTM_ref       604 AGGTGGAAGCGGCTTCAGCGTGGACACCAGTGCCCTGCTGGACCTGTTCA    653
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          701 AGGTGGAAGCGGCTTCAGCGTGGACACCAGTGCCCTGCTGGACCTGTTCA    750

NPHPTM_ref       654 GCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGACAGCAGC    703
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          751 GCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGACAGCAGC    800

NPHPTM_ref       704 CTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCAGGCCTCC    753
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          801 CTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCAGGCCTCC    850

NPHPTM_ref       754 CGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCACTACA    803
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          851 CGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCACTACA    900

NPHPTM_ref       804 CAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCGGGAGC    853
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          901 CAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCGGGAGC    950

NPHPTM_ref       854 AACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTCTCCGA    903
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2          951 AACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTCTCCGA   1000

NPHPTM_ref       904 AGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGGCTCGG    953
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_2         1001 AGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGGCTCGG   1050

NPHPTM_ref       954 AGCCTCCCAAAGCCAAGGACCCCACTGTCTCC--------------    985
                    |||||||||||||||||| ||||||||||||
clone_2         1051 AGCCTCCCAAAGCCAAGGA-CCCACTGTCTCCGGGGTCGGTGGGTK   1095
```

```
SEQ_ID: 011
Name: pNLS-p65-HSF1-PAL-T2A-mCherry_cor_3
Clone: 3
Service: SeqLab
Primer: CMV-Fwd
Direction: Forward

NPHPTM_ref         1 ------cgccaccatgcctaagaaaaagaggaaggtggcggccgctggat     44
                       ||||||||||||||||||||||||||||||||||||||||||||||
clone_3          101 ACCGGTCGCCACCATGCCTAAGAAAAAGAGGAAGGTGGCGGCCGCTGGAT    150

NPHPTM_ref        45 ccCCTTCAGGGCAGATCAGCAACCAGGCCCTGGCTCTGGCCCCTAGCTCC     94
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          151 CCCCTTCAGGGCAGATCAGCAACCAGGCCCTGGCTCTGGCCCCTAGCTCC    200

NPHPTM_ref        95 GCTCCAGTGCTGGCCCAGACTATGGTGCCCTCTAGTGCTATGGTGCCTCT    144
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          201 GCTCCAGTGCTGGCCCAGACTATGGTGCCCTCTAGTGCTATGGTGCCTCT    250

NPHPTM_ref       145 GGCCCAGCCACCTGCTCCAGCCCCTGTGCTGACCCCAGGACCACCCCAGT    194
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          251 GGCCCAGCCACCTGCTCCAGCCCCTGTGCTGACCCCAGGACCACCCCAGT    300

NPHPTM_ref       195 CACTGAGCGCTCCAGTGCCCAAGTCTACACAGGCCGGCGAGGGGACTCTG    244
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          301 CACTGAGCGCTCCAGTGCCCAAGTCTACACAGGCCGGCGAGGGGACTCTG    350

NPHPTM_ref       245 AGTGAAGCTCTGCTGCACCTGCAGTTCGACGCTGATGAGGACCTGGGAGC    294
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          351 AGTGAAGCTCTGCTGCACCTGCAGTTCGACGCTGATGAGGACCTGGGAGC    400

NPHPTM_ref       295 TCTGCTGGGGAACAGCACCGATCCCGGAGTGTTCACAGATCTGGCCTCCG    344
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          401 TCTGCTGGGGAACAGCACCGATCCCGGAGTGTTCACAGATCTGGCCTCCG    450

NPHPTM_ref       345 TGGACAACTCTGAGTTTCAGCAGCTGCTGAATCAGGGCGTGTCCATGTCT    394
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          451 TGGACAACTCTGAGTTTCAGCAGCTGCTGAATCAGGGCGTGTCCATGTCT    500

NPHPTM_ref       395 CATAGTACAGCCGAACCAATGCTGATGGAGTACCCCGAAGCCATTACCCG    444
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          501 CATAGTACAGCCGAACCAATGCTGATGGAGTACCCCGAAGCCATTACCCG    550

NPHPTM_ref       445 GCTGGTGACCGGCAGCCAGCGGCCCCCCGACCCCGCTCCAACTCCCCTGG    494
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          551 GCTGGTGACCGGCAGCCAGCGGCCCCCCGACCCCGCTCCAACTCCCCTGG    600

NPHPTM_ref       495 GAACCAGCGGCCTGCCTAATGGGCTGTCCGGAGATGAAGACTTCTCAAGC    544
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          601 GAACCAGCGGCCTGCCTAATGGGCTGTCCGGAGATGAAGACTTCTCAAGC    650

NPHPTM_ref       545 ATCGCTGATATGGACTTTAGTGCCCTGCTGTCACAGATTTCCTCTAGTGG    594
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          651 ATCGCTGATATGGACTTTAGTGCCCTGCTGTCACAGATTTCCTCTAGTGG    700

NPHPTM_ref       595 GCAGGGAGGAGGTGGAAGCGGCTTCAGCGTGGACACCAGTGCCCTGCTGG    644
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          701 GCAGGGAGGAGGTGGAAGCGGCTTCAGCGTGGACACCAGTGCCCTGCTGG    750

NPHPTM_ref       645 ACCTGTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTT    694
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          751 ACCTGTTCAGCCCCTCGGTGACCGTGCCCGACATGAGCCTGCCTGACCTT    800

NPHPTM_ref       695 GACAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCC    744
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          801 GACAGCAGCCTGGCCAGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCC    850

NPHPTM_ref       745 CAGGCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGG    794
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          851 CAGGCCTCCCGAGGCAGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGG    900
```

```
NPHPTM_ref       795 TGCACTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGAC    844
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          901 TGCACTACACAGCGCAGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGAC    950

NPHPTM_ref       845 ACCGGGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTA    894
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3          951 ACCGGGAGCAACGACCTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTA   1000

NPHPTM_ref       895 CTTCTCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGA    944
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone_3         1001 CTTCTCCGAAGGGGACGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGA   1050

NPHPTM_ref       945 CAGGCTCGGAGCCTCCCAAAGCCAAGGACCCCACTGTCTCC---------    985
                     |||||||||||||||||||||||||||| ||||||||||||
clone_3         1051 CAGGCTCGGAGCCTCCCAAAGCCAAGGA-CCCACTGTCTCCGTGGTCGGT   1099
```

SEQ_ID: 012
Name: pdCas9_eGFP
Clone: 3
Service: SeqLab
Primer: dCas9_rev_seq primer
Direction: Reverse

```
ref_dCas9        951 CCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGG   1000
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone3           853 CCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCTGG    902

ref_dCas9       1001 GAGGCGACAGCGCTGGAGGAGGTGGAAGCGGAGGAGGAGGAAGCGGAGGA   1050
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone3           903 GAGGCGACAGCGCTGGAGGAGGTGGAAGCGGAGGAGGAGGAAGCGGAGGA    952

ref_dCas9       1051 GGAGGTAGCGGACCTAAGAAAAAGAGGAAGGTGGCGGCCATTAACGCTAG   1100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone3           953 GGAGGTAGCGGACCTAAGAAAAAGAGGAAGGTGGCGGCCATTAACGCTAG   1002

ref_dCas9       1101 CGGCAGTGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGG   1150
                     |||||||||||||||||||||||||..|
clone3          1003 CGGCAGTGGAGAGGGCAGAGGAAGnnt----------------------   1029
```

SEQ_ID: 013
Name: pENTR.hU6.sgRNA.CMV.native
Clone: 2
Service: GATC
Primer: KAN2-RP
Direction: Reverse

```
pENTR_empty        1 -------------------------GGGAAAGGTCGAGAAACTGCAAAGT     25
                                               |||||||||||||||||||||||||
1_2              901 TATCTTGTGGAAAGGACGAAACACCGGGAAAGGTCGAGAAACTGCAAAGT    950

pENTR_empty       26 TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTT     75
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
1_2              951 TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTT   1000

pENTR_empty       76 GAAAAAGTGGCACCGAGTCGGTGCTTTTT----------------------    104
                     ||||||||||||||||||||||||||||||
1_2              1001 GAAAAAGTGGCACCGAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTAC   1050
```

SEQ_ID: 014
Name: pENTR.hU6.sgRNA.CMV.sgRNA2.0
Clone: 3
Service: GATC
Primer: KAN2-RP
Direction: Reverse

```
ref20              1 ----------------------------------------------GGGA      4
```

```
                                                                ||||
2_3             801 TTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGGA    850

ref20             5 AAGGTCGAGAAACTGCAAAGTTTTAGAGCTAGGCCAACATGAGGATCACC     54
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
2_3             851 AAGGTCGAGAAACTGCAAAGTTTTAGAGCTAGGCCAACATGAGGATCACC    900

ref20            55 CATGTCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACT    104
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
2_3             901 CATGTCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACT    950

ref20           105 TGGCCAACATGAGGATCACCCATGTCTGCAGGGCCAAGTGGCACCGAGTC    154
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
2_3             951 TGGCCAACATGAGGATCACCCATGTCTGCAGGGCCAAGTGGCACCGAGTC   1000

ref20           155 GGTGCTTTTT----------------------------------------    164
                    ||||||||||
2_3            1001 GGTGCTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAA   1050
```

SEQ_ID: 015
Name: pENTR.hU6.sgRNA.CMV.SG6
Clone: 1
Service: GATC
Primer: KAN2-RP
Direction: Reverse

```
refm2s6           1 --------------------------------GGGAAAGGTCGAGAAAC     17
                                                     |||||||||||||||||
3_1             801 GCTTTATATATCTTGTGGAAAGGACGAAACACCGGGAAAGGTCGAGAAAC    850

refm2s6          18 TGCAAAGTTTTAGAGCTAGGCCAGGCCGCGTACAGCAGCGATGCGCGGCC     67
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
3_1             851 TGCAAAGTTTTAGAGCTAGGCCAGGCCGCGTACAGCAGCGATGCGCGGCC    900

refm2s6          68 CTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCC    117
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
3_1             901 CTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCC    950

refm2s6         118 AGCGCGGGTACAGCAGCGATGCCCGCGCCTGCAGGGCCAAGTGGCACCGA    167
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
3_1             951 AGCGCGGGTACAGCAGCGATGCCCGCGCCTGCAGGGCCAAGTGGCACCGA   1000

refm2s6         168 GTCGGTGCTTTTT-------------------------------------    180
                    |||||||||||||
3_1            1001 GTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAA   1050
```

SEQ_ID: 016
Name: pENTR.hU6.sgRNA.CMV.SG9
Clone: 3
Service: GATC
Primer: KAN2-RP
Direction: Reverse

```
refm2s9           1 ------GGGAAAGGTCGAGAAACTGCAAAGTTTTAGAGCTAGGCCAGGCC     44
                          ||||||||||||||||||||||||||||||||||||||||||||
4_3             651 AACACCGGGAAAGGTCGAGAAACTGCAAAGTTTTAGAGCTAGGCCAGGCC    700

refm2s9          45 GCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAGGGCCTAGCAAGTTAA     94
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
4_3             701 GCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAGGGCCTAGCAAGTTAA    750

refm2s9          95 AATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCGGCGGGTACAGCAGCG    144
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
4_3             751 AATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCGGCGGGTACAGCAGCG    800
```

```
refm2s9          145 ATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTTT--    192
                     ||||||||||||||||||||||||||||||||||||||||||||||||||
4_3              801 ATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTTTCT    850
```

SEQ_ID: 017
Name: pENTR.hU6.sgRNA.CMV.SG12_mutated
Clone: 7
Service: GATC
Primer: KAN2-RP
Direction: Reverse

```
M2S12_c7         551 AAAGGACGAaACACCGGGAAAGGTCGAGAAACT-CAAAGTTTTAGAGCTA    599
                         ||||||||||||||||||| ||||||||||||||||||
refm2s12           1 --------------GGGAAAGGTCGAGAAACTGCAAAGTTTTAGAGCTA     35

M2S12_c7         600 GgcCAGGCCGCCGCCGCGTACAGCAGCGATGCGCGGCGGCGGCCCTGCAG    649
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
refm2s12          36 GGCCAGGCCGCCGCCGCGTACAGCAGCGATGCGCGGCGGCGGCCCTGCAG     85

M2S12_c7         650 GGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCG    699
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
refm2s12          86 GGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCG    135

M2S12_c7         700 GCGGCGGGTACAGCAGCGATGCCCGCCGCCGCGCCTGCAGGGCCAAGTGG    749
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
refm2s12         136 GCGGCGGGTACAGCAGCGATGCCCGCCGCCGCGCCTGCAGGGCCAAGTGG    185

M2S12_c7         750 CACCGAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCA    799
                     ||||||||||||||||||
refm2s12         186 CACCGAGTCGGTGCTTTTT------------------------------    204
```

SEQ_ID: 018
Name: pENTR.hU6.sgRNA.CMV.SG12
Clone: 2
Service: GATC
Primer: KAN2-RP
Direction: Reverse

```
clone2           251 TACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTT    300

M2S12_ref          1 ------------------------------------------ACTTGAAA      8
                                                               ||||||||
clone2           301 TTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAA    350

M2S12_ref          9 GTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG     58
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone2           351 GTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG    400

M2S12_ref         59 GGAAAGGTCGAGAAACTGCAAAGTTTTAGAGCTAGGCCAGGCCGCCGCCG    108
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone2           401 GGAAAGGTCGAGAAACTGCAAAGTTTTAGAGCTAGGCCAGGCCGCCGCCG    450

M2S12_ref        109 CGTACAGCAGCGATGCGCGGCGGCGGCCCTGCAGGGCCTAGCAAGTTAAA    158
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone2           451 CGTACAGCAGCGATGCGCGGCGGCGGCCCTGCAGGGCCTAGCAAGTTAAA    500

M2S12_ref        159 ATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCGGCGGCGGGTACAGCAG    208
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone2           501 ATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCGGCGGCGGGTACAGCAG    550

M2S12_ref        209 CGATGCCCGCCGCCGCGCCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTT    258
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone2           551 CGATGCCCGCCGCCGCGCCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTT    600

M2S12_ref        259 TTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCATTG    308
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
clone2          601 TTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCATTG    650

M2S12_ref       309 CTTATCA-------------------------------------------    315
                    |||||||
clone2          651 CTTAtcAATTTGTTGTAACGAACAGGTCACTATCAGTCAAAATAAAATCA    700
```

SEQ_ID: 019
Name: pENTR.hU6.sgRNA.CMV.SG6M21
Clone: 2
Service: GATC
Primer: KAN2-RP
Direction: Reverse

```
M2S6M14_2       651 ATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGC    700

eBFP_sgRNA_M2     1 ------------------------------gggaaaggtcgagaaactgc     20
                                                  ||||||||||||||||||||
M2S6M14_2       701 TTTATATATTTGTGGAAAGGACGAAACACCGGGAAAGGTCGAGAAACTGC    750

eBFP_sgRNA_M2    21 aaagttttagagctaggccaggccgcgtacagctgcgatgcgcggccctg     70
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S6M14_2       751 AAAGTTTTAGAGCTAGGCCAGGCCGCGTACAGCTGCGATGCGCGGCCCTG    800

eBFP_sgRNA_M2    71 cagggcctagcaagttaaaataaggctagtccgttatcaacttggccagc    120
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S6M14_2       801 CAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAGC    850

eBFP_sgRNA_M2   121 gcgggtacagctgcgatgcccgcgcctgcagggccaagtggcaccgagtc    170
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S6M14_2       851 GCGGGTACAGCTGCGATGCCCGCGCCTGCAGGGCCAAGTGGCACCGAGTC    900

eBFP_sgRNA_M2   171 ggtgcttttt---------------------------------------    180
                    ||||||||||
M2S6M14_2       901 GGTGCTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAA    950
```

SEQ_ID: 020
Name: pENTR.hU6.sgRNA.CMV.SG9M21
Clone: 2
Service: GATC
Primer: KAN2-RP
Direction: Reverse

```
M2S9M14_ref       1 ---------------------------------------gggaaaggtc     10
                                                            ||||||||||
M2S9M14_2       601 TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGGAAAGGTC    650

M2S9M14_ref      11 gagaaactgcaaagttttagagctaggccaggccgccgcgtacagctgcg     60
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S9M14_2       651 GAGAAACTGCAAAGTTTTAGAGCTAGGCCAGGCCGCCGCGTACAGCTGCG    700

M2S9M14_ref      61 atgcgcggcggccctgcagggcctagcaagttaaaataaggctagtccgt    110
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S9M14_2       701 ATGCGCGGCGGCCCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGT    750

M2S9M14_ref     111 tatcaacttggccagcgcggcgggtacagctgcgatgcccgccgcgcctg    160
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S9M14_2       751 TATCAACTTGGCCAGCGCGGCGGGTACAGCTGCGATGCCCGCCGCGCCTG    800

M2S9M14_ref     161 cagggccaagtggcaccgagtcggtgcttttt------------------    192
                    ||||||||||||||||||||||||||||||||
M2S9M14_2       801 CAGGGCCAAGTGGCACCGAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTG    850
```

SEQ_ID: 021
Name: pENTR.hU6.sgRNA.CMV.SG12M21
Clone: 7
Service: GATC
Primer: KAN2-RP
Direction: Reverse

```
clone7           301 AAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCGATTTCTT    350

M2S12M14_ref       1 ----------------------------------GGGAAAGGTCGAGAAA     16
                                                       ||||||||||||||||
clone7           351 GGCTttATATATCTTGTGGAAAGGACGAAACACCGGGAAAGGTCGAGAAA    400

M2S12M14_ref      17 CTGCAAAGTTTTAGAGCTAGGCCAGGCCGCCGCCGCGTACAGCTGCGATG     66
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone7           401 CTGCAAAGTTTTAGAGCTAGGCCAGGCCGCCGCCGCGTACAGCTGCGATG    450

M2S12M14_ref      67 CGCGGCGGCGGCCCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGT    116
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone7           451 CGCGGCGGCGGCCCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGT    500

M2S12M14_ref     117 TATCAACTTGGCCAGCGCGGCGGCGGGTACAGCTGCGATGCCCGCCGCCG    166
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone7           501 TATCAACTTGGCCAGCGCGGCGGCGGGTACAGCTGCGATGCCCGCCGCCG    550

M2S12M14_ref     167 ------------------------------------------------     166

clone7           551 CGCCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTTTCTAGACCCAGCT    600
```

SEQ_ID: 022
Name: pENTR.hU6.sgRNA.CXCR4-#6.SG9M21
Clone: 2
Service: GATC
Primer: KAN2-RP
Direction: Reverse

```
refCXCR4_M2S9      1 -------------------------GCAGACGCGAGGAAGGAGGGCGCGT     25
                                              ||||||||||||||||||||||||||
9M_2             701 TATCTTGTGGAAAGGACGAAACACCGCAGACGCGAGGAAGGAGGGCGCGT    750

refCXCR4_M2S9     26 TTTAGAGCTAGGCCAGGCCGCCGCGTACAGCTGCGATGCGCGGCGGCCCT     75
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
9M_2             751 TTTAGAGCTAGGCCAGGCCGCCGCGTACAGCTGCGATGCGCGGCGGCCCT    800

refCXCR4_M2S9     76 GCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAG    125
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
9M_2             801 GCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAG    850

refCXCR4_M2S9    126 CGCGGCGGGTACAGCTGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCA    175
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
9M_2             851 CGCGGCGGGTACAGCTGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCA    900

refCXCR4_M2S9    176 CCGAGTCGGTGCTTTTT---------------------------------    192
                     |||||||||||||||||
9M_2             901 CCGAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATT    950
```

SEQ_ID: 023
Name: pENTR.hU6.sgRNA.CXCR4-#6.sgRNA2.0
Clone: 4
Service: GATC
Primer: KAN2-RP
Direction: Reverse

```
CXCR4_2.0          1 ----------------GCAGACGCGAGGAAGGAGGGCGCGTTTTAGAGCT     34
                                     |||||||||||||||||||||||||||||||||||
mut_20_4         851 GAAAGGACGAAACACCGCAGACGCGAGGAAGGAGGGCGCGTTTTAGAGCT    900

CXCR4_2.0         35 AGGCCAACATGAGGATCACCCATGTCTGCAGGGCCTAGCAAGTTAAAATA     84
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
mut_20_4         901 AGGCCAACATGAGGATCACCCATGTCTGCAGGGCCTAGCAAGTTAAAATA    950

CXCR4_2.0         85 AGGCTAGTCCGTTATCAACTTGGCCAACATGAGGATCACCCATGTCTGCA    134
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
mut_20_4         951 AGGCTAGTCCGTTATCAACTTGGCCAACATGAGGATCACCCATGTCTGCA   1000

CXCR4_2.0        135 GGGCCAAGTGGCACCGAGTCGGTGCTTTTT--------------------    164
                     |||||||||||||||||||||||||||||
```

```
mut_20_4        1001 GGGCCAAGTGGCACCGAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTA   1050


SEQ_ID: 024
Name: pENTR.hU6.sgRNA.CXCR4-#6.SG9
Clone: 1
Service: GATC
Primer: KAN2-RP
Direction: Reverse


clone1           551 AACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGAC    600

refCXCR4_M2S9      1 --------GCAGACGCGAGGAAGGAGGGCGCGTTTTAGAGCTAGGCCAGG     42
                         |||||||||||||||||||||||||||||||||||||||||||
clone1           601 GAAACACCGCAGACGCGAGGAAGGAGGGCGCGTTTTAGAGCTAGGCCAGG    650

refCXCR4_M2S9     43 CCGCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAGGGCCTAGCAAGTT     92
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           651 CCGCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAGGGCCTAGCAAGTT    700

refCXCR4_M2S9     93 AAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCGGCGGGTACAGCAG    142
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           701 AAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCGGCGGGTACAGCAG    750

refCXCR4_M2S9    143 CGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTTT    192
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           751 CGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTTT    800


SEQ_ID: 025
Name: pENTR.hU6.sgRNA.CXCR4-#1.SG9
Clone: 1
Service: GATC
Primer: KAN2-RP
Direction: Reverse


ref                1 ----------------GGCTAGGAACGCGTCTCTCTGGTTTTAGAGCTAG     34
                                     ||||||||||||||||||||||||||||||||||
EMBOSS_001       651 AAAGGACGAAACACCGGGCTAGGAACGCGTCTCTCTGGTTTTAGAGCTAG    700

ref               35 GCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAGGGCCTAG     84
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       701 GCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAGGGCCTAG    750

ref               85 CAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCGGCGGGTA    134
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       751 CAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCGGCGGGTA    800

ref              135 CAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGAGTCGGTG    184
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
EMBOSS_001       801 CAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGAGTCGGTG    850

ref              185 CTTTTT----------------------------------------------    190
                     ||||||
EMBOSS_001       851 CTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCA    900
```

```
SEQ_ID: 026
Name: pENTR.hU6.sgRNA.CXCR4-#2.SG9
Clone: 1
Service: GATC
Primer: KAN2-RP
Direction: Reverse
```

```
ref               1 --------------GCCTGAAGACAGGTGGGAAGCGCGTTTTAGAGCTAG     36
                                  ||||||||||||||||||||||||||||||||||
seed2           701 AGGACGAAACACCGGCCTGAAGACAGGTGGGAAGCGCGTTTTAGAGCTAG    750

ref              37 GCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAGGGCCTAG     86
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed2           751 GCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAGGGCCTAG    800

ref              87 CAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCGGCGGGTA    136
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed2           801 CAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCGGCGGGTA    850

ref             137 CAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGAGTCGGTG    186
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed2           851 CAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGAGTCGGTG    900

ref             187 CTTTTT-------------------------------------------     192
                    ||||||
seed2           901 CTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCA    950
```

```
SEQ_ID: 026
Name: pENTR.hU6.sgRNA.CXCR4-#3.SG9
Clone: 1
Service: GATC
Primer: KAN2-RP
Direction: Reverse
```

```
ref               1 ----------GAGCCGGACAGGACCTCCCAGGTTTTAGAGCTAGGCCAGG     40
                              ||||||||||||||||||||||||||||||||||||||||
seed3           801 CGAAACACCGGAGCCGGACAGGACCTCCCAGGTTTTAGAGCTAGGCCAGG    850

ref              41 CCGCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAGGGCCTAGCAAGTT     90
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed3           851 CCGCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAGGGCCTAGCAAGTT    900

ref              91 AAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCGGCGGGTACAGCAG    140
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed3           901 AAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCGGCGGGTACAGCAG    950

ref             141 CGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTTT    190
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed3           951 CGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTTT   1000
```

```
SEQ_ID: 027
Name: pENTR.hU6.sgRNA.CXCR4-#4.SG9
Clone: 2 (after re-picking)
Service: GATC
Primer: KAN2-RP
Direction: Reverse
```

```
CXCR4_4_ref       1 ----------------------------------------GCGGGTGGT      9
                                                            |||||||||
clone           551 TTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGCGGGTGGT    600

CXCR4_4_ref      10 CGGTAGTGAGTCGTTTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCGA     59
```

```
                    ||||||||||||||||||||||||||||||||||||||||||||||||||
clone             601 CGGTAGTGAGTCGTTTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCGA    650

CXCR4_4_ref        60 TGCGCGGCGGCCCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTT    109
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone             651 TGCGCGGCGGCCCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTT    700

CXCR4_4_ref       110 ATCAACTTGGCCAGCGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTGC    159
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone             701 ATCAACTTGGCCAGCGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTGC    750

CXCR4_4_ref       160 AGGGCCAAGTGGCACCGAGTCGGTGCTTTTT----------------------    190
                    |||||||||||||||||||||||||||||||
clone             751 AGGGCCAAGTGGCACCGAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTGT    800
```

SEQ_ID: 028
Name: pENTR.hU6.sgRNA.CXCR4-#5.SG9
Clone: 1
Service: GATC
Primer: KAN2-RP
Direction: Reverse

```
ref                 1 --------GGACCCTGCTGTTTGCGGGTGGTGTTTTAGAGCTAGGCCAGG     42
                      |||||||||||||||||||||||||||||||||||||||||||
seed5             751 AAACACCGGGACCCTGCTGTTTGCGGGTGGTGTTTTAGAGCTAGGCCAGG    800

ref                43 CCGCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAGGGCCTAGCAAGTT     92
                      ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed5             801 CCGCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAGGGCCTAGCAAGTT    850

ref                93 AAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCGGCGGGTACAGCAG    142
                      ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed5             851 AAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCGGCGGGTACAGCAG    900

ref               143 CGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTTT    192
                      ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed5             901 CGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTTT    950
```

SEQ_ID: 029
Name: pENTR.hU6.sgRNA.CXCR4-#7.SG9
Clone: 1
Service: GATC
Primer: KAN2-RP
Direction: Reverse

```
ref                 1 ------------------------GCAAGTCACTCCCCTTCCCTGTTTTA     26
                                              |||||||||||||||||||||||||||
seed7             551 TCTTGTGGAAAGGACGAAACACCGGCAAGTCACTCCCCTTCCCTGTTTTA    600

ref                27 GAGCTAGGCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAG     76
                      ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed7             601 GAGCTAGGCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAG    650

ref                77 GGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCG    126
                      ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed7             651 GGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCG    700

ref               127 GCGGGTACAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGA    176
                      ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed7             701 GCGGGTACAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGA    750

ref               177 GTCGGTGCTTTTT-------------------------------------    189
                      |||||||||||||
seed7             751 GTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAA    800
```

```
SEQ_ID: 030
Name: pENTR.hU6.sgRNA.CXCR4-#8.SG9
Clone: 1
Service: GATC
Primer: KAN2-RP
Direction: Reverse


ref               1 ---------------------------------GAATTCCATCCACTTTA     17
                                                      |||||||||||||||||
seed8           701 CTTTATATATCTTGTGGAAAGGACGAAACACCGGAATTCCATCCACTTTA    750

ref              18 GCAAGGAGTTTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCGATGCGC     67
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed8           751 GCAAGGAGTTTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCGATGCGC    800

ref              68 GGCGGCCCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAA    117
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed8           801 GGCGGCCCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAA    850

ref             118 CTTGGCCAGCGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTGCAGGGC    167
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed8           851 CTTGGCCAGCGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTGCAGGGC    900

ref             168 CAAGTGGCACCGAGTCGGTGCTTTTT------------------------    193
                    |||||||||||||||||||||||||
seed8           901 CAAGTGGCACCGAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTACAAA    950

SEQ_ID: 031
Name: pENTR.hU6.sgRNA.CXCR4-#9.SG9
Clone: 1
Service: GATC
Primer: KAN2-RP
Direction: Reverse



ref               1 -------------------------GCCCGCGCTTCCCACCTGTCTTCGT     25
                                              |||||||||||||||||||||||||
seed9           651 ATCTTGTGGAAAGGACGAAACACCGGCCCGCGCTTCCCACCTGTCTTCGT    700

ref              26 TTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGCCCT     75
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed9           701 TTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGCCCT    750

ref              76 GCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAG    125
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed9           751 GCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAG    800

ref             126 CGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCA    175
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed9           801 CGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCA    850

ref             176 CCGAGTCGGTGCTTTTT---------------------------------    192
                    |||||||||||||||||
seed9           851 CCGAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATT    900



SEQ_ID: 032
Name: pENTR.hU6.sgRNA.CXCR4-#10.SG9
Clone: 1
Service: GATC
Primer: KAN2-RP
Direction: Reverse


ref               1 ------------------------GCCTCTGGGAGGTCCTGTCCGGCTCG     26
                                             |||||||||||||||||||||||||||
seed10          701 TCTTGTGGAAAGGACGAAACACCGGCCTCTGGGAGGTCCTGTCCGGCTCG    750
```

```
ref              27 TTTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGCCC     76
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed10          751 TTTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGCCC    800

ref              77 TGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCA    126
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed10          801 TGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCA    850

ref             127 GCGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGC    176
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seed10          851 GCGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGC    900

ref             177 ACCGAGTCGGTGCTTTTT--------------------------------    194
                    ||||||||||||||||||
seed10          901 ACCGAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCAT    950



SEQ_ID: 033
Name: pENTR.hU6.sgRNA.ASCL1-#2.sgRNA2.0
Clone: 2
Service: SeqLab
Primer: pENTattL2-rev
Direction: Reverse

ref               1 --------------------------------ATCTTGTGGAAAGGACGAA     19
                                                    ||||||||||||||||||||
12              901 TTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAA    950

ref              20 ACACC-gcagccgctcgctgcagcagGTTTTAGAGCTAGGCCAACATGAG     68
                    ||||| ||||||||||||||||||||||||||||||||||||||||||||||
12              951 ACACCGGCAGCCGCTCGCTGCAGCAGGTTTTAGAGCTAGGCCAACATGAG   1000

ref              69 GATCACCCATGTCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTT    118
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
12             1001 GATCACCCATGTCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTT   1050

ref             119 ATCAACTTGGCCAACATGAGGATCACCCATGTCTGCAGGGCCAAGTGGCA    168
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
12             1051 ATCAACTTGGCCAACATGAGGATCACCCATGTCTGCAGGGCCAAGTGGCA   1100

ref             169 CCGAGTCGGTGCTTTTT--------------------------------    185
                    |||||||||||||||||
12             1101 CCGAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATT   1150



SEQ_ID: 034
Name: pENTR.hU6.sgRNA.ASCL1-#1.SG9
Clone: 2
Service: GATC
Primer: KAN2-RP
Direction: Reverse


9_22            851 TAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAAG    900

M2S9              1 ------------------------ATCTTGTGGAAAGGACGAAACACC-g     25
                                            |||||||||||||||||||||||| |
9_22            901 TATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGG    950

M2S9             26 gctgggtgtcccattgaaaGTTTTAGAGCTAGGCCAGGCCGCCGCGTACA     75
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
9_22            951 GCTGGGTGTCCCATTGAAAGTTTTAGAGCTAGGCCAGGCCGCCGCGTACA   1000

M2S9             76 GCAGCGATGCGCGGCGGCCCTGCAGGGCCTAGCAAGTTAAAATAAGGCTA    125
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
9_22            1001 GCAGCGATGCGCGGCGGCCCTGCAGGGCCTAGCAAGTTAAAATAAGGCTA   1050

M2S9            126 GTCCGTTATCAACTTGGCCAGCGCGGCGGGTACAGCAGCGATGCCCGCCG    175
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
9_22            1051 GTCCGTTATCAACTTGGCCAGCGCGGCGGGTACAGCAGCGATGCCCGCCG   1100
```

```
M2S9             176 CGCCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTTT------------    213
                     |||||||||||||||||||||||||||||||||||||
9_22            1101 CGCCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTTTCTAGACCCAGCT   1150
```

```
SEQ_ID: 035
Name: pENTR.hU6.sgRNA.ASCL1-#2.SG9
Clone: 3
Service: GATC
Primer: KAN2-RP
Direction: Reverse
```

```
M2S9_2_ref         1 ------------------ATCTTGTGGAAAGGACGAAACACC-gcagccg     31
                                       ||||||||||||||||||||||| |||||||
92_3             801 GATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCGGCAGCCG    850

M2S9_2_ref        32 ctcgctgcagcagGTTTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCG     81
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
92_3             851 CTCGCTGCAGCAGGTTTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCG    900

M2S9_2_ref        82 ATGCGCGGCGGCCCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGT    131
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
92_3             901 ATGCGCGGCGGCCCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGT    950

M2S9_2_ref       132 TATCAACTTGGCCAGCGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTG    181
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
92_3             951 TATCAACTTGGCCAGCGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTG   1000

M2S9_2_ref       182 CAGGGCCAAGTGGCACCGAGTCGGTGCTTTTT------------------    213
                     |||||||||||||||||||||||||||||||||
92_3            1001 CAGGGCCAAGTGGCACCGAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTG   1050
```

```
SEQ_ID: 036
Name: pENTR.hU6.sgRNA.ASCL1-#3.SG9
Clone: 1
Service: GATC
Primer: KAN2-RP
Direction: Reverse
```

```
M2S9_3_ref         1 -----ATCTTGTGGAAAGGACGAAACACC-gtgtttattcagccgggagt     44
                          ||||||||||||||||||||||| ||||||||||||||||||||
M2S9_3_c1        551 TATATATCTTGTGGAAAGGACGAAACACCGGTGTTTATTCAGCCGGGAGT    600

M2S9_3_ref        45 cGTTTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGC     94
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S9_3_c1        601 CGTTTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGC    650

M2S9_3_ref        95 CCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGC    144
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S9_3_c1        651 CCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGC    700

M2S9_3_ref       145 CAGCGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTG    194
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S9_3_c1        701 CAGCGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTG    750

M2S9_3_ref       195 GCACCGAGTCGGTGCTTTTT------------------------------    214
                     ||||||||||||||||||||
M2S9_3_c1        751 GCACCGAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGC  800
```

```
SEQ_ID: 037
Name: pENTR.hU6.sgRNA.ASCL1-#4.SG9
Clone: 1
Service: GATC
Primer: KAN2-RP
```

```
Direction: Reverse

M2S9_4_ref         1 ---------------ATCTTGTGGAAAGGACGAAACACC-gatggagagt     34
                                     ||||||||||||||||||||||| ||||||||||
M2S9_4_c1        451 TTTCTTGGCTTATATATCTTGTGGAAAGGACGAAACACCGGATGGAGAGT    500

M2S9_4_ref        35 ttgcaaggagcGTTTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCGAT     84
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S9_4_c1        501 TTGCAAGGAGCGTTTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCGAT    550

M2S9_4_ref        85 GCGCGGCGGCCCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTA    134
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S9_4_c1        551 GCGCGGCGGCCCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTA    600

M2S9_4_ref       135 TCAACTTGGCCAGCGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTGCA    184
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S9_4_c1        601 TCAACTTGGCCAGCGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTGCA    650

M2S9_4_ref       185 GGGCCAAGTGGCACCGAGTCGGTGCTTTTT--------------------    214
                     ||||||||||||||||||||||||||||||
M2S9_4_c1        651 GGGCCAAGTGGCACCGAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTA  700



SEQ_ID: 038
Name: pENTR.hU6.sgRNA.ASCL1-#5.SG9
Clone: 1
Service: GATC
Primer: KAN2-RP
Direction: Reverse

M2S9_5_ref         1 -----ATCTTGTGGAAAGGACGAAACACC-gcaggagggacaaatggctc     44
                         ||||||||||||||||||||||| ||||||||||||||||||||
M2S9_5_c1        501 TATATATCTTGTGGAAAGGACGAAACACCGGCAGGAGGGACAAATGGCTC    550

M2S9_5_ref        45 tGTTTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGC     94
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S9_5_c1        551 TGTTTTAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGC    600

M2S9_5_ref        95 CCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGC    144
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S9_5_c1        601 CCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGC    650

M2S9_5_ref       145 CAGCGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTG    194
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S9_5_c1        651 CAGCGCGGCGGGTACAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTG    700

M2S9_5_ref       195 GCACCGAGTCGGTGCTTTTT------------------------------    214
                     ||||||||||||||||||||
M2S9_5_c1        701 GCACCGAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGC    750



SEQ_ID: 039
Name: pENTR.hU6.sgRNA.ASCL1-#6.SG9
Clone: 1
Service: GATC
Primer: KAN2-RP
Direction: Reverse

M2S9_6_ref         1 -------------------------------------------------AT      2
                                                                       ||
M2S9_6_c1        651 CATATGCTTAACGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATAT    700

M2S9_6_ref         3 CTTGTGGAAAGGACGAAACACC-gagtgaggagagagagaaaacGTTTTA     51
                     ||||||||||||||||||||| ||||||||||||||||||||||||||||
M2S9_6_c1        701 CTTGTGGAAAGGACGAAACACCGGAGTGAGGAGAGAGAGAAAACGTTTTA    750

M2S9_6_ref        52 GAGCTAGGCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAG    101
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S9_6_c1        751 GAGCTAGGCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGCCCTGCAG    800

M2S9_6_ref       102 GGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCG    151
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
M2S9_6_c1        801 GGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCG    850

M2S9_6_ref       152 GCGGGTACAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGA    201
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
M2S9_6_c1        851 GCGGGTACAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGA    900

M2S9_6_ref       202 GTCGGTGCTTTTT-------------------------------------    214
                     |||||||||||||
M2S9_6_c1        901 GTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAA    950


SEQ_ID: 040
Name: pENTR.hU6.sgRNA.CMV.SG9-NM60
Clone: 3
Service: SeqLab
Primer: pENTattL2-rev
Direction: Reverse


ref_nm60           1 ------------------------------gggaaaggtcgagaaactgca     21
                                                   ||||||||||||||||||||
NM60_3           851 TATATATCTTGTGGAAAGGACGAAACACCGGGAAAGGTCGAGAAACTGCA    900

ref_nm60          22 aagttttagagctaggccaggccgccgcttcagcagcgaggcggcggccc     71
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
NM60_3           901 AAGTTTTAGAGCTAGGCCAGGCCGCCGCTTCAGCAGCGAGGCGGCGGCCC    950


ref_nm60          72 tgcagggcctagcaagttaaaataaggctagtccgttatcaacttggcca    121
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
NM60_3           951 TGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCA   1000

ref_nm60         122 gcgcggcggttcagcagcgagccgccgcgcctgcagggccaagtggcacc    171
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
NM60_3          1001 GCGCGGCGGTTCAGCAGCGAGCCGCCGCGCCTGCAGGGCCAAGTGGCACC   1050

ref_nm60         172 gagtcggtgcttttt-----------------------------------    186
                     ||||||||||||||||
NM60_3          1051 GAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTAT   1100




SEQ_ID: 041
Name: pENTR.hU6.sgRNA.CMV.SG9M21_TL
Clone: 1
Service: SeqLab
Primer: pENTattL2-rev
Direction: Reverse


ref_m14a           1 --------------gggaaaggtcgagaaactgcaaagttttagagctag     36
                                   ||||||||||||||||||||||||||||||||||||
c1               851 AAGGACGAAACACCGGGAAAGGTCGAGAAACTGCAAAGTTTTAGAGCTAG    900

ref_m14a          37 gccaggccgccgcgtacagctgcgatgcgcggcggccctgcagggcctag     86
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c1               901 GCCAGGCCGCCGCGTACAGCTGCGATGCGCGGCGGCCCTGCAGGGCCTAG    950

ref_m14a          87 caagttaaaataaggctagtccgttatcaacttggccagcgcggcgggta    136
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c1               951 CAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCGCGGCGGGTA   1000

ref_m14a         137 cagcagcgatgcccgccgcgcctgcagggccaagtggcaccgagtcggtg    186
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c1              1001 CAGCAGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACCGAGTCGGTG   1050

ref_m14a         187 cttttt--------------------------------------------    192
                     ||||||
c1              1051 CTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTATAAGAAAGCA   1100
```

```
SEQ_ID: 041
Name: pENTR.hU6.sgRNA.CMV.SG9M21_SL2
Clone: 1
Service: SeqLab
Primer: pENTattL2-rev
Direction: Reverse


ref_m14b           1 ----------------------gggaaaggtcgagaaactgcaaagttt     27
                                            ||||||||||||||||||||||||||
c3               901 TCTTGTGGAAAGGACGAAACACCGGGAAAGGTCGAGAAACTGCAAAGTTT    950

ref_m14b          28 tagagctaggccaggccgccgcgtacagcagcgatgcgcggcggccctgc     77
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3               951 TAGAGCTAGGCCAGGCCGCCGCGTACAGCAGCGATGCGCGGCGGCCCTGC   1000

ref_m14b          78 agggcctagcaagttaaaataaggctagtccgttatcaacttggccagcg    127
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3              1001 AGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCCAGCG   1050

ref_m14b         128 cggcgggtacagctgcgatgcccgccgcgcctgcagggccaagtggcacc    177
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3              1051 CGGCGGGTACAGCTGCGATGCCCGCCGCGCCTGCAGGGCCAAGTGGCACC   1100

ref_m14b         178 gagtcggtgcttttt-----------------------------------    192
                     ||||||||||||||||
c3              1101 GAGTCGGTGCTTTTTCTAGACCCAGCTTTCTTGTACAAAGTTGGCATTAT   1150


SEQ_ID: 042
Name: pB-UniSAM-PAL
Clone: 3
Service: GATC
Primer: NPHPTM_seq_1_fwd
Direction: Forward


ref             3851 CTCGACGTCACCGAGCGGGTCGAGCGCGATCAGCAGTTGGAGCAGCTGGC   3900
                                 |||| ..||||||||||||||||||||||||||||
c3_fwd1            1 -----------------GGTc-nncGCGATcAGCAGTTGGAGCAGCTGGC     32

ref             3901 GTCGGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATC   3950
                     |||||||||||||||||||||||||||||| ||||||||||||||||||||
c3_fwd1           33 GTCGGAGGGCAGAGGAAGTCTGCTAACATGC-GTGACGTCGAGGAGAATC     81

ref             3951 CTGGCCCAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAG   4000
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd1           82 CTGGCCCAATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAG    131

ref             4001 GAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGA   4050
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd1          132 GAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGA    181

ref             4051 GTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGA   4100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd1          182 GTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGA    231

ref             4101 CCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGAC   4150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd1          232 CCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGAC    281

ref             4151 ATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCC   4200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd1          282 ATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCC    331

ref             4201 CGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGT   4250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd1          332 CGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGT    381

ref             4251 GGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAG   4300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
c3_fwd1          382 GGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAG    431

ref             4301 GACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGG   4350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd1          432 GACTCCTCCCTGCAGGACGGCGAGTTCATCTACAAGGTGAAGCTGCGCGG    481

ref             4351 CACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCT   4400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd1          482 CACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCT    531

ref             4401 GGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGC   4450
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd1          532 GGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGC    581

ref             4451 GAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGA   4500
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd1          582 GAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGA    631

ref             4501 GGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCT   4550
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd1          632 GGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCT    681

ref             4551 ACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACC   4600
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd1          682 ACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACC    731

ref             4601 ATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCAT   4650
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd1          732 ATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCAT    781

ref             4651 GGACGAGCTGTACAAGtaaTGATAAAAGCGGCCGCAATAAAATATCTTTA   4700
                     ||||||||||||||||||.||||||||||||||||||||||||||||||||||
c3_fwd1          782 GGACGAGCTGTACAAGTAGTGATAAAAGCGGCCGCAATAAAATATCTTTA    831

ref             4701 TTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAACGCCAGCAACGCGG   4750
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd1          832 TTTTCATTACATCTGTGTGTTGGTTTTTTGTGTGAACGCCAGCAACGCGG    881
```

SEQ_ID: 043
Name: pB-UniSAM-PAL
Clone: 3
Service: GATC
Primer: NPHPTM_seq_2_fwd
Direction: Forward

```
ref             2301 CTAATGGGCTGTCCGGAGATGAAGACTTCTCAAGCATCGCTGATATGGAC   2350
                                                              |||||
c3_fwd2            1 ---------------------------------------------atGGa-      5

ref             2351 TTTAGTGCCCTGCTGTCACAGATTTCCTCTAGTGGGCAGGGAGGAGGTGG   2400
                     ||||||| ||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2            6 tTTAGTG-CCTGCTGTCAcAGATTTCCTCTAGTGGGCAGGGAGGAGGTGG     54

ref             2401 AAGCGGCTTCAGCGTGGACACCAGTGCCCTGCTGGACCTGTTCAGCCCCT   2450
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2           55 AAGCGGCTTCAGCGTGgACACCAGTGCCCTGCTGGACCTGTTCAGCCCCT    104

ref             2451 CGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGACAGCAGCCTGGCC   2500
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2          105 CGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGACAGCAGCCTGGCC    154

ref             2501 AGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCAGGCCTCCCGAGGC   2550
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2          155 AGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCAGGCCTCCCGAGGC    204

ref             2551 AGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCACTACACAGCGC   2600
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2          205 AGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCACTACACAGCGC    254

ref             2601 AGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCGGGAGCAACGAC   2650
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2          255 AGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCGGGAGCAACGAC    304
```

```
ref            2651 CTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTCTCCGAAGGGGA   2700
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2         305 CTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTCTCCGAAGGGGA    354

ref            2701 CGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGGCTCGGAGCCTC   2750
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2         355 CGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGGCTCGGAGCCTC    404

ref            2751 CCAAAGCCAAGGACCCCACTGTCTCCggtggttcgggtggtggttcgggt   2800
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2         405 CCAAAGCCAAGGACCCCACTGTCTCCGGTGGTTCGGGTGGTGGTTCGGGT    454

ref            2801 ggtggttcgggtAAGGTGAACCGGCCCGCGGAGCGCGCCTCGTTCGGTTC   2850
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2         455 GGTGGTTCGGGTAAGGTGAACCGGCCCGCGGAGCGCGCCTCGTTCGGTTC    504

ref            2851 GTTCGTGCTCGACGCCGGGTCCGCCCGGTTCGTCGGCTCCGACGAGTTGG   2900
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2         505 GTTCGTGCTCGACGCCGGGTCCGCCCGGTTCGTCGGCTCCGACGAGTTGG    554

ref            2901 CGCTGGTGCTTGGCTTCGCCCCGGGTGATGTGGTGCTGACGCCGGCCGTG   2950
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2         555 CGCTGGTGCTTGGCTTCGCCCCGGGTGATGTGGTGCTGACGCCGGCCGTG    604

ref            2951 GTGCTGGCCCACCTGCATCCGGACGATCGGCTGGAATGGCAGGCCGGTCT   3000
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2         605 GTGCTGGCCCACCTGCATCCGGACGATCGGCTGGAATGGCAGGCCGGTCT    654

ref            3001 GCAGCGATGCCTGGCCACCGGGCGACCCGTGGTGGTCAACCACCTGCTGC   3050
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2         655 GCAGCGATGCCTGGCCACCGGGCGACCCGTGGTGGTCAACCACCTGCTGC    704

ref            3051 TGACCGCGGAGGCCGAGCCGCGCCCCGCGATGACGACGCTGACCGCGCTG   3100
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2         705 TGACCGCGGAGGCCGAGCCGCGCCCCGCGATGACGACGCTGACCGCGCTG    754

ref            3101 ACCGAGCAGGACCGGGTGCGAGCGGTCACCGGTGTGATCACCGACCTGAG   3150
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2         755 ACCGAGCAGGACCGGGTGCGAGCGGTCACCGGTGTGATCACCGACCTGAG    804

ref            3151 CGACCGCGTCCGGCGGGCGACCGAGGCCGAGATCCGCCAGGCGGTCCGGG   3200
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2         805 CGACCGCGTCCGGCGGGCGACCGAGGCCGAGATCCGCCAGGCGGTCCGGG    854

ref            3201 CGGCGGCCGCGACCCGCAGCGAGATCGACCAGGCCAAGGGCATCGTGATG   3250
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2         855 CGGCGGCCGCGACCCGCAGCGAGATCGACCAGGCCAAGGGCATCGTGATG    904

ref            3251 GCCGCCTTCGACGTCGACGCCGACCAGGCGTTTGCCCTGCTCAAGTGGCA   3300
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2         905 GCCGCCTTCGACGTCGACGCCGACCAGGCGTTTGCCCTGCTCAAGTGGCA    954

ref            3301 CTCCTCGCAGAGCAACCGCAAGCTGCGGGACCTGGCCACCGGGATGATCG   3350
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2         955 CTCCTCGCAGAGCAACCGCAAGCTGCGGGACCTGGCCACCGGGATGATCG   1004

ref            3351 AGGGCTTGGCAGCGGCCAACTCCGCGTTGCCACTGCGGCGGCGGCTGTCC   3400
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2        1005 AGGGCTTGGCAGCGGCCAACTCCGCGTTGCCACTGCGGCGGCgCTGTCC   1054

ref            3401 ACCGTCTTCACCGACATGGGTTGCCCGGCGCCCTCCACCAAGGGCTGGAC   3450
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2        1055 ACCGTCTTCaCcGACATGGGTTGCCcGGCGCCCTCCAccAAGGGCTGGAC   1104

ref            3451 CGTGCCGGTCACCGACATCGGCCTGCCGCCCACGTCGGGGCTGATCCCGA   3500
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_fwd2        1105 CGTGCCGGTCACCGAcATCGGCCTGccGCCcACGTCGGGGCTGATCCCGA   1154

ref            3501 CCGCGTTGCTCCCGGGCATCCTGACCCGGGCCGCGCACG-ACGCATCGGT   3549
                    ||||||||||||||||||||||||||||||||||...|..| |..|||||||
c3_fwd2        1155 CCGcgTtGCTCCcGGGCATCCTGACCCGGgCCnnncnngAAnncATCGGT   1204

ref            3550 TGCCATCACCGTCGCCGACGTCACCGCACCGGACCAGCCGCTGGTCTACG   3599
                    |||||||.||||||||||.||.
c3_fwd2        1205 TGCCATcncCGTcgccGAagtt----------------------------   1226
```

```
SEQ_ID: 044
Name: pB-UniSAM-PAL
Clone: 3
Service: GATC
Primer: NPHPTM_seq_1_rev
Direction: Reverse


ref             751 GTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGG    800
                                                    ||||||||||||
c3_rev2           1 ------------------------------------tgCTGGTggtGG     12

ref             801 CCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTG    850
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2          13 CCAAAgtgGaaAAgGGCAAGtCCAAGaAACtgaAGAGTgTGAAAGAGctg     62

ref             851 CTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGA    900
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2          63 CTGGGGATCAccaTCATGGaAAgAAGCAGCTTCGAGAAGAATCCCatCGA    112

ref             901 CTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCA    950
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         113 CttTctGGaAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCA    162

ref             951 AGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATG   1000
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         163 AGCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATG    212

ref            1001 CTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTC   1050
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         213 CTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTC    262

ref            1051 CAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGG   1100
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         263 CAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGG    312

ref            1101 GCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAG   1150
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         313 GCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAG    362

ref            1151 CACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGT   1200
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         363 CACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGT    412

ref            1201 GATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGC   1250
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         413 GATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGC    462

ref            1251 ACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTT   1300
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         463 ACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTT    512

ref            1301 ACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCAC   1350
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         513 ACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCAC    562

ref            1351 CATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCC   1400
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         563 CATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCC    612

ref            1401 TGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCT   1450
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         613 TGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCT    662

ref            1451 CAGCTGGGAGGCGACAGCGCTGGAGGAGGTGGAAGCGGAGGAGGAGGAAG   1500
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         663 CAGCTGGGAGGCGACAGCGCTGGAGGAGGTGGAAGCGGAGGAGGAGGAAG    712

ref            1501 CGGAGGAGGAGGTAGCGGACCTAAGAAAAAGAGGAAGGTGGCGGCCGCTG   1550
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
c3_rev2          713 CGGAGGAGGAGGTAGCGGACCTAAGAAAAAGAGGAAGGTGGCGGCCGCTG    762

ref             1551 GATCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGA   1600
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2          763 GATCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGA    812

ref             1601 AGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCT   1650
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2          813 AGTGACGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCT    862

ref             1651 TGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATTTCG   1700
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2          863 TGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTTGATGATTTCG    912

ref             1701 ACCTGGACATGCTGATTAACGGAGGAAGCGGAGCTACTAACTTCAGCCTG   1750
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2          913 ACCTGGACATGCTGATTAACGGAGGAAGCGGAGCTACTAACTTCAGCCTG    962

ref             1751 CTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACCTGGAAGCGGAat   1800
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2          963 CTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACCTGGAAGCGGAAT   1012

ref             1801 gcctaagaaaaagaggaaggtggcggccgctggatccCCTTCAGGGCAGA   1850
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         1013 GCCTAAGAAAAAGAGGAAGGTGGCGGCCGCTGGATCCCCTTCAGGGCAGA   1062

ref             1851 TCAGCAACCAGGCCCTGGCTCTGGCCCCTAGCTCCGCTCCAGTGCTGGCC   1900
                     ||||||||||||||||||||||||||||||||.||||.|||
c3_rev2         1063 TCAGCAACCAGGCCCTGGCtctGGcCCCtAgntCCgntcc----------   1102
```

SEQ_ID: 045
Name: pB-UniSAM-PAL
Clone: 3
Service: GATC
Primer: NPHPTM_seq_2_rev
Direction: Reverse

```
ref             2001 TGCCCAAGTCTACACAGGCCGGCGAGGGGACTCTGAGTGAAGCTCTGCTG   2050
                                                                    |||
c3_rev2            1 ----------------------------------------------ctg      3

ref             2051 CACCTGCAGTTCGACGCTGATGAGGACCTGGGAGCTCTGCTGGGGAACAG   2100
                     |||..|||||||||||||||||||||...|||||...||||||||||||
c3_rev2            4 CAcnngCAGTTCGACgCTGATGAgGACcnnnGAgctnnnCTGGGGAACAG     53

ref             2101 CACCGATCCCGGAGTGTTCACAGATCTGGCCTCCGTGGACAACTCTGAGT   2150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2           54 CACcGATCCCGGAGTGTTCACAGATCTGGCCTCCGTGGACAACTCTGAGT    103

ref             2151 TTCAGCAGCTGCTGAATCAGGGCGTGTCCATGTCTCATAGTACAGCCGAA   2200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2          104 TTCAGCAGCTGCTGAATCAGGGCgTGTCCATGTCTCATAGTACAGCCGAA    153

ref             2201 CCAATGCTGATGGAGTACCCCGAAGCCATTACCCGGCTGGTGACCGGCAG   2250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2          154 CCAATGCTGATGGAGTACCCCGAAGCCATTACCCGGCTGGTGACCGGCAG    203

ref             2251 CCAGCGGCCCCCCGACCCCGCTCCAACTCCCCTGGGAACCAGCGGCCTGC   2300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2          204 CCAGCGGCCCCCCGACCCCGCTCCAACTCCCCTGGGAACCAGCGGCCTGC    253

ref             2301 CTAATGGGCTGTCCGGAGATGAAGACTTCTCAAGCATCGCTGATATGGAC   2350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2          254 CTAATGGGCTGTCCGGAGATGAAGACTTCTCAAGCATCGCTGATATGGAC    303

ref             2351 TTTAGTGCCCTGCTGTCACAGATTTCCTCTAGTGGGCAGGGAGGAGGTGG   2400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2          304 TTTAGTGCCCTGCTGTCACAGATTTCCTCTAGTGGGCAGGGAGGAGGTGG    353
```

```
ref            2401 AAGCGGCTTCAGCGTGGACACCAGTGCCCTGCTGGACCTGTTCAGCCCCT   2450
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         354 AAGCGGCTTCAGCGTGGACACCAGTGCCCTGCTGGACCTGTTCAGCCCCT    403

ref            2451 CGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGACAGCAGCCTGGCC   2500
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         404 CGGTGACCGTGCCCGACATGAGCCTGCCTGACCTTGACAGCAGCCTGGCC    453

ref            2501 AGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCAGGCCTCCCGAGGC   2550
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         454 AGTATCCAAGAGCTCCTGTCTCCCCAGGAGCCCCCCAGGCCTCCCGAGGC    503

ref            2551 AGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCACTACACAGCGC   2600
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         504 AGAGAACAGCAGCCCGGATTCAGGGAAGCAGCTGGTGCACTACACAGCGC    553

ref            2601 AGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCGGGAGCAACGAC   2650
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         554 AGCCGCTGTTCCTGCTGGACCCCGGCTCCGTGGACACCGGGAGCAACGAC    603

ref            2651 CTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTCTCCGAAGGGGA   2700
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         604 CTGCCGGTGCTGTTTGAGCTGGGAGAGGGCTCCTACTTCTCCGAAGGGGA    653

ref            2701 CGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGGCTCGGAGCCTC   2750
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
c3_rev2         654 CGGCTTCGCCGAGGACCCCACCATCTCCCTGCTGACAGGCTCGGAGCCTC    703

ref            2751 CCAAAGCCAAGGA-CCCCACTGTCT-CCggtggttcgggtggtggttcgg   2798
                    ||||||||||||| |||||||||||| ||||| ||||
c3_rev2         704 CCAAAGCCAAGGACCCCCACtGtcTccCGGt-gttc--------------    738
```

```
SEQ_ID: 046
Name: pB-UniSAM_empty
Clone: 1
Service: GATC
Primer: hU6_seq_fwd primer
Direction: Forward
```

```
hU6_empty         1 GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGC     50
                                    ||||.|.|||||||||||||||||||||||
seq               1 ----------------------ccTTnanaTTTGCATATACGATACAAGGC     29

hU6_empty        51 TGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAG    100
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              30 TGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAG     79

hU6_empty       101 TACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTT    150
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              80 TACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTT    129

hU6_empty       151 TTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAA    200
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq             130 TTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAA    179

hU6_empty       201 GTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACAGAA    250
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq             180 GTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACAGAA    229

hU6_empty       251 TTCTTTAATTAAGCTAGCCAACAAGCTCGTCATCGCTTTGCAGAAGAGCA    300
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq             230 TTCTTTAATTAAGCTAGCCAACAAGCTCGTCATCGCTTTGCAGAAGAGCA    279

hU6_empty       301 GAGAGGATATGCTCATCGTCTAAAGAACTACCCATTTTATTATATATT--    348
                    ||||||||||||||||||||||||||||||||||||||||||||||||
seq             280 GAGAGGATATGCTCATCGTCTAAAGAACTACCCATTTTATTATATATTAG    329
```

```
SEQ_ID: 047
Name: pB-UniSAM-PAL_empty
Clone: 2
Service: GATC
Primer: hU6_seq_fwd primer
Direction: Forward


hU6_empty          1 GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGC     50
                                 |||  |||.|.|||||||||||||||||||||||
seq                1 -----------------gaT--CTTnanaTTTGCATATACGATACAAGGC     31

hU6_empty         51 TGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAG    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq               32 TGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAG     81

hU6_empty        101 TACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTT    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq               82 TACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTT    131

hU6_empty        151 TTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAA    200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              132 TTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAA    181

hU6_empty        201 GTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACAGAA    250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              182 GTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACAGAA    231

hU6_empty        251 TTCTTTAATTAAGCTAGCCAACAAGCTCGTCATCGCTTTGCAGAAGAGCA    300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              232 TTCTTTAATTAAGCTAGCCAACAAGCTCGTCATCGCTTTGCAGAAGAGCA    281

hU6_empty        301 GAGAGGATATGCTCATCGTCTAAAGAACTACCCATTTTATTATATATT--    348
                     ||||||||||||||||||||||||||||||||||||||||||||||||
seq              282 GAGAGGATATGCTCATCGTCTAAAGAACTACCCATTTTATTATATATTAG    331



SEQ_ID: 048
Name: pB-UniSAM-PAL_SG6
Clone: 2
Service: GATC
Primer: hU6_seq_fwd primer
Direction: Forward



s6ref              1 ---------------------------------GGGAAAGGTCGAGAAAC     17
                                                       |||||||||||||||||
seq              201 TGGCTTTATATATCTTGTGGAAAGGACGAAACAGGGAAAGGTCGAGAAAC    250

s6ref             18 TGCAAAGTTTTAGAGCTAGGCCAGGCCGCGTACAGCAGCGATGCGCGGCC     67
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              251 TGCAAAGTTTTAGAGCTAGGCCAGGCCGCGTACAGCAGCGATGCGCGGCC    300

s6ref             68 CTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCC    117
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              301 CTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGGCC    350

s6ref            118 AGCGCGGGTACAGCAGCGATGCCCGCGCCTGCAGGGCCAAGTGGCACCGA    167
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              351 AGCGCGGGTACAGCAGCGATGCCCGCGCCTGCAGGGCCAAGTGGCACCGA    400

s6ref            168 GTCGGTGCTTTTT-------------------------------------    180
                     |||||||||||||
seq              401 GTCGGTGCTTTTTGAATTCTTTAATTAAGCTAGCCAACAAGCTCGTCATC    450



SEQ_ID: 049
Name: pB-UniSAM-PAL_SG12
```

```
Clone: 5
Service: GATC
Primer: hU6_seq_fwd primer
Direction: Forward


ref_s12            1 --------------------------------GGGAAAGGTCGAGAAACT     18
                                                      ||||||||||||||||||
s12_5            201 GGCTTTATATATCTTGTGGAAAGGACGAAACAGGGAAAGGTCGAGAAACT    250

ref_s12           19 GCAAAGTTTTAGAGCTAGGCCAGGCCGCCGCCGCGTACAGCAGCGATGCG     68
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
s12_5            251 GCAAAGTTTTAGAGCTAGGCCAGGCCGCCGCCGCGTACAGCAGCGATGCG    300

ref_s12           69 CGGCGGCGGCCCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTA    118
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
s12_5            301 CGGCGGCGGCCCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTA    350

ref_s12          119 TCAACTTGGCCAGCGCGGCGGCGGGTACAGCAGCGATGCCCGCCGCCGCG    168
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
s12_5            351 TCAACTTGGCCAGCGCGGCGGCGGGTACAGCAGCGATGCCCGCCGCCGCG    400

ref_s12          169 CCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTT---------------     203
                     ||||||||||||||||||||||||||||||||||
s12_5            401 CCTGCAGGGCCAAGTGGCACCGAGTCGGTGCTTTTTGAATTCTTTAATTA    450



SEQ_ID: 050
Name: pB-UniSAM-sgRNA2.0
Clone: 1
Service: GATC
Primer: hU6_seq_fwd primer
Direction: Forward


ref                1 GAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACAAGGC     50
                           |||  |||| ||||||||||||||||||||||||
seq                1 -----------------gat--CTTc-taTTTGCATATACGATACAAGGC     30

ref               51 TGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAG    100
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq               31 TGTTAGAGAGATAATTGGAATTAATTTGACTGTAAACACAAAGATATTAG     80

ref              101 TACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTT    150
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq               81 TACAAAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTT    130

ref              151 TTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAA    200
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              131 TTAAAATTATGTTTTAAAATGGACTATCATATGCTTACCGTAACTTGAAA    180

ref              201 GTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACAGGG    250
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              181 GTATTTCGATTTCTTGGCTTTATATATCTTGTGGAAAGGACGAAACAGGG    230

ref              251 AAAGGTCGAGAAACTGCAAAGTTTTAGAGCTAGGCCAACATGAGGATCAC    300
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              231 AAAGGTCGAGAAACTGCAAAGTTTTAGAGCTAGGCCAACATGAGGATCAC    280

ref              301 CCATGTCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC    350
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              281 CCATGTCTGCAGGGCCTAGCAAGTTAAAATAAGGCTAGTCCGTTATCAAC    330

ref              351 TTGGCCAACATGAGGATCACCCATGTCTGCAGGGCCAAGTGGCACCGAGT    400
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
seq              331 TTGGCCAACATGAGGATCACCCATGTCTGCAGGGCCAAGTGGCACCGAGT    380

ref              401 CGGTGCTTTTT---TT---------------------------------    413
                     .||||||||||   ||
seq              381 tGGTGCTTTTTGAATTCTTTAATTAAGCTAGCCAACAAGCTCGTCATCGC    430
```

```
SEQ_ID: 051
Name: pmCherryPAL_L192K
Clone: 1
Service: SeqLab
Primer: SV40-pA-rev
Direction: Reverse


KL_ref             1 ----------------------------------------------ATG      3
                                                                   |||
clone1            51 AGCTGTACAAGTCCGGACTCAGATTTCGAGCTCAAGCTTCGAATTCTATG    100

KL_ref             4 AAGGTGAACCGGCCCGCGGAGCGCGCCTCGTTCGGTTCGTTCGTGCTCGA     53
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           101 AAGGTGAACCGGCCCGCGGAGCGCGCCTCGTTCGGTTCGTTCGTGCTCGA    150

KL_ref            54 CGCCGGGTCCGCCCGGTTCGTCGGCTCCGACGAGTTGGCGCTGGTGCTTG    103
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           151 CGCCGGGTCCGCCCGGTTCGTCGGCTCCGACGAGTTGGCGCTGGTGCTTG    200

KL_ref           104 GCTTCGCCCCGGGTGATGTGGTGCTGACGCCGGCCGTGGTGCTGGCCCAC    153
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           201 GCTTCGCCCCGGGTGATGTGGTGCTGACGCCGGCCGTGGTGCTGGCCCAC    250

KL_ref           154 CTGCATCCGGACGATCGGCTGGAATGGCAGGCCGGTCTGCAGCGATGCCT    203
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           251 CTGCATCCGGACGATCGGCTGGAATGGCAGGCCGGTCTGCAGCGATGCCT    300

KL_ref           204 GGCCACCGGGCGACCCGTGGTGGTCAACCACCTGCTGCTGACCGCGGAGG    253
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           301 GGCCACCGGGCGACCCGTGGTGGTCAACCACCTGCTGCTGACCGCGGAGG    350

KL_ref           254 CCGAGCCGCGCCCCGCGATGACGACGCTGACCGCGCTGACCGAGCAGGAC    303
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           351 CCGAGCCGCGCCCCGCGATGACGACGCTGACCGCGCTGACCGAGCAGGAC    400

KL_ref           304 CGGGTGCGAGCGGTCACCGGTGTGATCACCGACCTGAGCGACCGCGTCCG    353
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           401 CGGGTGCGAGCGGTCACCGGTGTGATCACCGACCTGAGCGACCGCGTCCG    450

KL_ref           354 GCGGGCGACCGAGGCCGAGATCCGCCAGGCGGTCCGGGCGGCGGCCGCGA    403
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           451 GCGGGCGACCGAGGCCGAGATCCGCCAGGCGGTCCGGGCGGCGGCCGCGA    500

KL_ref           404 CCCGCAGCGAGATCGACCAGGCCAAGGGCATCGTGATGGCCGCCTTCGAC    453
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           501 CCCGCAGCGAGATCGACCAGGCCAAGGGCATCGTGATGGCCGCCTTCGAC    550

KL_ref           454 GTCGACGCCGACCAGGCGTTTGCCCTGCTCAAGTGGCACTCCTCGCAGAG    503
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           551 GTCGACGCCGACCAGGCGTTTGCCCTGCTCAAGTGGCACTCCTCGCAGAG    600

KL_ref           504 CAACCGCAAGCTGCGGGACCTGGCCACCGGGATGATCGAGGGCTTGGCAG    553
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           601 CAACCGCAAGCTGCGGGACCTGGCCACCGGGATGATCGAGGGCTTGGCAG    650

KL_ref           554 CGGCCAACTCCGCGTTGCCATTGCGGCGGCGGCTGTCCACCGTCTTCACC    603
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           651 CGGCCAACTCCGCGTTGCCATTGCGGCGGCGGCTGTCCACCGTCTTCACC    700

KL_ref           604 GACATGGGTTGCCCGGCGCCCTCCACCAAGGGCTGGACCGTGCCGGTCAC    653
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           701 GACATGGGTTGCCCGGCGCCCTCCACCAAGGGCTGGACCGTGCCGGTCAC    750

KL_ref           654 CGACATCGGCCTGCCGCCCACGTCGGGGCTGATCCCGACCGCGTTGCTCC    703
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           751 CGACATCGGCCTGCCGCCCACGTCGGGGCTGATCCCGACCGCGTTGCTCC    800

KL_ref           704 CGGGCATCCTGACCCGGGCCGCGCACGACGCATCGGTTGCCATCACCGTC    753
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1           801 CGGGCATCCTGACCCGGGCCGCGCACGACGCATCGGTTGCCATCACCGTC    850

KL_ref           754 GCCGACGTCACCGCACCGGACCAGCCGCTGGTCTACGCCAACCCGGCCTT    803
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
clone1          851 GCCGACGTCACCGCACCGGACCAGCCGCTGGTCTACGCCAACCCGGCCTT    900

KL_ref          804 CGAACGCCTGACCGGTTACGCCGCGGCCGAGGTGCTGGGCCGCAACTGCC    853
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1          901 CGAACGCCTGACCGGTTACGCCGCGGCCGAGGTGCTGGGCCGCAACTGCC    950

KL_ref          854 GCTTCCTGCAGGCCGAGTCCGGCGATCCGCACGAGCGGTCGGCGATTCGC    903
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1          951 GCTTCCTGCAGGCCGAGTCCGGCGATCCGCACGAGCGGTCGGCGATTCGC   1000

KL_ref          904 TCGGCGATCGCCAACGGCGACGCGGTCACCACCCTGATCCGCAACTTCCG    953
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1         1001 TCGGCGATCGCCAACGGCGACGCGGTCACCACCCTGATCCGCAACTTCCG   1050

KL_ref          954 CCAGGACGGGCACGCCTTCTGGAACGAGTTCCACCTGTCCCCCGTGCGCA   1003
                    ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1         1051 CCAGGACGGGCACGCCTTCTGGAACGAGTTCCACCTGTCCCCCGTGCGCA   1100

KL_ref          1004 ACGGGGCCGGCCGGGTCACCCATTACATCGGCTACCAGCTCGACGTCACC   1053
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1          1101 ACGGGGCCGGCCGGGTCACCCATTACATCGGCTACCAGCTCGACGTCACC   1150

KL_ref          1054 GAGCGGGTCGAGCGCGATCAGCAGTTGGAGCAGCTGGCGTCGTAG-----   1098
                     ||||||||||||||||||||||||||||||||||||||||||||||||||||
clone1          1151 GAGCGGGTCGAGCGCGATCAGCAGTTGGAGCAGCTGGCGTCGTAGGGATC   1200
```

XI.6 Python program code image to x-axis intensity converter

```
#!/usr/bin/env python
#-*- coding: utf-8 -*-

from PIL import Image
from PIL import ImageOps
im = Image.open('stripe.png').convert('L')
#im=ImageOps.invert(im) #invert black/white
width, height = im.size
results={} # results[position]=[intensities]
#print width
#print height
pixels = list(im.getdata())
matrix=im.load()

for position in range(1, width): # go through the pixels
        results[position]=[]
        for line in range(1,height):
                results[position].append(matrix[position, line])

for key,value in results.iteritems():
        print "%s,%s" %(key, sum(value) / float(len(value)))
```

XI.7 Abbreviations

Table 40: Abbreviations

Abbreviation	Explanation
°C	Degree Celsius
µg	Microgram
µl	Microliter
µm	Micrometer
µM	Micromolar
ANTAR	AmiR and NasR transcription antitermination regulators
APS	Ammonium perfulfate
AQUA	Advanced quick assembly
ARNT	Per-aryl hydrocarbon receptor nuclear translocator
bp	Base pair
Cas9	CRISPR associated protein 9
CMV	Cytomegalovirus
CRISPR	Clustered regularly interspaced short palindromic repeats
CRY2	Cryptochrome 2
CIB1	Cryptochrome-interacting basic-helix-loop-helix
CPTS	CRISPR-Cas9-based photoactivatable transcription system
cv	Column volume
dCas9	Catalytically dead Cas9
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleoside triphosphate
DSMZ	German Collection of Microorganisms and Cell Culture
DTT	Dithiothreitol
EBFP	Enhanced blue fluorescent protein
E. Coli	Escherichia Coli
EGFP	Enhanced green fluorescent protein
EtBr	Ethidium bromide
FACS	Fluorescence activated cell sorting
FCS	Fetal calf serum
FITC	Fluorescein isothiocyanate
FMN	Flavin mononucleotide
fmol	Femtomole
FSC	Forward scatter
fwd	Forward
g	Gram
gRNA	Guide RNA

HSF1	Heat shock factor 1
ICB	Intracellular buffer
IPTG	Isopropyl-β-D-thiogalactopyranosid
lncRNA	Long non-coding ribonucleic acid
LOV	Light-oxygen-voltage
M	Molar
MEM	Modified Eagle Medium
mg	Milligram
min	Minute
mL	Milliliter
mM	Millimolar
mRNA	Messenger ribonucleic acid
ncRNA	Non-coding ribonucleic acid
NEAA	Non-essential amino acids
ng	Nanogram
NLS	Nuclear localization sequence
nm	Nanometer
nt	Nucleotide
NTC	No template control
NTP	Nucleoside triphosphate
OD600	Optical density at 600nm
PAA	Polyacrylamide
PAS	Per-ARNT-Sim
PAM	Protospacer adjacent motif
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PNK	Polynucleotide Kinase
PenStrep	Penicillin/Streptomycin
rev	Reverse
RNA	Ribonucleic acid
RNAP II	RNA Polymerase II
RNAP III	RNA Polymerase III
rpm	Rounds per minute
rRNA	Ribosomal ribonucleic acid
RT	Room temperature
SAM	synergistic activation mediator
SD	Standard deviation
SDS-PAGE	Sodiumdodecylsulfate polyacrylamide gel electrophoresis
SELEX	systematic evolution of ligands by exponential enrichment’
sgRNA	Single guide RNA
snRNA	Small nuclear ribonucleic acid

SOE	Splicing by overlapping extension
SSC	Sideward scatter
T2A	Thosea asigna virus 2A
T7 RNAP	T7 RNA Polymerase
TBE	Tris/Borate/EDTA
TEMED	Tetramethylethylendiamin
TF	Transcription factor
tRNA	Transfer ribonucleic acid
ULR	Ultra Low Range
UT	Untransfected
UV	Ultraviolet
V	Volt
W	Watts

XI.8 Table of figures

XI.9 Table of supporting figures

XI.10 List of tables

XI.11 Acknowledgments

Ich danke:

Herrn Prof. Dr. Mayer, LIMES Institut der Universität Bonn, für die Vergabe des Themas und die enge Betreuung der Dissertation.

Herrn Prof. Dr. Pankratz, LIMES Institut der Universität Bonn, für die Übernahme des Koreferates.

Den Mitarbeitern der AG Mayer, AG Famulok und AG Kath-Schorr für die gemeinsame Zeit im Labor, die vielen netten und auch intensiven Gespräche und Diskussionen. Insbesondere danke ich Georg Pietruschka, Anna Maria Weber, Dr. Ankana Kakoti sowie Sebastian Pilsl für die Zusammenarbeit im „Optoribo“-Team.

Sonia Renzl und meiner Familie, ohne deren Hilfe diese Arbeit nicht möglich gewesen wäre.