



**Universidade de Brasília – UnB
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Biotecnologia e Biodiversidade**

**CONSTRUÇÃO DE INTERRUPTORES GENÉTICOS BASEADOS EM SERINA
INTEGRASES PARA CONTROLE DA EXPRESSÃO GÊNICA EM CÉLULAS DE
MAMÍFEROS**

Thais Torquato Sales

**Brasília
2024**

Universidade de Brasília – UnB
Instituto de Ciências Biológicas
Programa de Pós-Graduação em Biotecnologia e Biodiversidade

Thais Torquato Sales

**CONSTRUÇÃO DE INTERRUPTORES GENÉTICOS BASEADOS EM SERINA
INTEGRASES PARA CONTROLE DA EXPRESSÃO GÊNICA EM CÉLULAS DE
MAMÍFEROS**

Tese apresentada ao Programa de Pós-graduação
em Biotecnologia e Biodiversidade da Universidade
de Brasília como parte dos requisitos para obtenção
do título de Doutora em Biotecnologia e
Biodiversidade.

Área de Concentração: Desenvolvimento de
Produtos, Processos e Serviços Biotecnológicos.

Orientador: Prof. Dr. Elibio Leopoldo Rech Filho.

Brasília
2024

FOLHA DE APROVAÇÃO

Thais Torquato Sales

CONSTRUÇÃO DE INTERRUPTORES GENÉTICOS BASEADOS EM SERINA INTEGRASES PARA CONTROLE DA EXPRESSÃO GÊNICA EM CÉLULAS DE MAMÍFEROS

Tese apresentada ao Programa de Pós-graduação em Biotecnologia e Biodiversidade da Universidade de Brasília como parte dos requisitos para obtenção do título de Doutora em Biotecnologia e Biodiversidade.

Orientador: Prof. Dr. Elibio Leopoldo Rech Filho.

Aprovada em 19 de dezembro de 2024.

BANCA EXAMINADORA:

Prof. Dr. Elibio Leopoldo Rech Filho
(Presidente – Embrapa Recursos Genéticos e Biotecnologia)

Prof. Dr. Nicolau Brito da Cunha
(Membro 1 - Universidade de Brasília, UnB)

Profa. Dra. Gracia Maria Soares Rosinha
(Membro 2 - Embrapa Recursos Genéticos e Biotecnologia)

Profa. Dra. Danielle Biscaro Pedrolli
(Membro 3 - Universidade Estadual Paulista, UNESP)

Prof. Dr. Marcelo de Macedo Brígido
(Suplente - Universidade de Brasília, UnB)

À minha amada família e amigos cujo amor, apoio e encorajamento têm sido fundamentais em minha jornada acadêmica e *in memoriam* meu avô Domingos que sempre me chamava de Doutora desde quando iniciei a graduação em Ciências Biológicas.

AGRADECIMENTOS

Primeiramente, agradeço a DEUS pelas bênçãos, saúde e força para superar todas as dificuldades enfrentadas durante essa jornada.

À Embrapa Recursos Genéticos e Biotecnologia e à Universidade de Brasília, que não apenas me acolheram, mas também se tornaram minha segunda casa durante um período significativo da minha vida.

Ao Professor Dr. Elibio Rech, gostaria de expressar minha imensa gratidão pela oportunidade de crescimento profissional e pela orientação dedicada. Seu comprometimento, paciência e dedicação foram fundamentais para o sucesso deste trabalho.

Ao Coordenador Professor Dr. Marcelo Brígido por sua dedicação e orientação exemplares. Sua liderança foi fundamental para o meu crescimento acadêmico.

Aos professores membros da banca examinadora, Prof. Dr. Nicolau Brito da Cunha, Profa. Dra. Gracia Maria Soares Rosinha e Profa. Dra. Danielle Biscaro Pedrolli, por dedicarem seu tempo e expertise para avaliar este trabalho e fornecer sugestões valiosas para aprimorá-lo. Aos professores membros da banca de qualificação, Profa. Dra. Marcella Lemos Brettas Carneiro e Prof. Dr. Fernando Araripe G. Torres. Suas contribuições críticas e *insights* foram de grande importância para o desenvolvimento desta tese.

Aos meus colegas do Laboratório de Biologia Sintética: Mariana, Lilian, Marco, Rayane (grande apoio na minha reta final), Hermerson, Valkíria, Raquel Bonnet e Raquel Sampaio, pela colaboração, discussões estimulantes e apoio mútuo ao longo dessa jornada acadêmica. Nossas trocas de ideias e debates enriqueceram minha compreensão sobre o tema e tornaram-se uma experiência mais gratificante. Em especial, gostaria de agradecer pelos cafezinhos da Lilian, que sempre me davam um impulso extra! Vocês não apenas foram meus companheiros de pesquisa, mas também verdadeiros amigos!

À Mayna tenho um agradecimento extremamente especial. Foi uma parceira de laboratório e se tornou uma grande amiga, a mineira mais brasiliense que conheço. Gostaria de expressar minha profunda gratidão por sua colaboração constante e sua disposição incansável em compartilhar informações que foram absolutamente essenciais para o desenvolvimento deste trabalho. Sem as contribuições generosas e voluntárias dela, este estudo não teria sido possível.

À Dra. Leila e à Profa. Dra. Cíntia Coelho, minha profunda gratidão por doarem seu valioso tempo para me ajudar na interpretação de dados, análise de resultados e por compartilharem seus vastos conhecimentos em Biologia Molecular.

À Dra. Daniela e à Dra. Gracia do Laboratório de Biologia Sintética, expresso minha sincera gratidão por estarem sempre dispostas a ajudar em meu trabalho, com simpatia e carinho exemplares.

Ao Dr. Cristiano Lacorte, gostaria de agradecer por compartilhar generosamente seus conhecimentos e por emprestar valiosos instrumentos laboratoriais. Sua contribuição foi essencial para o avanço do meu trabalho.

Ao Martín e aos demais membros do laboratório do Inca e Fiocruz, gostaria de expressar minha gratidão pela colaboração nos ensaios e análise dos resultados que foram fundamentais para o sucesso deste trabalho.

Aos queridos amigos do grupo Amanita (Eriquita, Vanessinha, Dudu e Milton), minha profunda gratidão pela amizade sincera que cultivamos ao longo dos anos e pelos momentos de descontração inesquecíveis que compartilhamos juntos.

À minha amiga Deborah Bambil, gostaria de expressar minha gratidão por nossa amizade que nasceu no dia da seleção do Doutorado e tem crescido ao longo dos anos com muito carinho, empatia e confiança mútua.

Aos queridos amigos de longa data, Milene e Marcos Emmanuel (Killocan), expresso minha profunda gratidão por sempre acreditarem no meu sucesso e por estarem constantemente torcendo por mim ao longo dessa jornada.

Às queridas Mary-Ann, Chris e Simoneide do Arte Consciência, gostaria de expressar minha imensa gratidão. Vocês são mais do que parceiras Arteiras, são cientistas dedicadas e possuem um coração lindo, sempre dispostas a ajudar as pessoas. A presença de vocês e comprometimento são verdadeiramente inspiradores.

Ao Gabriel, meu amado marido, gostaria de expressar minha profunda gratidão. Sua presença constante ao meu lado, apoiando, enaltecedo e protegendo, tem sido uma fonte de força e conforto inestimável. Você tem sido um companheiro incrível, cuidando de mim em todos os aspectos. Em especial, durante as crises de ansiedade que enfrentei ao concluir este trabalho, sua compreensão foi fundamental. Sua presença e maravilhoso senso de humor foram verdadeiros fortalecedores nesta fase de conclusão do artigo/tese. Agradeço por todo o seu amor incondicional! Te amo!

A todos vocês, minha família querida, dedico essa conquista com todo o meu coração. Sei que sem o amor, o apoio e a motivação de vocês, eu não teria chegado tão longe. Cada passo que dei foi impulsionado pelo amor que compartilhamos e pela crença que vocês têm em mim. Mãe e pai, vocês têm sido meus pilares desde o primeiro dia, me encorajando a seguir meus sonhos. Minha querida irmã e cunhado, vocês me lembram que posso superar qualquer obstáculo e que sempre terei o amor e o suporte de vocês. Queridas primas e amada avó Graça, agradeço por sempre estarem ao meu lado, compartilhando alegrias e incentivando-me a buscar a excelência. Amo cada um de vocês de todo o meu ser, e serei eternamente grata pelo apoio inabalável que sempre recebi.

Ao CNPq, CAPES, FAPDF e INCT- Biosyn pelo generoso apoio financeiro fornecido.

“Na vida, não existe nada a temer, mas a entender”
(Marie Curie)

RESUMO

As serina integrases são proteínas de bacteriófagos responsáveis pela integração do DNA viral no genoma bacteriano. Essa inserção ocorre por meio de uma sequência denominada *attP* (*phago attachment site*), com o DNA bacteriano em um sítio denominado *attB* (*bacterial attachment site*). Após a recombinação mediada pela integrase, o DNA viral é incorporado ao genoma da bactéria, os sítios são mesclados, e os novos sítios *attL* (*left attachment site*) e *attR* (*right attachment site*) são formados. Quando os sítios *attB* e *attP* são posicionados em sentidos opostos flanqueando uma sequência gênica de interesse, os sítios reconhecidos pela serina integrase faz uma inversão da sequência, ou seja, realiza um giro de 180° no DNA que se encontra entre os dois sítios, resultando na posição reverso complementar à inicial. Diversas serina integrases já foram utilizadas para controlar um gene repórter em células procarióticas. No entanto, a utilização de integrases em células eucarióticas como ferramenta para regulação da expressão gênica ainda é limitada. Assim, o objetivo deste trabalho foi avaliar a funcionalidade de seis integrases no controle da expressão gênica em células de mamíferos, visando a construção de um interruptor genético. Células de fibroblastos de pele bovina foram transfetadas com as seis integrases (Int2, Int4, Int5, Int7, Int9 e Int13) em um sistema de cotransfecção plasmidial, como segue: um vetor de expressão da integrase nomeado como pIE e um segundo vetor contendo a sequência codificadora (CDS) do gene repórter *gfp* em orientação reverso complementar, flanqueado pelos sítios *attB/attP* da respectiva integrase, intitulado pSG. Após a cotransfecção, as células foram incubadas durante 48 h e acúmulo de GFP foi avaliada por meio de microscopia de fluorescência e mensurada por citometria de fluxo. A inversão da CDS foi amplificada por meio da PCR e confirmada por sequenciamento do tipo Sanger. A fluorescência da GFP foi detectada por microscopia e citometria para Int9 e Int13. Além disso, o sequenciamento evidenciou a correta inversão do *gfp* e a formação dos sítios *attL/attR* previstos para todas as integrases avaliadas.

Palavras-chave: bacteriófago, serina recombinases, circuito genético, interruptores genéticos, recombinação sítio-específica.

ABSTRACT

Serine integrases are bacteriophage proteins responsible for the integration of viral DNA into the bacterial genome. This insertion occurs via a sequence called *attP* (*phage attachment site*), with the bacterial DNA at a site called *attB* (*bacterial attachment site*). After recombination mediated by the integrase, the viral DNA is incorporated into the bacterial genome, the sites are merged, and new sites *attL* (*left attachment site*) and *attR* (*right attachment site*) are formed. When the *attB* and *attP* sites are positioned in opposite orientations flanking a target gene sequence, the sites recognized by the serine integrase perform a sequence inversion, rotating the DNA located between the two sites by 180°, resulting in a reverse complementary orientation relative to the original. Various serine integrases have already been used to control a reporter gene in prokaryotic cells. However, the use of integrases in eukaryotic cells as a tool for gene expression regulation remains limited. Thus, the aim of this study was to evaluate the functionality of six integrases in controlling gene expression in mammalian cells, aiming to construct a genetic switch. Bovine skin fibroblast cells were transfected with the six integrases (Int2, Int4, Int5, Int7, Int9, and Int13) in a plasmid cotransfection system, as follows: an integrase expression vector named pIE and a second vector containing the coding sequence (CDS) of the reporter gene *gfp* in reverse complementary orientation, flanked by *attB/attP* sites of the respective integrase, titled pSG. After cotransfection, the cells were incubated for 48 h, and GFP accumulation was assessed by fluorescence microscopy and measured by flow cytometry. The CDS inversion was amplified by PCR and confirmed by Sanger sequencing. GFP fluorescence was detected by microscopy and cytometry for Int9 and Int13. Additionally, sequencing confirmed the correct inversion of *gfp* and the formation of the expected *attL/attR* sites for all evaluated integrases.

Keywords: bacteriophage, serine recombinases, genetic circuit, genetic switches, site-specific recombination.

LISTA DE FIGURAS

Figura 1. Ciclo de vida do bacteriófago e importância das serina integrases no processo.....	14
Figura 2. Ilustração esquemática da atividade de recombinação promovida por serina integrases e os diferentes resultados possíveis.....	16
Figura 3. Árvore filogenética e representação esquemática da arquitetura de domínios das serina integrases.....	17
Figura 4. Alinhamento de sequência de aminoácidos de distintas Serina Integrases.....	19
Figura 5. Representação da estrutura tridimensional do domínio catalítico N-terminal da Integrase 13 (Int13).....	20
Figura 6. Representação esquemática dos cassetes de expressão construídos para avaliar a funcionalidade das integrases em células de fibroblastos bovino e um plasmídeo putativo resultado da atuação das integrases.....	25
Figura 7. Imagens representativas da expressão de <i>gfp</i> nos fibroblastos bovinos.....	30
Figura 8. Caracterização funcional das integrases em fibroblastos bovinos avaliadas por citometria de fluxo.....	31
Figura 9. Reação do MTT em células de fibroblasto bovino após tratamento com todas as integrases.....	33
Figura 10. Efeito das integrases sobre a viabilidade de células de fibroblasto bovino. O gráfico representa a média e o desvio padrão das células cotransfetadas com o vetor de expressão de integrase (pIE) mais o vetor repórter (pSG).....	34
Figura 11. Análise da funcionalidade das integrases quanto à formação dos sítios <i>attL/attR</i> e rotação da sequência codificadora do gene repórter <i>egfp</i>	35
Figura 12. Alinhamento dos sítios <i>attL/attR</i> obtidos por sequenciamento com os sítios de integração previstos.....	36

LISTA DE TABELAS

Tabela 1. Integrases selecionadas, código do GenBank e organismo de origem.....	23
Tabela 2. Sequências de iniciadores utilizados para se ligarem aos sítios <i>attL</i> e <i>attR</i> formados após a atividade das integrases em fibroblastos bovinos. Os valores de Tm foram calculados utilizando a ferramenta <i>OligoAnalyzer</i> do site da <i>Integrated DNA Technologies</i> (IDT) https://www.idtdna.com/calc/analyzer	28

SUMÁRIO

1. INTRODUÇÃO	12
1.1 A Família das Recombinases.....	14
1.2 Serina integrases: Mecanismo, Estrutura e Aplicações em Biotecnologia.....	15
1.3 Estrutura Tridimensional	17
1.4 Aplicações promissoras de serina integrases em sistemas eucarióticos	20
2. OBJETIVOS.....	22
2.1 Geral.....	22
2.2 Específicos	22
3. METODOLOGIA	23
3.1 Construção dos vetores	23
3.2 Cultivo celular e condições de cultura	25
3.3 Cotransfecção das células	26
3.4 Microscopia de fluorescência.....	26
3.5 Quantificação da população celular GFP positiva por citometria de fluxo	26
3.6 Avaliação da citotoxicidade das integrases por método MTT	26
3.7 Extração, amplificação e sequenciamento do DNA.....	27
3.8 Análise estatística	29
4. RESULTADOS	29
4.1 Caracterização funcional das integrases.....	29
4.2 Avaliação da citotoxicidade das integrases	32
4.3 PCR e sequenciamento dos sítios <i>attL</i> e <i>attR</i>	34
5. DISCUSSÃO E CONCLUSÃO	37
6. REFERÊNCIAS BIBLIOGRÁFICAS	39
Anexo I	47
Anexo II	48

1. INTRODUÇÃO

Os vírus são as entidades infecciosas submicroscópicas mais abundantes e diversificadas na natureza, desempenhando um papel essencial nos ecossistemas. Estima-se que os oceanos contenham aproximadamente 10^{30} partículas virais, com uma densidade média de 10^7 vírus por mililitro de água superficial do mar (Suttle, 2005; Suttle, 2007). Entre esses, os bacteriófagos, ou simplesmente fagos, destacam-se pela sua capacidade de infectar uma ampla diversidade de procariotos. Eles são encontrados em uma vasta gama de ambientes, incluindo solos férteis, águas oceânicas, corpos de água doce, o ecossistema intestinal de animais e locais extremos, como fontes hidrotermais e lagos hipersalinos (Chibani-Chennoufi *et al.*, 2004; Mills *et al.*, 2013; Alkhalil, 2023). A relevância dos fagos vai além da biologia, com suas enzimas possuindo múltiplas funcionalidades e um papel crucial no meio ambiente, influenciando o metabolismo celular e os ciclos biogeoquímicos de elementos como nitrogênio, oxigênio e carbono (Clokie *et al.*, 2011; Warwick-Dugdale *et al.*, 2019; Wang *et al.*, 2022; Bisen *et al.*, 2024). Esses vírus apresentam uma diversidade marcante em termos de tamanho, morfologia e organização genômica (Hatfull e Hendrix, 2011; Simmonds e Aiewsakun, 2018; Zhu *et al.*, 2024). São classificados de acordo com características morfológicas, tipo de material genético, locais de ocorrência e as espécies bacterianas que infectam (Hatfull e Hendrix, 2011). Todos os fagos possuem um genoma de DNA encapsulado em uma estrutura proteica de capsídeo, codificada pelo próprio fago, o que protege seu material genético e facilita a entrega à célula hospedeira seguinte (Grigson *et al.*, 2023; Turner *et al.*, 2024). O advento da microscopia eletrônica permitiu a observação detalhada de estruturas complexas desses vírus, que podem apresentar "cabeças", "caudas" e "pernas" (Figura 1A). Apesar de sua aparência complexa, os fagos são imóveis, deslocando-se em direção ao hospedeiro por movimento Browniano ou pedese (Kasman e Porter, 2023). Após bilhões de anos de coevolução com seus hospedeiros, os fagos e as bactérias desenvolveram diversos mecanismos de ataque e defesa (Bernheim e Sorek, 2018), resultando em uma constante interação de adaptação mútua (Piel *et al.*, 2022; Butt *et al.*, 2024; Patel *et al.*, 2024; Siedentop *et al.*, 2024). Na natureza, esse "conflito" gera uma grande diversidade de fagos e bactérias, cada um adaptado a ambientes ou hospedeiros específicos. Os fagos se dividem em duas principais categorias: os fagos líticos (virulentos) e os

lisogênicos (temperados) (Baaziz *et al.*, 2024), com ciclos biológicos distintos. No ciclo lítico, a replicação dos fagos leva à lise da célula hospedeira, promovendo a liberação de novos vírus que podem infectar outras bactérias, influenciando a mortalidade bacteriana e o ciclo de nutrientes. Já no ciclo lisogênico, o material genético viral (profago) é integrado ao cromossomo do hospedeiro, sendo transmitido para as células filhas durante a divisão celular. Esse processo pode proporcionar vantagens ao hospedeiro, como resistência a outros fagos ou aumento da virulência (Bailey *et al.*, 2024; Rostøl *et al.*, 2024). Em certas condições, os fagos temperados podem alternar entre o ciclo lisogênico e o lítico (Young, 2013; Fogg *et al.*, 2014; Doore e Fane, 2016), e a atividade lisogênica, embora menos comum, pode ser essencial para a propagação de genes entre bactérias. O mecanismo de integração do genoma do fago na célula bacteriana é mediado por integrases codificadas pelo próprio fago. Essas enzimas catalisam reações de recombinação altamente específicas e unidireccionais, sem necessidade de fatores adicionais (Figura 1B). A integração e a excisão do DNA viral dependem de integrases de serina de grande porte (*Large serine recombinases - LSRs*), responsáveis por inserir e excisar o DNA do cromossomo bacteriano (Groth *et al.*, 2004; Smith, 2015). Esse processo ocorre com a formação dos sítios de ancoragem *attL* e *attR*, localizados nas extremidades do DNA integrado. Diversas integrases de serina já foram extensivamente descritas e têm encontrado aplicações em sistemas biotecnológicos que exigem recombinação genética precisa.

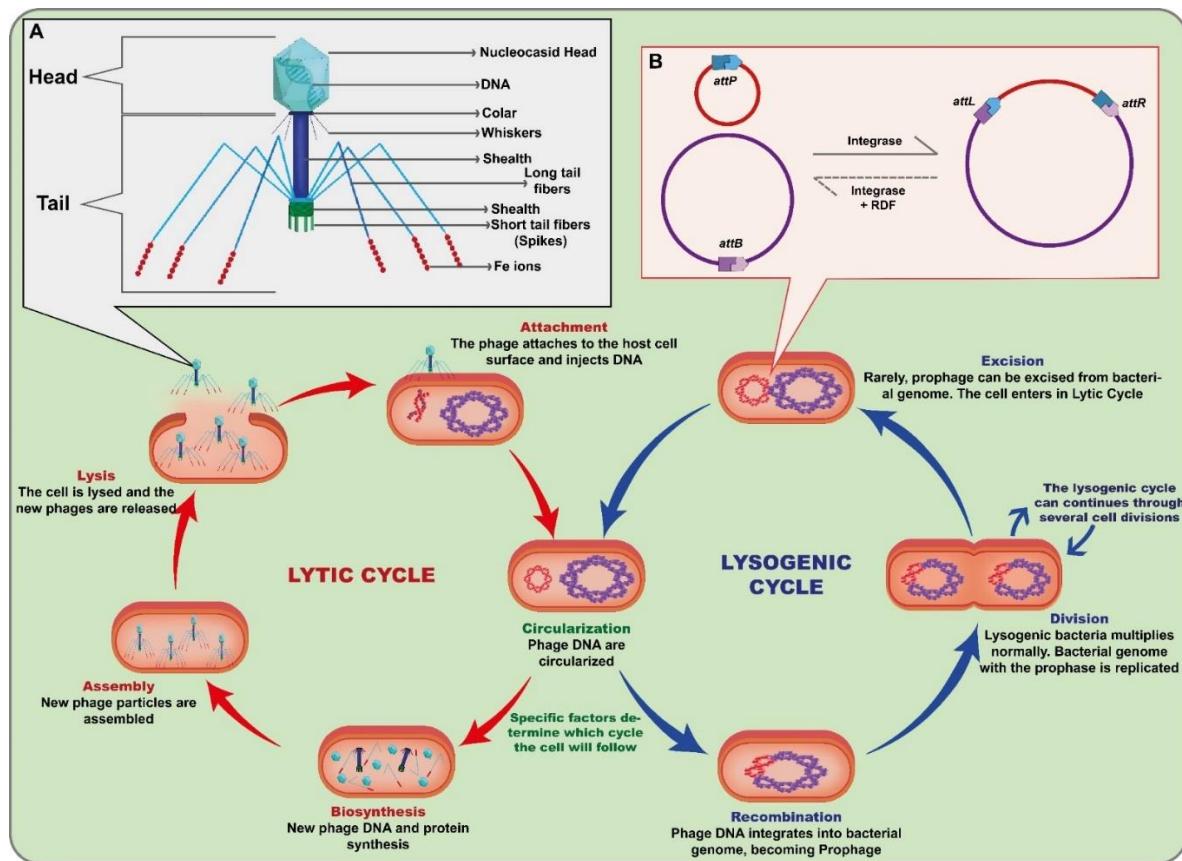


Figura 1. Ciclo de vida do bacteriófago e importância das serina integrases no processo. (A) O bacteriófago possui uma estrutura composta por uma cápsula proteica que envolve seu material genético. A cauda do fago tem a função de ligação específica e entrada do vírus na bactéria hospedeira. Ela consiste em uma estrutura tubular longa chamada bainha, que se conecta à cabeça do fago. No final da cauda, estruturas especializadas, como fibras ou fibrilas, que permitem ao fago se ligar a receptores específicos na superfície da célula hospedeira. Durante o ciclo lítico, o bacteriófago se adere à célula e injeta seu material genético, que assume o controle da maquinaria celular, resultando na produção de novos vírus e na lise da célula hospedeira, liberando os vírus recém-formados. No ciclo lisogênico, em vez de iniciar imediatamente o processo lítico, o material genético do bacteriófago integra-se ao genoma bacteriano, formando uma estrutura chamada profago. (B) Esquema de recombinação sítio-específica mediado pelo fago no hospedeiro bacteriano. A integração realiza uma recombinação precisa entre um sítio *attB* localizado no genoma bacteriano e um sítio *attP* localizado no genoma do fago. O resultado é a integração do fago no genoma do hospedeiro e a formação de sequências híbridas *attL* e *attR*.

1.1 A Família das Recombinases

A recombinação sítio-específica (SSR, do inglês *site-specific recombination*) refere-se a processos que envolvem trocas de sequências de DNA em locais específicos do genoma, sendo essenciais para diversas funções biológicas, como a mudança de fase de fatores de virulência bacterianos e a integração de fagos no genoma do hospedeiro (Grindley *et al.*, 2006; Wang *et al.*, 2011). Tais sistemas foram inicialmente descritos em bactérias e leveduras, com um papel fundamental em

processos biológicos como a regulação da virulência e a manutenção da diversidade genética (Wang *et al.*, 2011). Dependendo da configuração dos sítios de recombinação, a recombinação pode resultar em integração, excisão ou inversão do material genético (Ross e Landy, 1982; Grindley *et al.*, 2006).

A superfamília das recombinases é dividida em dois grupos principais, baseados no aminoácido ativo dentro do domínio catalítico: as tirosina (Tyr) e as serina (Ser) recombinases. Cada grupo é subdividido em famílias que variam em tamanho e mecanismo de ação (Wang *et al.*, 2011). As tirosina recombinases são amplamente distribuídas em procariotos e eucariotos, sendo responsáveis por processos de recombinação bidirecional e unidirecional, com aplicações notáveis em sistemas como Cre-lox e FLP-FRT (Grindley *et al.*, 2006). Já as serina recombinases, embora também mediadoras de recombinação unidirecional, exigem proteínas acessórias para reverter a reação, sendo cruciais para a integração do DNA viral no genoma bacteriano (Brown *et al.*, 2011).

1.2 Serina integrases: Mecanismo, Estrutura e Aplicações em Biotecnologia.

As serina integrases são enzimas de bacteriófagos responsáveis pela inserção e excisão de sequências de DNA no genoma do hospedeiro, com alta especificidade e eficiência. Estas enzimas realizam recombinação entre um sítio de fixação viral (*attP*) e um sítio bacteriano (*attB*), resultando na formação de dois sítios híbridos (*attL* e *attR*) que flanqueiam o DNA integrado. A reação de integração é essencial para a formação de um fago profago, enquanto a excisão permite a liberação do material genético viral (Groth e Calos, 2004; Smith, 2015). A simplicidade e precisão dessas reações tornam as serina integrases ferramentas promissoras em biotecnologia. Além da integração e excisão, as serina integrases também podem mediar a inversão de sequências de DNA. Esse mecanismo é particularmente útil em estratégias de engenharia genética, como a troca de cassetes de DNA (*Recombinase-Mediated Cassette Exchange - RMCE*), que permite a inserção de novas sequências em locais específicos do genoma sem causar danos à sequência original (Figura 2). Para aumentar a flexibilidade dessas reações, proteínas adicionais codificadas pelo fago, como o fator de direcionamento de recombinação (RDF), podem inverter a direção da recombinação, permitindo o controle preciso da sequência genética desejada (Figura 1B e 2) (Olorunniji *et al.*, 2016; 2017).

As serina integrases têm sido amplamente exploradas na biotecnologia e na terapia genética, oferecendo soluções inovadoras para a edição e regulação de genes. Elas são essenciais para o desenvolvimento de ferramentas moleculares capazes de manipular genomas com precisão, facilitando desde a construção de modelos de doenças até a aplicação em terapias baseadas na regulação genética. Entre as serina integrases mais estudadas estão as de fagos como phiC31, Bxb1, e R4, cujas propriedades têm sido aplicadas na modulação da expressão gênica e na criação de modelos celulares personalizados (Brown *et al.*, 2011).

Com essas avançadas possibilidades de manipulação genética, as serina integrases se destacam como elementos fundamentais na engenharia genética e na biologia sintética, ampliando o potencial de terapias inovadoras.

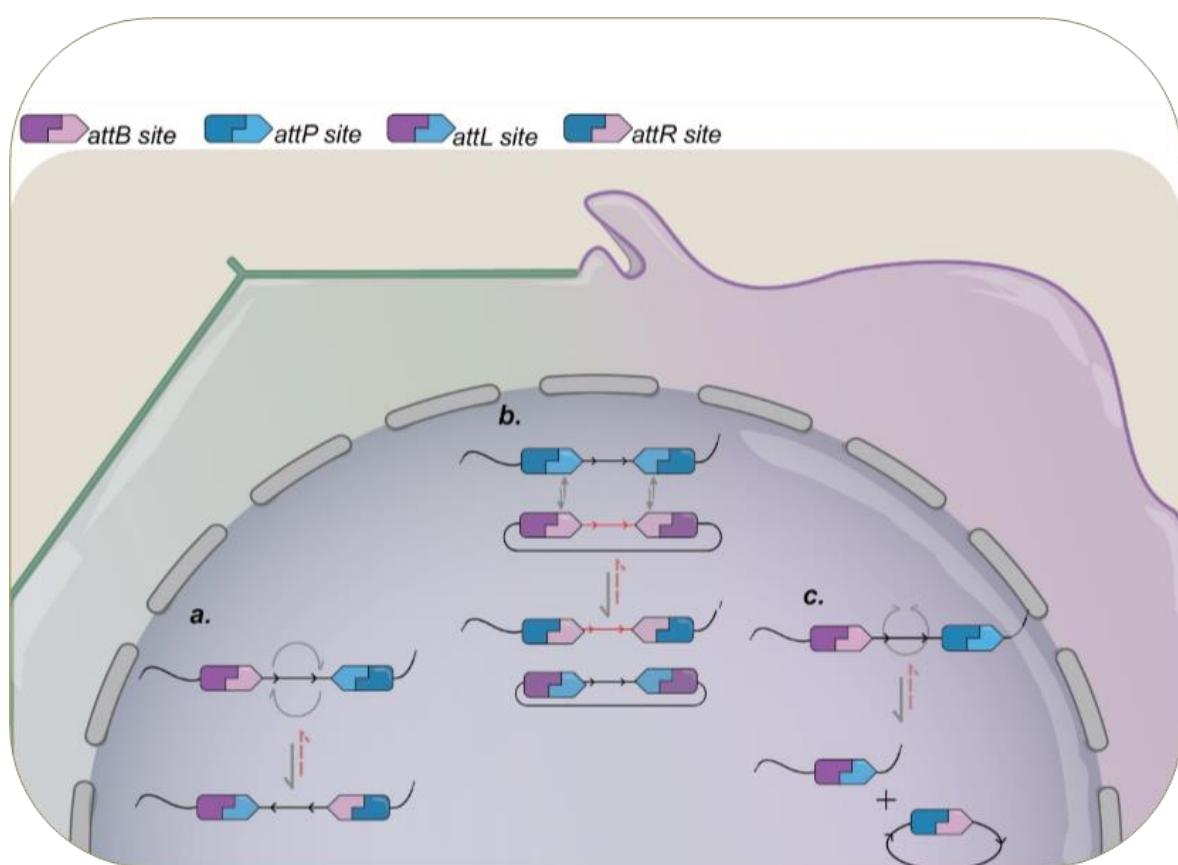


Figura 2. Ilustração esquemática da atividade de recombinação promovida por serina integrases e os diferentes resultados possíveis. Já utilizado em uma variedade de organismos, o resultado de um rearranjo por integração de serina pode ser controlado pelo design sintético de seus sítios *attB/P*. **(A)** Quando ambos os sítios apresentam a mesma molécula, mas com orientações opostas, a recombinação segue a uma inversão de 180º da sequência de DNA flanqueada por eles. **(B)** A Troca de Cassetes Mediada por Recombinase (RMCE) pode ser alcançada quando duplicatas de cada sítio são apresentadas em moléculas diferentes. Após a recombinação, as sequências de DNA flanqueadas serão trocadas entre as duas construções. **(C)** Diferente da inversão de DNA, quando ambos os sítios *att* estão presentes na mesma molécula, mas com a mesma orientação, a recombinação resultará na excisão da sequência alvo, com a formação de *attL* na molécula.

original, além de uma circular de DNA contendo o DNA excisado e o sítio *attR* formado. Em todos os casos, a ocorrência reversa (linhas pontilhadas em vermelho) é possível apenas na presença do cognato RDF.

1.3 Estrutura Tridimensional

A família das recombinases de serina é composta por uma ampla diversidade de enzimas (Grindley *et al.*, 2006; Smith, 2015). Até o momento, mais de 500 tipos de estruturas de serina integrases foram identificadas, embora muitos domínios permaneçam desconhecidos (domínio DUFF) (Wang *et al.*, 2023). No entanto, as serina integrases mais utilizadas em biotecnologia apresentam três domínios bem conhecidos: *Ser_recom* [pfam 00239]; *Recombinase* [pfam 07508] e *Zn ribbon* [pfam13408] (Figura 3). Vale notar que apenas algumas integrases possuem o domínio *Zn-ribbon* (Figura 3), o que se deve à presença ou ausência de quatro resíduos de cisteína ligantes de zinco (estrelas azuis), mas todas as integrases são funcionais (Figura 4) (McEwan *et al.*, 2011).

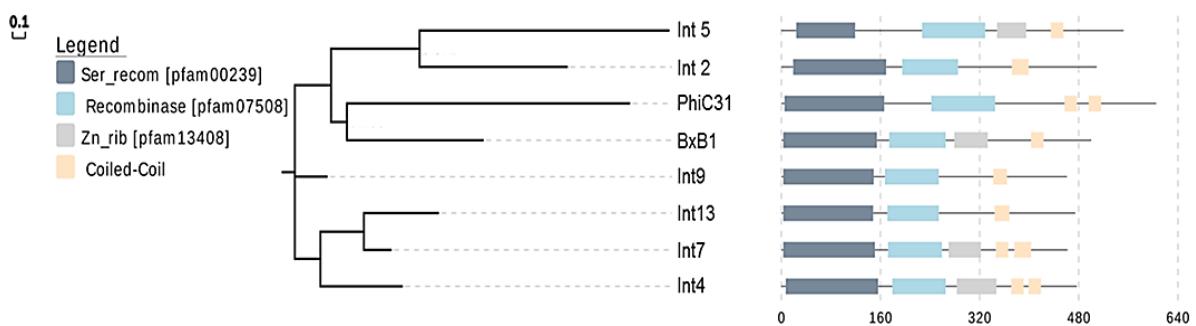


Figura 3. Árvore filogenética e representação esquemática da arquitetura de domínios das serina integrases. A árvore filogenética das integrases de profagos foi inferida usando a sequência de aminoácidos do domínio recombinado de serina, com representação esquemática dos domínios conservados das integrases em sua forma completa. Cada domínio é representado em escala. *Ser_recom* indica o domínio recombinase de serina, e *Zn_rib* representa o domínio beta da fita de zinco da recombinase.

As características estruturais e bioquímicas dessas recombinases foram estudadas em detalhe para compreender seus mecanismos de ação e desenvolver novas ferramentas para a engenharia genética (Keravala *et al.*, 2006; Bonnet *et al.*, 2012; Van Duyne and Rutherford, 2013; Fogg *et al.*, 2014; Rutherford and Van Duyne, 2014; Fan *et al.*, 2016; Fogg *et al.*, 2017; Stark, 2017).

As estruturas tridimensionais de algumas serina integrases foram determinadas por meio de cristalografia de raios X e técnicas de microscopia eletrônica (Ghosh *et al.*,

2006; Rutherford *et al.*, 2013; Gupta *et al.*, 2017; Li *et al.*, 2018; Mandali e Johnson, 2021). Estas estruturas revelam que as integrases adotam uma conformação homodimérica com um domínio catalítico conservado que contém o resíduo de serina do sítio ativo. O domínio catalítico está conectado a um domínio de ligação ao DNA, que reconhece os locais específicos de recombinação na molécula de DNA. A enzima é composta por dois domínios: domínio catalítico N-terminal (NTD) e um domínio de ligação ao DNA C-terminal (CTD) (McEwan *et al.*, 2011). O domínio catalítico contém um resíduo de serina conservado que é essencial para a reação de recombinação, enquanto os domínios de ligação ao DNA reconhecem sequências de DNA específicas para recombinação (sítios *att*) (McEwan *et al.*, 2009).

Na Figura 5A, é possível observar a presença de resíduos conservados (triângulos amarelos): o motivo catalítico N-terminal RxS...(S/D) RxxR. Além disso, essas integrases possuem uma região de articulação flexível que permite a formação de diferentes conformações durante o ciclo catalítico (Figura 5B). Essa flexibilidade permite que a integração adote diferentes configurações para o reconhecimento e ligação aos substratos de DNA, resultando na formação de diferentes produtos de recombinação. Assim, as serina integrases reconhecem e ligam os sítios *attB* e *attP*, resultando na integração, inversão ou excisão do DNA adjacente. Estudos adicionais sobre a estrutura e função dessas integrases ajudaram no desenvolvimento de novas tecnologias para manipulação genética e terapia gênica (Stark, 2017; Merrick *et al.*, 2018).

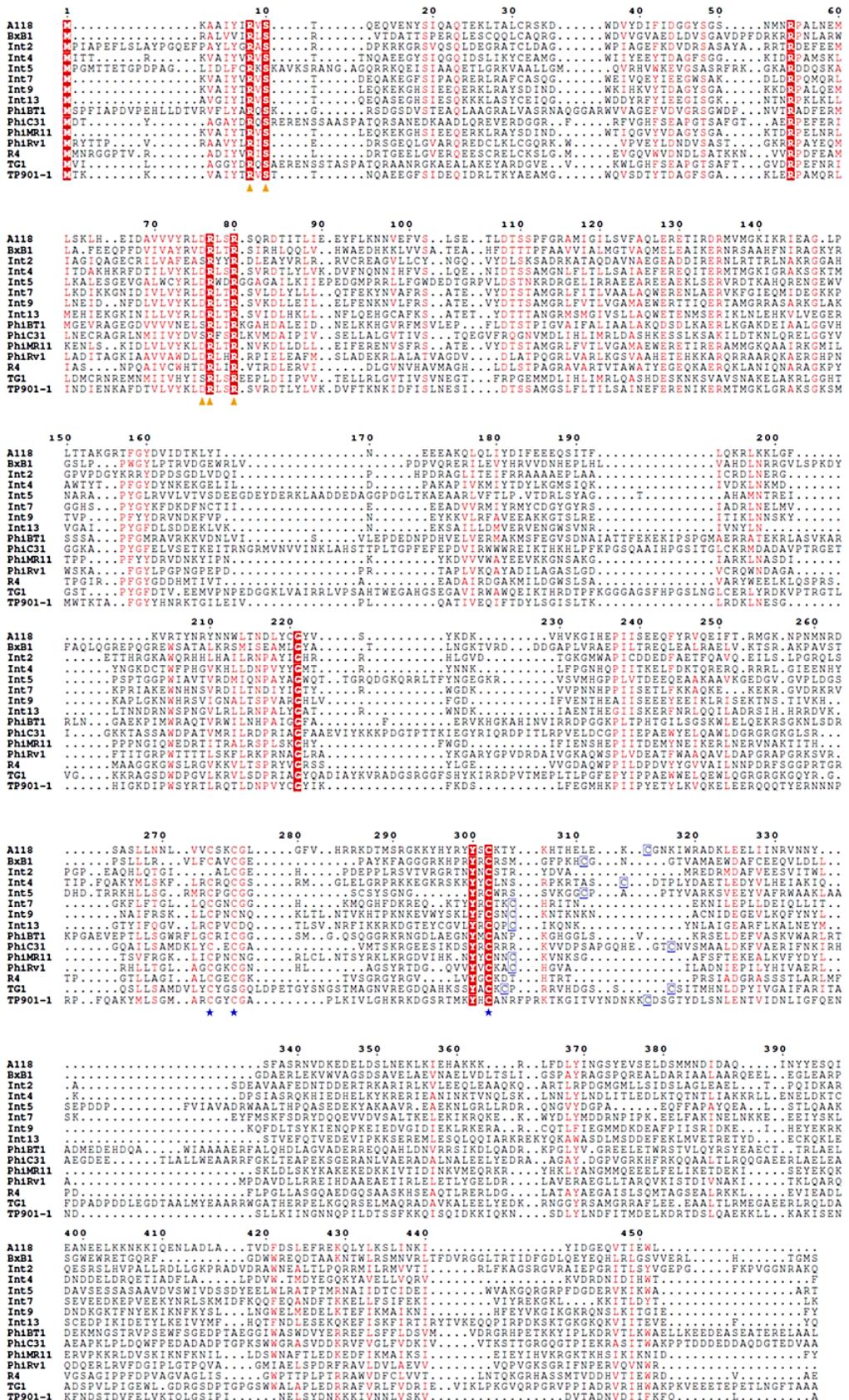


Figura 4. Alinhamento de sequência de aminoácidos de distintas Serina Integrases. Resíduos de domínio catalítico importantes estão indicados com triângulos amarelos e resíduos de cisteína são indicados com estrelas/quadrados azuis. Resíduos conservados estão destacados em vermelho.

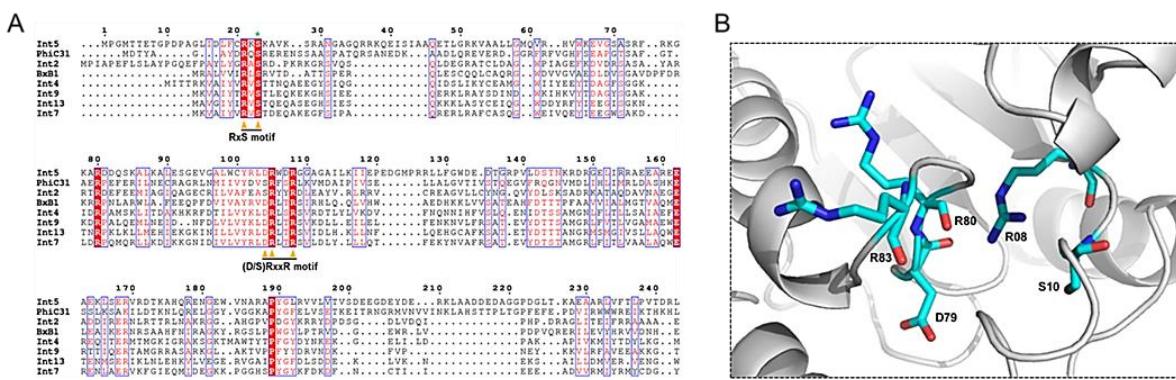


Figura 5. Representação da estrutura tridimensional do domínio catalítico N-terminal da Integrase 13 (Int13). (A) Alinhamento da sequência de aminoácidos da porção N-terminal de diferentes Serina Integrases. Resíduos importantes do domínio catalítico são indicados com triângulos. O nucleófilo serina (S10) está indicado com o asterisco verde. Os resíduos conservados estão destacados em vermelho. **(B)** Modelagem tridimensional por homologia estrutural da arquitetura do motivo catalítico: RxS... (D/S) RxxR da Int13.

1.4 Aplicações promissoras de serina integrases em sistemas eucarióticos

Entre as recombinases, as serina integrases são consideradas uma subfamília particularmente fascinante. Estudos pioneiros com essas enzimas foram realizados para modificar os genomas da bactéria *Streptomyces* em locais específicos conhecidos como *attB*. Usando a serina integrase, os pesquisadores conseguiram integrar de forma estável o plasmídeo pSAM2 no cromossomo bacteriano em um local predeterminado. Essa integração bem-sucedida possibilitou a transmissão do plasmídeo integrado para células descendentes, ampliando as possibilidades de aplicação dessas ferramentas biotecnológicas (Merrick *et al.*, 2018; Abioye *et al.*, 2023). Ao longo de mais de trinta anos, uma ampla variedade de aplicações utilizando serina integrases em vários organismos eucarióticos, incluindo células humanas, pode ser encontrada na literatura (Fogg *et al.*, 2014; Snoeck *et al.*, 2019). O uso dessas proteínas tem sido aplicado de diversas formas, como a remoção de transgenes ou a inserção de sequências desejadas em sítios específicos, denominados nesses casos como “locais de ancoragem” ou “plataformas de aterrissagem”. Além disso, essas proteínas são empregadas em mecanismos de troca de cassetes, conhecidos como RMCE, montagem de DNA e na ativação ou desativação da expressão gênica por meio de inversão de sequência, possibilitando a construção de interruptores genéticos. (Groth *et al.*, 2000; Stark, 2017). Mais recentemente, circuitos de amplificação baseados na expressão de integrases têm sido integrados a plataformas de diagnóstico baseadas em sistemas livres de células (CFS), aumentando significativamente a sensibilidade dessas plataformas,

especialmente quando combinados com a tecnologia de *toehold switch*. Esse avanço permite o desenvolvimento de sistemas de diagnóstico mais precisos e eficientes, ampliando as aplicações biotecnológicas dessas integrases (Franco et al., 2023). Uma importante vantagem em seu estudo é a disponibilidade de uma variedade dessas enzimas que reconhecem sítios com sequências diversas, permitindo o uso de várias delas concomitantemente, e desta forma, a construção de circuitos genéticos. Bonnet et al., (2013) evidencia sua semelhança a um transistor, o nomeando de *transcriptor*. Neste artigo os autores relatam a construção de seis portas lógicas (*Boolean gates*) em *Escherichia coli* usando as serina integrases Bxb1 e TP901 para controlar a orientação de terminadores e promotores e, assim, permitir ou bloquear a expressão do gene repórter *gfp* (*Green fluorescent protein*). Uma robusta busca em bancos de dados por outras integrases foi realizada por Yang et al., (2014), encontrando 34 novas e seus sítios *attP* e *attB* cognatos, das quais 11 foram descritas funcionais em *E. coli* DH10b, com propriedade ortogonal, ou seja, uma integrase não reconhece o sítio de outra. Além disso, demonstraram que várias integrases podem ser utilizadas sequencialmente em um mesmo segmento de DNA para mudar um estado de uma porta lógica transcricional (Yang et al., 2014). Um software para o *design* de circuitos genéticos, nomeado de Cello, foi desenvolvido por Nielsen et al., (2016), o qual possibilitou a testagem de 60 serina integrases em *E. coli*, das quais 45 funcionaram da forma esperada. Entretanto, a maioria dos trabalhos são voltados para células procarióticas. Recentemente, Weinberg et al. (2017), utilizando diferentes enzimas recombinases, desenharam 113 circuitos genéticos para células eucarióticas, como as células de rim embrionário humano (HEK293T) e células Jukart T, alcançando 96,5% de funcionalidade.

Esses dados demonstram a capacidade das serina integrases como ferramentas biológicas para diversas aplicações, desde a engenharia de genomas até a construção de biocomputadores (Brown et al., 2011). Devido à sua direcionalidade, eficiência de recombinação e requisitos simples de sequência de DNA, têm sido adotadas para aplicações nas áreas de genética molecular, biotecnologia e biologia sintética (Fogg et al., 2014).

Para a biologia sintética, os interruptores genéticos com dois estados (ligado/desligado) são componentes essenciais, com aplicações potenciais em biotecnologia, biosensores e biocomputadores (Pokhilko et al., 2018). Entretanto, os estudos da funcionalidade de integrases em células de mamíferos são ainda

escassos, ratificando a necessidade de investigações, pois o foco das pesquisas atualmente tem se voltado principalmente para bactérias e leveduras. Desta forma, os estudos com células de mamíferos apresentam-se como um desafio para a obtenção de células contendo informações genéticas sintéticas, que desempenhem uma nova função e não apresentem efeitos colaterais nocivos (Ye e Fussenegger, 2014). Tal sistema de ativação/desativação poderá ser empregado para investigação da atuação de genes de organismos de interesse biotecnológico ou para controlar a expressão de proteínas heterólogas, por meio dos circuitos genéticos. As aplicações poderão englobar áreas diversas como tratamento de distúrbios metabólicos, doenças imunomediadas, terapia anticâncer, área reprodutiva humana ou animal, melhoramento e clonagem animal no campo agropecuário, entre outras. Além disso, a necessidade de reconhecimento de dois sítios (*attB/attP*) para sua ação pode ser um fator minimizador de edição de genes não alvo. Dessa forma, partindo do estudo de Yang *et al.* (2014), a proposta inicial desse projeto de tese foi avaliar a funcionalidade das serina integrases Int2, Int4, Int5, Int7, Int9, Int13 e das integrases phiC31 e Bxb1 utilizando células primárias de fibroblasto bovino como modelo de estudo em sistema de expressão transiente.

2. OBJETIVOS

2.1 Geral

Avaliar, em sistema transiente, a funcionalidade de seis serina integrases (Int2, Int4, Int5, Int7, Int9 e Int13) em células primárias de fibroblasto bovino visando expandir o número de proteínas dessa classe capazes de ativar/desativar interruptores genéticos em células de mamíferos.

2.2 Específicos

- Construir um sistema de plasmídeos para expressão das integrases e de plasmídeos contendo o gene repórter *gfp* em orientação reverso complementar com os sítios *attB/attP* das respectivas integrases;
- Transfектar células de fibroblasto bovino em sistema de cotransfecção transiente;
- Quantificar a ativação do gene repórter (*gfp*) por microscopia de fluorescência e citometria de fluxo;

- Avaliar o desempenho e citotoxicidade de cada integrase na célula testada;
- Verificar a correta inversão do gene repórter (*gfp*) e a esperada formação dos sítios *attL* e *attR* obtidos após a ação da integrase.

3. METODOLOGIA

3.1 Construção dos vetores

Seis serina integrases (Int2, Int4, Int5, Int7, Int9 e Int13) foram selecionadas para testes em células de fibroblasto bovino. As serina integrases selecionadas foram escolhidas dentre 11 integrases funcionais em *E. coli* descritas por Yang *et al.*, (2014), considerando as seguintes características: a) não formar o códon ATG no sítio *attL* que pudesse retirar de frame a sequência do GFP; e b) não formar os códons de parada da tradução (UAG, UAA, ou UGA). As integrases Bxb1 e phiC31 (Tabela 1), também foram avaliadas, por serem funcionais como interruptores genéticos em células humanas e de planta (Weinberg *et al.*, 2017; Rubtsova *et al.*, 2008).

Tabela 1. Integrases selecionadas, código do *GenBank* e organismo de origem.

Nome	Nº NCBI	Organismo
Int2	CBG73463	Vírus de <i>Streptomyces scabiei</i> 87.22
Int4	YP_002747001	Vírus de <i>Streptococcus equi</i> subsp. <i>equi</i> 4047
Int5	BAF03598	Vírus de <i>Streptomyces</i> phage PhiK38-1
Int7	YP_003251752	Vírus de <i>Geobacillus</i> sp. Y412MC61
Int9	BAF67264	Vírus de <i>Staphylococcus aureus</i> str. Newman
Int13	YP_001376196	Virus de <i>Bacillus cytotoxicus</i> NVH 391-98
phiC31	AFR23342	Virus de <i>Streptomyces</i>
Bxb1	NC_002656	Vírus de <i>Mycobacterium</i>

Fonte: Adaptado de Yang *et al.* (2014).

Previamente, as integrases Int2, Int4, Int5, Int7, Int9, Int13, phiC31 e Bxb1 foram códon otimizadas por meio da ferramenta de otimização de códons (IDT,

<https://www.idtdna.com/pages>), para a expressão em sistema eucariótico de mamíferos (otimização para *Homo sapiens*).

Para a construção dos vetores de expressão, as integrases Int2, Int4, Int5, Int7, Int9, Int13, phiC31 e Bxb1 foram clonadas sob o controle do promotor de ubiquitina e do terminador da β-globina no plasmídeo pUB-GFP (Addgene, 11155), substituindo a sequência codificadora do *gfp*, o que resultou no conjunto dos vetores de expressão da integrase chamados pIE IntX (X = 2, 4, 5, 7, 9, 13, phiC31 ou Bxb1) (Figura 6A). Para os vetores do repórter, a sequência codificadora EGFP (Addgene, 11154) foi clonada na orientação reverso complementar flanqueada pelo sítio *attB* (posição 5') e pelo sítio *attP* (posição 3') de uma das integrases Int2, Int4, Int5, Int7, Int9, Int13, phiC31 e Bxb1. Esses cassetes foram clonados no plasmídeo pEF-GFP (Addgene, 11154) substituindo a sequência codificadora do EGFP original sob a regulação do promotor alfa EF1 e do terminador de sinal poli(A) da β-globina. Os vetores repórteres resultantes foram denominados pSG IntX (X = 2, 4, 5, 7, 9, 13), phiC31 ou Bxb1 (Figura 6B). O plasmídeo pEF-GFP original (Addgene, 11154) foi utilizado como controle positivo de expressão de *gfp* em todos os experimentos. As sínteses e clonagens das sequências descritas foram realizadas pela Empresa Epoch Life Science Inc. e para a amplificação dos plasmídeos foi utilizado o *kit* comercial QIAGEN Plasmid Maxi kit (QIAGEN, Hilden, Alemanha), de acordo com as recomendações do fabricante.

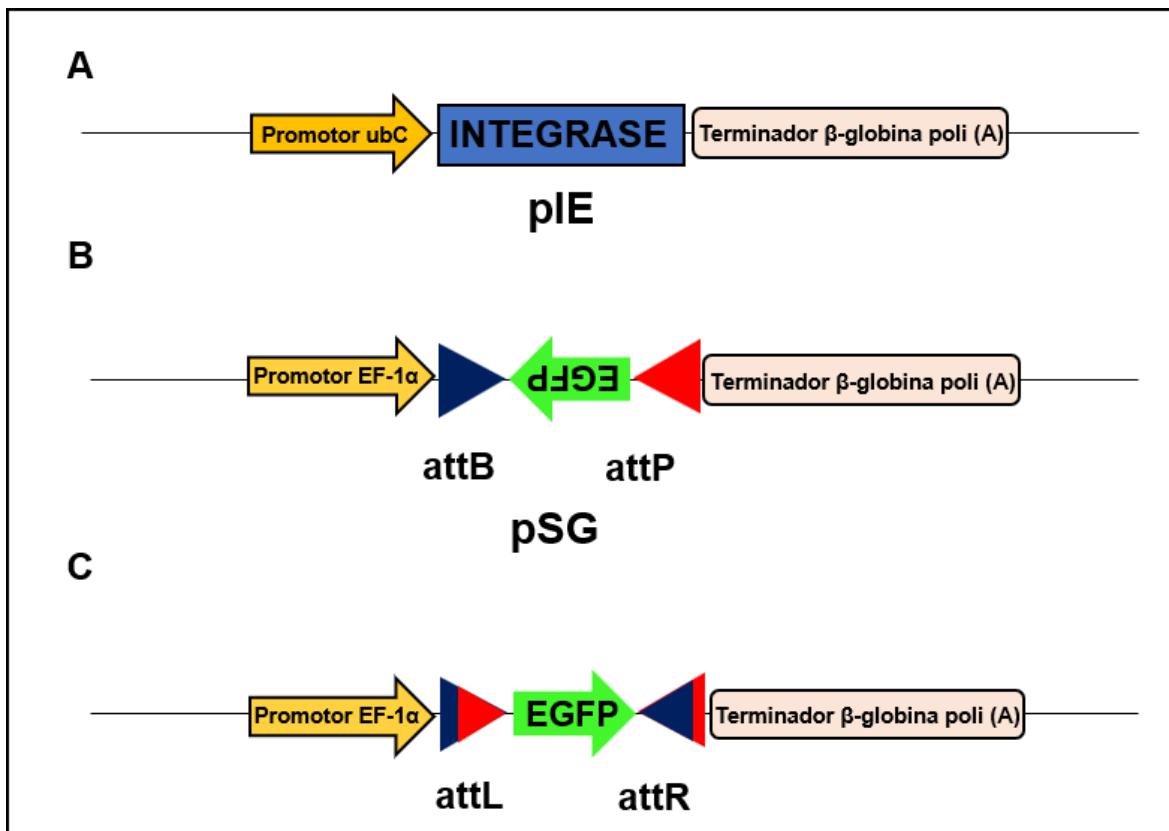


Figura 6. Representação esquemática dos cassetes de expressão construídos para avaliar a funcionalidade das integrases em células de fibroblastos bovino e um plasmídeo putativo resultado da atuação das integrases. **(A)** plasmídeo de expressão da integrase (pIE); **(B)** plasmídeo repórter (pSG) com os sítios *attB* e *attP* das integrases flanqueando o gene *egfp*克隆ada na orientação reverso complementar; **(C)** Plasmídeo putativo resultado da inversão pela ação da integrase, formando os sítios *attL* e *attR*.

3.2 Cultivo celular e condições de cultura

Fibroblastos bovinos foram obtidos a partir de biópsias retiradas da orelha de boi da raça nelore (*Bos indicus*). As linhagens foram cultivadas em meio Dulbecco's Modified Medium (DMEM, GIBCO), suplementado com 10% de soro fetal bovino (GIBCO). As linhagens foram incubadas a 37°C em atmosfera gasosa de 5% de CO₂, com troca do meio de cultura a cada três dias (Freshney, 1994). A contagem de células foi realizada em câmara de Neubauer com azul de tripan, plaqueando 5,0x10⁵ células por frascos de cultura com área de superfície de 75 cm². Um microscópio invertido equipado com uma plataforma aquecida para prevenir choque térmico foi utilizado para acompanhar a taxa de crescimento celular. Cultivos celulares com 60% a 70% de confluência foram submetidos a passagens (repiques). Após o crescimento, as células foram tripsinizadas com solução de tripsina 0,05% e EDTA, seguida por lavagens com Dulbecco's Phosphate-Buffered Saline (DPBS, GIBCO).

3.3 Cotransfecção das células

Os fibroblastos bovinos primários foram cotransfetadas com um vetor pIE, contendo o gene de uma das integrases, e um segundo vetor pSG contendo o gene repórter em orientação invertida, flanqueado por sítios *attB/attP* da integrase correspondente. Alíquota de 350 ng de cada um dos dois plasmídeos (pIE e pSG) foram estruturadas em lipossomas catiônicos (Lipofectamine® LTX & Plus Reagent Invitrogen, Carlsbad, EUA) e meio Opti-MEM (Invitrogen, California, EUA), de acordo com as orientações do fabricante. Os ensaios foram realizados em triplicatas técnicas e biológicas. A expressão transiente de GFP foi analisada 48 h após a cotransfecção.

3.4 Microscopia de fluorescência

A análise da fluorescência de GFP e aquisição de imagens foram realizadas em microscópio de fluorescência Axiovert 135M (Carls Zeiss). As imagens foram adquiridas em três campos aleatórios, usando uma câmera digital conectada DS-Ri1 (Nikon) e o software de captura Nikon Digital sight DS-L3 (Nikon) sob luz UV, com as seguintes configurações: conjunto de filtros nº15 (Carls Zeiss); excitação: BP 546; divisor de feixe: FT 580; emissão LP 590.

3.5 Quantificação da população celular GFP positiva por citometria de fluxo

As células foram cultivadas em placas de vinte e quatro poços contendo 5×10^5 células/poço. Um total de 10.000 células foram analisadas em três experimentos independentes, após 48 h de transfeção, em citômetro de fluxo de imagem (Amnis 582 FlowSight) sob excitação a laser de 488 nm e potência de 60 e 10 mW. A aquisição das células foi realizada sob a intensidade do canal 3 (filtros a 566-635) versus a intensidade do canal 2 (emissão verde do repórter GFP, 532-555 nm). Os resultados foram processados e os gráficos obtidos usando o software IDEAS®.

3.6 Avaliação da citotoxicidade das integrases por método MTT

Células de fibroblasto bovino foram cultivadas em microplacas de 96 poços a uma densidade de 1×10^5 células/ml, a 37°C com 5% de CO₂, em estufa. As células foram cotransfetadas com os vetores pIE + pSG ou apenas com um desses vetores mais um plasmídeo *mock* (controles), como descrito no item 3.3. No controle negativo foram adicionados 10% de dimetilsulfóxido (DMSO; Sigma-Aldrich), para um volume

final de 200 µl/poço. Realizou-se também controles de células com lipofectamina LTX, transfecção com *mock* e poços contendo células somente com meio de cultura (dados não apresentados). Após 48 h do tratamento, todos os grupos foram incubados com 20 µl de brometo de 3- (4,5-dimetiltiazol-2-il) -2,5-difeniltetrazólio (MTT, Thermo Fisher Scientific) na concentração de 5 mg/ml durante 4 h a 37°C. Posteriormente, o meio foi removido e 200 µl de DMSO foi adicionado em cada poço para dissolver os cristais de formazan resultante da atividade das células viáveis. A quantificação do produto formazan foi realizada em espectrofotômetro com leitor de placa (Sunrise Reader, Tecan Magellan™) a 595 nm, com valores processados pelo programa Magellan. Os ensaios foram realizados em triplicata com três repetições em dias independentes.

3.7 Extração, amplificação e sequenciamento do DNA

O DNA total foi extraído das células após 48 h de transfecção, utilizando o *kit* comercial DNeasy Blood & Tissue (QIAGEN, Hilden, Alemanha). Para confirmar a funcionalidade das integrases pela inversão da sequência de *gfp* e correta formação dos sítios *attL/attR*, foram realizadas análises por PCR e sequenciamento do DNA. A PCR foi feita utilizando-se dois pares de oligonucleotídeos (Figura 11A): um par anelando na sequência do promotor (senso) e no sítio *attR* (antisenso), e o outro par anelando no sítio *attL* (senso) e em uma região do *gfp* próxima à sequência do terminador (antisenso) (Tabela 2). A amplificação somente ocorrerá se houver a ação da integrase invertendo o gene *gfp*. Para a reação de PCR foi utilizada a enzima Platinum Taq DNA Polymerase (Invitrogen) com as seguintes condições de termociclagem: 94°C por 3 minutos, 35 ciclos de 94°C por 30 segundos, 60°C por 30 segundos, 72°C por 1 minuto e 30 segundos e uma extensão final a 72°C por 10 minutos. A amplificação do DNA foi detectada por eletroforese em gel de agarose 1% com SYBR Green, a 80 V em tampão TAE 1X, e visualizado em transiluminador UV, utilizando o Sistema Chemidoc XRS Gel Chemiluminescente (Bio-Rad, California, EUA). Os produtos de PCR foram purificados com o *kit* Wizard®SV Gel (Promega), clonados no pGEM-T Easy Vector (Promega) de acordo com o protocolo do fabricante e introduzidos em células quimicamente competentes de *E. coli* (DH10B ou XL1-blue) por choque térmico. O DNA plasmidial foi extraído com o *kit* Wizard® Plus SV Minipreps DNA Purification System (Promega) e enviados para sequenciamento na empresa Macrogen, usando os pares de iniciadores universais

M13F e M13R ou SP13 e T6. As sequências obtidas foram analisadas usando o software Geneious (versão 7.0.6).

Tabela 2. Sequências de iniciadores utilizados para se ligarem aos sítios *attL* e *attR* formados após a atividade das integrases em fibroblastos bovinos. Os valores de Tm foram calculados utilizando a ferramenta *OligoAnalyzer* do site da *Integrated DNA Technologies* (IDT) <https://www.idtdna.com/calc/analyzer>.

Iniciadores para amplificar <i>attL</i>		Tm (°C)
Promotor	Iniciador senso 5'-> 3'	
EFa_966F	TTCTCGAGCTTTGGAGTACGTCGTCTTAGGTTG	71
<i>attR</i>	Iniciador antisenso 5'-> 3'	
<i>attR</i> _Int2_R	GTTGCTACGCGAGATTCTGCCGGACCGTCGACATACTGC	76,4
<i>attR</i> _Int4_R	AGTTTCAACCCTTGATTGAATAAGACTGCTGCTGTGT	72,2
<i>attR</i> _Int5_R	ATAACTCTCCTGGGAGCGCTACACGCTGTGGCTG	75,4
<i>attR</i> _Int7_R	CTGTGTGAGAGTTAACGTTACATGGCAAAGTTGATGAC	70,7
<i>attR</i> _Int9_R	TGGAAGTGTGTATCAGGTAACGGATACCTCATC	69,2
<i>attR</i> _Int13_R	GTTAGAACTTGACCAGTTGGCCTGTAAATATAAGCAATCC	70
<i>attR</i> _phiC31_2	CCAACTGGGTAACCTTGGCTCC	70,1
<i>attR</i> _Bxb1_R2	CTGGTCAACCACCGCGGTCTCCGTCAGGATC	76,3
Iniciadores para amplificar <i>attR</i>		Tm (°C)
<i>attL</i>	Iniciador senso 5'-> 3'	
<i>attL</i> _Int2_F	GGAGTAGCTTCGCCCGAGAACTTCTGCAAG	72,4
<i>attL</i> _Int4_F	CGACCTGAAATTGAATTAGCGGTCAAATAATTGTA	68,2
<i>attL</i> _Int5_F	GACGGCCTGGGAGCGTTGACAACTTGCGCACC	77,2
<i>attL</i> _Int7_F	GTCCGTCTGGTCAGTTGCCTAACCTTAACCTTAC	71,2
<i>attL</i> _Int9_F	ATAATTGGCGAACGAGGTATCTGCATAGTTATTCCGAAC	71,1
<i>attL</i> _Int13_F	TCCAGATCCAGTTGTTTAGTAACATAAATACA	65,3
<i>attL</i> _phiC31_F	TGCCAGGGCGTGCCCTTGAGTTCTCAGT	75,7
<i>attL</i> _Bxb1_F	TGTCGACGACGGCGGTCTCAGTGGTGTACGGT	76,6

Backbone	Iniciador antisenso 5'-> 3'	
TermiAni_205R	AATGATTGCCCTCCCATATGTCCCTCCGAGTG	71.6

3.8 Análise estatística

O conjunto de dados com a proporção de células geneticamente ativadas (EGFP +) obtido por citometria de fluxo foi analisado no software R (versão 3.6.0). O teste não paramétrico de Kruskal-Wallis foi utilizado para determinar diferenças estatísticas entre controles e condições de teste de cada grupo das integrases com 5% de nível de significância.

4. RESULTADOS

Esses dados foram publicados como parte do artigo *Genetic switches designed for eukaryotic cells and controlled by serine integrases* (DOI: <https://doi.org/10.1038/s42003-020-0971-8>) (Anexo I).

4.1 Caracterização funcional das integrases

Para testar a atividade das integrases 2, 4, 5, 7, 9, 13, phiC31 e Bxb1 em células de mamíferos, os fibroblastos bovinos foram cotransfetados com dois plasmídeos (pIE e pSG). O plasmídeo pIE é um vetor para expressão de uma integrase e o pSG contém o gene repórter *egfp* em sentido antisenso com os sítios *attB* à esquerda e o sítio *attP* à direita da região codificadora sob regulação do promotor EF-1 α (Figura 6). Após 48h da cotransfecção, as células de fibroblastos foram avaliadas por microscopia de fluorescência e citometria de fluxo. As células utilizadas como controles negativos foram cotransfetadas com apenas um dos dois plasmídeos (pIE ou pSG) mais um plasmídeo *mock*. Como controle positivo foi utilizado o vetor pGFP, que é um vetor de expressão do gene repórter, ou seja, contém o gene da *egfp* na orientação senso sob o controle do mesmo promotor presente no vetor pSG e mais um plasmídeo *mock*.

A análise qualitativa por microscopia demonstrou que houve fluorescência de EGFP nas células cotransfetadas com as integrases Int9, Int13, phiC31 e Bxb1 e seus respectivos vetores pSG, embora em porcentagem inferior às células transfectadas com o vetor controle (pGFP) (Figura 7). Além disso, um detalhe de uma célula de

fibroblasto bovino isoladamente cotransfetada com a integrase Int13 e apresentando fluorescência de EGFP é exibido na Figura 7B.

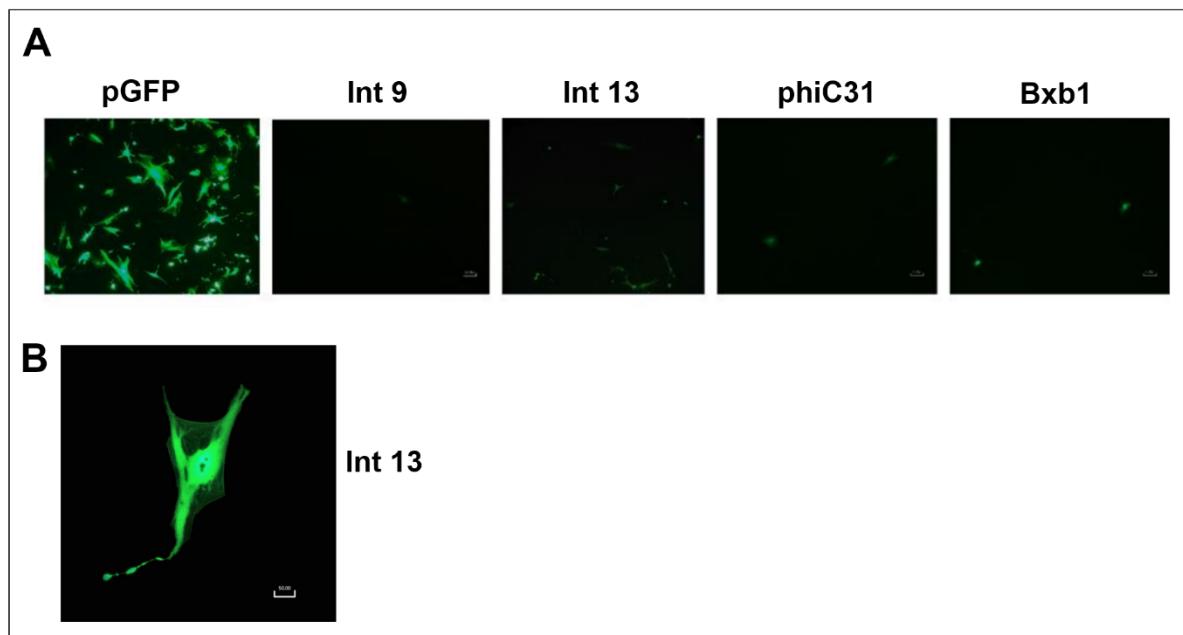


Figura 7. Imagens representativas da expressão de EGFP nos fibroblastos bovinos. (A) pGFP: controle positivo; Int9: células cotransfetadas com os vetores de expressão da Integrase 9 (pIE) e o vetor pSG; Int13: células cotransfetadas com o vetor de expressão da Integrase 13 (pIE) e o vetor pSG; phiC31: células cotransfetadas com o vetor de expressão da Integrase phiC31 (pIE) e o vetor pSG; Bxb1: células cotransfetadas com o vetor de expressão da Integrase Bxb1 (pIE) e o vetor pSG. (B) Detalhe de uma célula de fibroblasto bovino isolada cotransfetada com a integrase Int13, exibindo fluorescência de EGFP. As imagens foram obtidas em microscópio de fluorescência Axiovert 135M (Carls Zeiss) utilizando uma câmera digital DS-Ri1 (Nikon) acoplada e o software de captura Nikon Digital sight DS-L3 (Nikon) sob luz UV com o conjunto de filtros 15 (Carls Zeiss) - excitação: BP 546; divisor de feixe: FT 580; emissão: LP 590. As imagens foram adquiridas em campos aleatórios.

Os resultados obtidos pelas análises quantitativas em citômetro de fluxo corroboram a análise de microscopia, demonstrando que as integrases Int9, Int13, phiC31 e Bxb1 resultaram em acúmulo de GFP (Figura 8A). O gráfico com a porcentagem de células que expressam GFP e o desvio padrão das triplicatas biológica e técnicas está apresentado na Figura 8B. Os círculos no eixo y representam os valores médios da triplicata biológica e técnica de cada ensaio. O eixo x contém as diferentes condições do ensaio. Para cada grupo de dados das integrases, as letras diferentes indicam diferenças significativas ($p < 0,05$). As integrases Int9, Int13, phiC31 e Bxb1 se destacaram pelo maior percentual de células fluorescentes acumulando GFP (Figura 8 A e B).

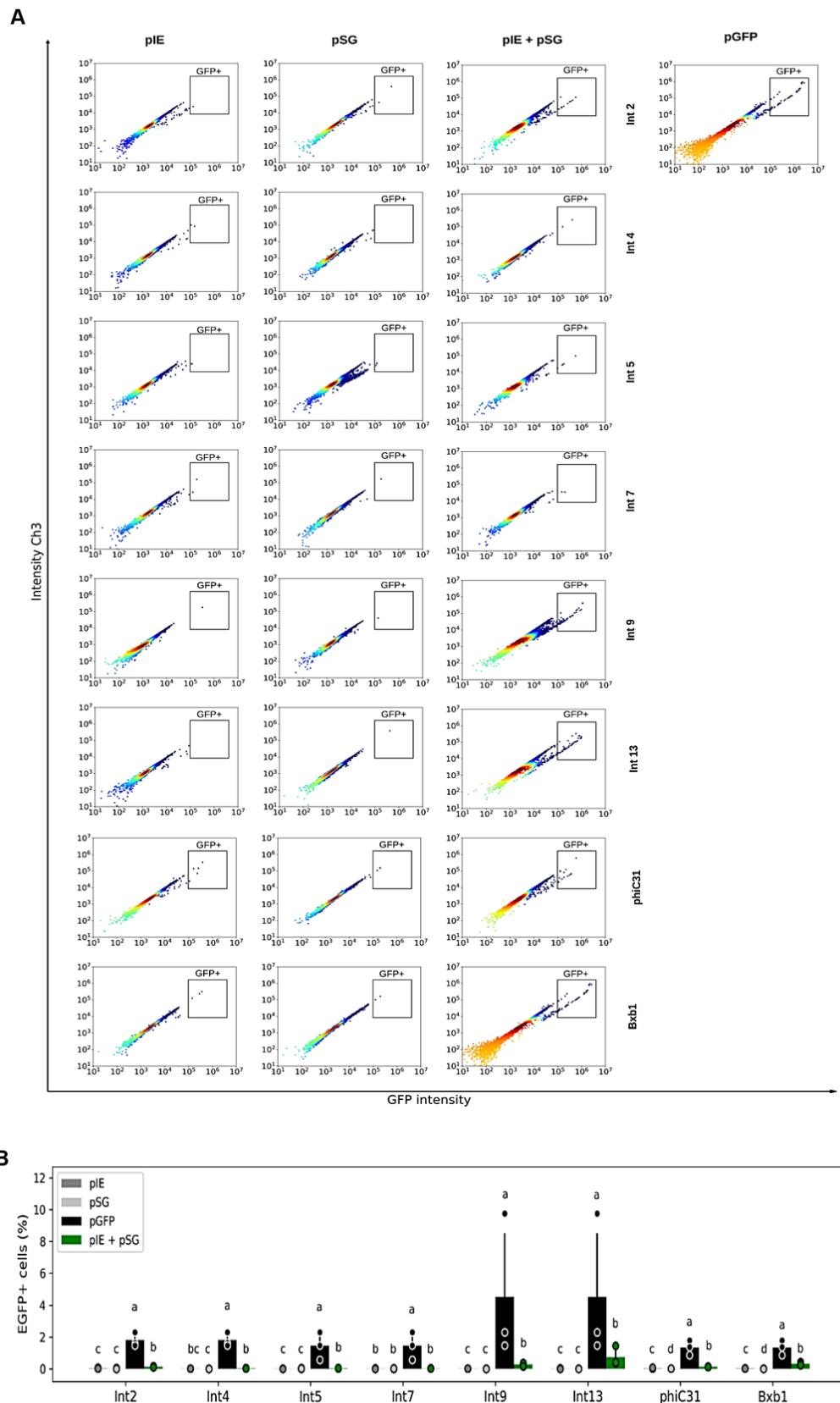


Figura 8. Caracterização funcional das integrases em fibroblastos bovinos avaliadas por citometria de fluxo. (A) Distribuição por citometria de fluxo das células após 48 h da cotransfecção para teste da inversão do gene EGFP pelas integrases Int2, Int4, Int5, Int7, Int9, Int13, phiC31 e Bxb1.

Int9, Int13, phiC31 e Bxb1. Cores quentes indicam maior concentração de células enquanto as cores mais frias concentrações menores. A população marcada nos quadrados são células que expressam GFP. Os controles negativos foram cotransfetadas com um dos plasmídeos do vetor de expressão da integrase (pIE) das Int2, Int4, Int5, Int7, Int9, Int13, phiC31 e Bxb1 ou um dos vetores repórter GFP (pSG) que contém o gene *egfp* em orientação reverso complementar flanqueado pelos sítios *attB/ attP* das integrases Int2, Int4, Int5, Int7, Int9, Int13, phiC31 e Bxb1 mais um plasmídeo *mock*. O grupo teste (pIE + pSG) indica as células cotransfetadas com vetores de expressão de integrase e vetores repórter GFP (tratamento). As células de controle positivo foram cotransfetadas com o plasmídeo pGFP contendo o gene EGFP na orientação senso (pGFP) + um plasmídeo *mock*. **(B)** Gráfico de barras representa a média da expressão de cada integrase em porcentagem e o desvio padrão da população de células que expressam GFP está representado pelos círculos (eixo y). Cada grupo de integrase do eixo x contém as diferentes condições avaliadas nesse estudo. Os experimentos foram realizados em triplicatas técnicas e experimentais ($n=3$). Para cada grupo de dados das integrases, letras diferentes indicam diferenças estatisticamente significativas ($p<0.05$).

4.2 Avaliação da citotoxicidade das integrases

Um importante fator limitante ao uso de integrases em interruptores genéticos (*switches*) é seu potencial citotóxico. Em células viáveis o reagente MTT, de coloração amarelada, é reduzido a cristais de formazan de cor púrpura, por meio da atividade da enzima desidrogenase mitocondrial. A concentração de cristais formados é diretamente proporcional à concentração de células viáveis. O solvente DMSO foi utilizado para solubilizar os cristais de formazan acumulados no citoplasma celular, com isto, houve a formação de uma solução de cor púrpura cuja intensidade da cor foi avaliada por espectrofotômetro no comprimento de onda de 595 nm e comparada com a absorbância das células cotransfetadas com o vetor pGFP e um plasmídeo “*mock*” (controle). Dessa forma, observa-se que as células de fibroblastos bovinos se apresentaram viáveis após 48 horas da cotransfecção para todas as integrases testadas (Figura 9).

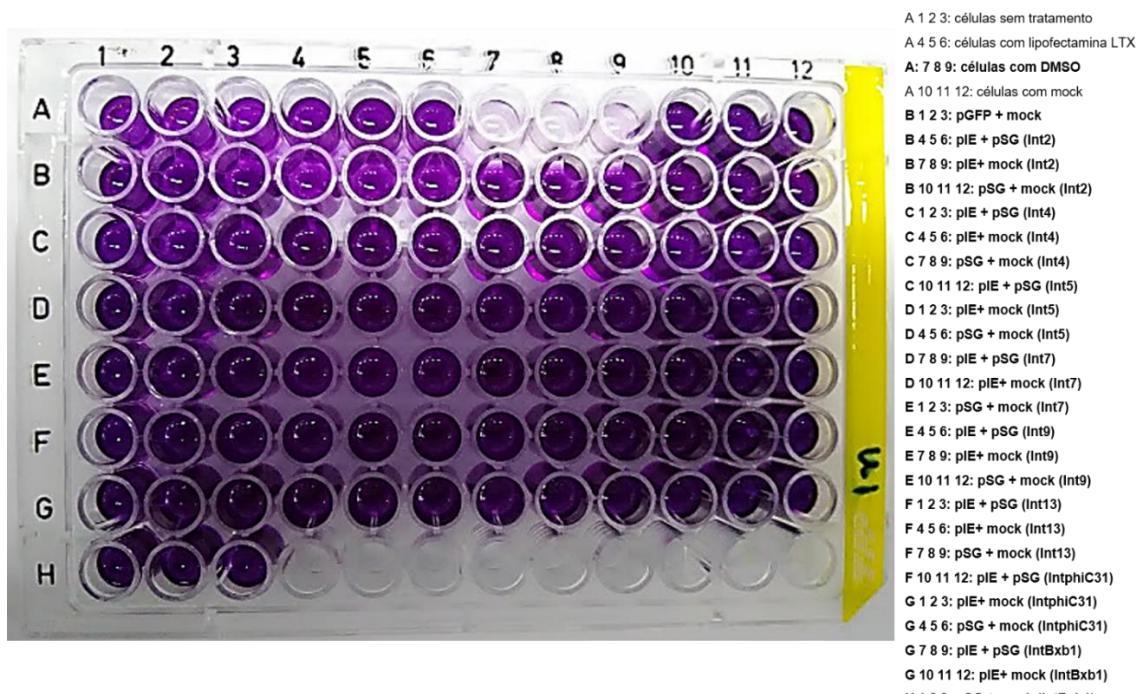


Figura 9. Reação do MTT em células de fibroblasto bovino após tratamento com todas as integrases.

Os valores de absorbâncias das diferentes cotransfecções foram normalizadas pelos valores das células controle (*pGFP + mock*), demonstrando que as células de fibroblasto bovino não apresentaram comprometimento significativo de viabilidade em comparação com o grupo controle (Figura 10).

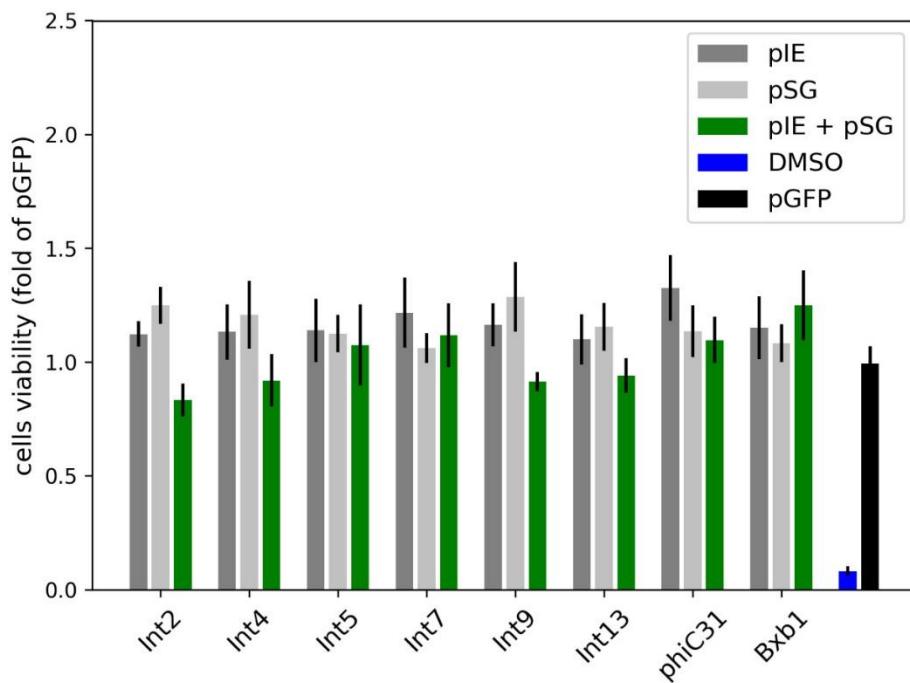


Figura 10. Efeito das integrases sobre a viabilidade de células de fibroblasto bovino. O gráfico representa a média e o desvio padrão das células cotransfetadas com o vetor de expressão de integrase (pIE) mais o vetor repórter (pSG). Os controles negativos foram cotransfetados apenas com um dos dois conjuntos de vetores e o plasmídeo mock. As células do controle positivo (pGFP + mock) foram arbitrariamente definidas como 100% de viabilidade, para fins de normalização. O grupo DMSO corresponde ao controle de citotoxicidade do experimento. Os experimentos foram realizados em triplicatas técnicas e biológicas ($n = 3$). A absorbância foi medida em OD = 595 nm.

4.3 PCR e sequenciamento dos sítios *attL* e *attR*

Os ensaios de PCR seguidos de sequenciamento dos *amplicons* demonstraram que houve formação dos sítios *attL/attR* previstos para todas as integrases testadas (Int2, Int4, Int5, Int7, Int9, Int13, phiC31 e Bxb1) (Figuras 11 e 12). Embora a fluorescência da proteína repórter GFP não tenha sido observada por microscopia ou citometria de fluxo nas células cotransfetadas com as integrases Int2, Int4, Int5 e Int7, a detecção da formação dos sítios *attL/attR* sugere que houve a correta inversão da sequência codificadora de GFP, apesar de as células não terem gerado fluorescência suficiente para ser detectada nos experimentos de análise qualitativa.

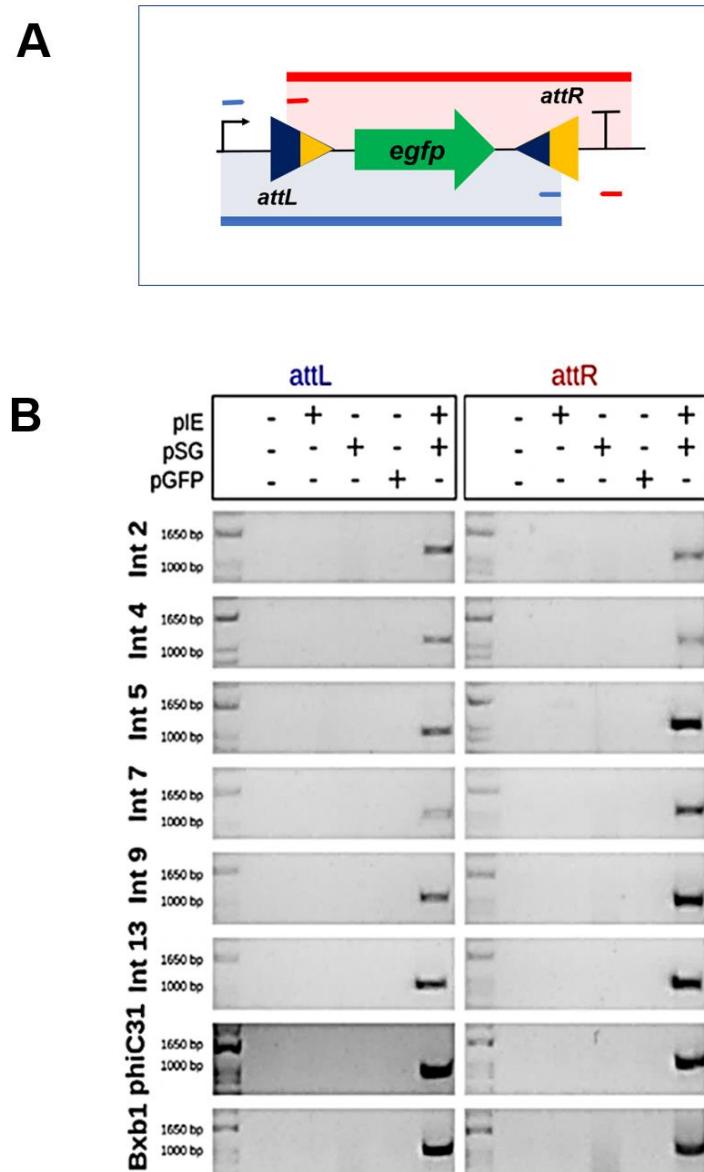


Figura 11. Análise da funcionalidade das integrases quanto à formação dos sítios attL/attR e rotação da sequência codificadora do gene repórter egfp. (A) Esquema representando a posição dos pares de iniciadores: um par anelando na sequência do promotor (senso) e no sítio attR (vermelho) e o outro par anelando no sítio attL (senso) e em uma região do GFP próxima à sequência do terminador (azul). (B) Eletroforese em gel de agarose 1% dos amplicons obtidos por PCR utilizando iniciadores específicos para amplificação de segmentos de DNA contendo os sítios attL/attR, formados após a ação das integrases invertendo a sequência codificadora (egfp). Foi utilizado o marcador 1 kb Plus DNA ladder (Ready-LoadTM- Invitrogen), para determinação do tamanho dos fragmentos. O tamanho esperado dos amplicons varia entre as diferentes integrase, o segmento que contém o attL varia de 1021 a 1104 pb e o que contém o attR de 1058 a 1084 pb.



Figura 12. Alinhamento dos sítios *attL/attR* obtidos por sequenciamento com os sítios previstos. Observa-se as sequências *attL* e *attR* representativas obtidas após a atuação das integrases em comparação com as sequências preditas (destacado em cinza). *attL1* e *attR1* correspondem às regiões dos sítios *attP* e *attB* que foram invertidas, respectivamente. *attL2* e *attR2* correspondem às regiões *attB* e *attP* que não sofreram alterações, respectivamente.

5. DISCUSSÃO E CONCLUSÃO

Os resultados obtidos neste estudo demonstram que as integrases testadas foram eficazes na recombinação da sequência do gene alvo (*egfp*), possibilitando sua expressão em células de fibroblasto bovino. Este achado evidenciou o potencial das serina integrases como interruptores genéticos sintéticos em células de mamíferos. As análises adicionais revelaram que as integrases Int9 e Int13 apresentaram maior atividade em fibroblasto bovino, e que as integrases não interferem na viabilidade dessas células. Embora as células cotransfetadas com as integrases Int2, Int4, Int5 e Int7 não tenham apresentado fluorescência visível tanto ao microscópio quanto no citômetro de fluxo, a PCR, seguida de sequenciamento, confirmou a inversão do gene repórter *gfp*. Além disso, o sequenciamento dos sítios de recombinação após a cotransfecção demonstrou que os sítios resultantes (*attL* e *attR*) estavam de acordo com os sítios previstos. Esses resultados reforçam que as serina integrases podem ser utilizadas como ferramentas de regulação gênica em sistemas eucarióticos transientes para aplicações biotecnológicas, atuando na forma de interruptores genéticos. Futuramente, esses interruptores podem se tornar uma estratégia avançada para a criação de novas rotas metabólicas reguladas por integrases, permitindo a adição de genes de interesse com os sítios *attP/attB* em diferentes tipos celulares. Essa abordagem poderia condicionar a expressão gênica à indução por fatores bióticos ou abióticos, contribuindo significativamente em diversas áreas, como a modificação de linhagens celulares, o desenvolvimento de organismos transgênicos e os avanços em terapias gênicas. É importante ressaltar que o uso de integrases na construção de interruptores e circuitos em células de mamíferos ainda é limitado. Isso destaca a necessidade de investigações adicionais e a relevância dos resultados apresentados neste projeto de tese. Em um contexto mais amplo, diversas técnicas estão sendo empregadas na pesquisa de edição genética, incluindo sistemas como TALEN (*Transcription activator-like effector nuclease*), ZFN (*Zinc Finger Nucleases*) e CRISPR/Cas9 (*Clustered regularly interspaced palindromic repeats/associated (Cas) protein 9*). Cada uma dessas metodologias utiliza diferentes proteínas para direcionar enzimas de restrição a locais específicos do DNA, apresentando vantagens e especificações que influenciam a escolha conforme o propósito da alteração genética. A identificação de nucleases com alta espessura e especificidade continua sendo um desafio no campo (Gaj *et al.*, 2013).

No entanto, novas abordagens estão sendo desenvolvidas na Biologia Sintética para facilitar a caracterização e geração de componentes moleculares, células, biomoléculas, além da expressão e modificação de material genético (DNA/RNA). A engenharia genética possui aplicações promissoras nas áreas de saúde, agricultura e meio ambiente, necessitando, portanto, de ferramentas que melhorem a modulação de múltiplos genes. O uso de serina integrases ganhou destaque em estudos de Biologia Sintética devido à sua capacidade de mediar recombinações específicas de sequências de DNA, permitindo a integração ou remoção de genes de interesse em locais específicos do genoma. A alta eficiência dessas recombinações na inserção ou remoção de sequências de DNA minimiza indesejadas no genoma, tornando-as promissoras para aplicações biotecnológicas, como a produção de organismos transgênicos e terapias genéticas. Concluindo, as serina integrases são ferramentas valiosas para promover a integração seletiva de sequências genômicas em células eucarióticas, especialmente quando utilizadas em conjunto com plasmídeos doadores que contêm o sítio de reconhecimento *attB* e sítios *attP* genômicos ou *pseudo-attP* (Merrick *et al.*, 2018). Estudos recentes têm demonstrado o sucesso das serina integrases em células vegetais, humanas e animais, abrindo novas possibilidades para aplicações terapêuticas e biotecnológicas. As aplicações potenciais incluem o tratamento de distúrbios metabólicos, doenças imunomediadas, terapia anticâncer, e melhoramento animal e clonagem na agricultura.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- Abioye, J., Lawson-Williams, M., Lecanda, A., Calhoon, B., McQue, A.L., Colloms, S.D., et al. 619 (2023). High fidelity one-pot DNA assembly using orthogonal serine integrases. *Biotechnol J* 620 18(3), e2200411. doi: 10.1002/biot.202200411.
- Alkhalil SS. The role of bacteriophages in shaping bacterial composition and diversity in the human gut. *Front Microbiol.* 2023 Sep 19;14:1232413. doi: 10.3389/fmicb.2023.1232413. PMID: 37795308; PMCID: PMC10546012.
- Baaziz, H., Makhlouf, R., McClelland, M., and Hsu, B.B. (2024). Bacterial resistance to temperate 637 phage is influenced by the frequency of lysogenic establishment. *iScience* 27(4), 109595. doi: 638 10.1016/j.isci.2024.109595.
- Bailey, Z.M., Igler, C., and Wendling, C.C. (2024). Prophage maintenance is determined by 640 environment-dependent selective sweeps rather than mutational availability. *Curr Biol.* doi: 641 10.1016/j.cub.2024.03.025.
- Bernheim, A., and Sorek, R. (2018). Viruses cooperate to defeat bacteria. *Nature* 559(7715), 482-649 484. doi: 10.1038/d41586-018-05762-1.
- Bisen, M., Kharga, K., Mehta, S., Jabi, N., and Kumar, L. (2024). Bacteriophages in nature: recent 657 advances in research tools and diverse environmental and biotechnological applications. 658 *Environ Sci Pollut Res Int* 31(15), 22199-22242. doi: 10.1007/s11356-024-32535-3.
- Bonnet J, Yin P, Ortiz ME, Subsoontorn P, Endy D. Amplifying genetic logic gates. *Science* 2013;340:599–603.
- Bonnet, J., Subsoontorn, P., and Endy, D. (2012). Rewritable digital data storage in live cells via 666 engineered control of recombination directionality. *Proc Natl Acad Sci U S A* 109(23), 8884-667 8889. doi: 10.1073/pnas.1202344109.
- Brown WRA, Lee NCO, Xu Z, Smith MCM. Serine recombinases as tools for genome engineering. *Methods* 2011;53:372–379.
- Butt, L., Meyer, J.R., Lindsay, R.J., Beardmore, R.E., and Gudelj, I. (2024). Bacterial resistance 676 response and resource availability mediate viral coexistence. *J Evol Biol* 37(4), 371-382. doi: 677 10.1093/jeb/voae022.

- Casini A, Storch M, Baldwin GS, Ellis T. Bricks and blueprints: methods and standards for DNA assembly. *Nat Rev Mol Cell Biol* 2015;16:568–576. doi: 10.1038/nrm4014.
- Chibani-Chennoufi, S., Bruttin, A., Dillmann, M.L., and Brüssow, H. (2004). Phage-host interaction: 682 an ecological perspective. *J Bacteriol* 186(12), 3677-3686. doi: 10.1128/jb.186.12.3677-683 3686.2004.
- Clokie MR, Millard AD, Letarov AV, Heaphy S. Phages in nature. *Bacteriophage*. 2011 Jan;1(1):31-45. doi: 10.4161/bact.1.1.14942. PMID: 21687533; PMCID: PMC3109452.
- Cohen EEW, Bell RB, Bifulco CB, Burtness B, Gillison ML, Harrington KJ, Le Q-T, Lee NY, Leidner R, Lewis RL, Licitra L, Mehanna H, Mell LK, Raben A, Sikora AG, Uppaluri R, Whitworth F, Zandberg DP, Ferris RL. The Society for Immunotherapy of Cancer consensus statement on immunotherapy for the treatment of squamous cell carcinoma of the head and neck (HNSCC). *J Immunother Cancer* 2019;7(184). doi: 10.1186/s40425-019-0662-5).
- Doore, S.M., and Fane, B.A. (2016). The microviridae: Diversity, assembly, and experimental evolution. *Virology* 491, 45-55. doi: 10.1016/j.virol.2016.01.020.
- Fan, H.F., Hsieh, T.S., Ma, C.H., and Jayaram, M. (2016). Single-molecule analysis of φC31 703 integrase-mediated site-specific recombination by tethered particle motion. *Nucleic Acids Res* 704 44(22), 10804-10823. doi: 10.1093/nar/gkw861.
- Fogg, P.C., Colloms, S., Rosser, S., Stark, M., and Smith, M.C. (2014). New applications for phage 716 integrases. *J Mol Biol* 426(15), 2703-2716. doi: 10.1016/j.jmb.2014.05.014.
- Fogg, P.C.M., Haley, J.A., Stark, W.M., and Smith, M.C.M. (2017). Genome Integration and 715 Excision by a New Streptomyces Bacteriophage, φJoe. *Appl Environ Microbiol* 83(5). doi: 716 10.1128/aem.02767-16.
- Franco, R. A. L.; Brenner, G.; Zocca, V. F. B.; de Paiva, G. B.; Lima, R. N.; Rech, E. L.; Amaral, D. T.; Lins, M. R. C. R.; Pedrolli, D. B. Signal amplification for cell-free biosensors, an analog-to-digital converter. *ACS Synth. Biol.* 2023, 12 (10), 2819–2826, DOI: 10.1021/acssynbio.3c00227.

Freshney RI. Culture of animal cells: a manual of basic technique. New York: Wiley-Liss, 1994.

Gaj, T., Gersbach, C. A., and Barbas, C. F. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering Trends Biotechnol. 31, 397– 405 DOI: 10.1016/j.tibtech.2013.04.004.

Ghosh, P., Wasil, L.R., and Hatfull, G.F. (2006). Control of phage Bxb1 excision by a novel 731 recombination directionality factor. PLoS Biol 4(6), e186. doi: 10.1371/journal.pbio.0040186.

Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, Algire MA, Benders GA, Montague MG, Ma L, Moodie MM, Merryman C, Vashee S, Krishnakumar R, Assad-Garcia N, Andrews-Pfannkoch C, Denisova EA, Young L, Qi ZQ, Segall-Shapiro TH, Calvey CH, Parmar PP, Hutchison CA, Smith HO, Venter JC. Creation of a bacterial cell controlled by a chemically synthesized genome. Science 2010;329:52–56. doi: 10.1126/science.1190719.

Gomide, M.S., Sales, T.T., Barros, L.R.C., Limia, C.G., de Oliveira, M.A., Florentino, L.H., et al. 736 (2020). Genetic switches designed for eukaryotic cells and controlled by serine integrases. 737 Communications Biology 3(1), 255. doi: 10.1038/s42003-020-0971-8.

Grigson, S.R., Giles, S.K., Edwards, R.A., and Papudeshi, B. (2023). Knowing and Naming: Phage 739 Annotation and Nomenclature for Phage Therapy. Clin Infect Dis 77(Suppl 5), S352-s359. 740 doi: 10.1093/cid/ciad539.

Grindley, N.D., Whiteson, K.L., and Rice, P.A. (2006). Mechanisms of site-specific recombination. 742 Annu Rev Biochem 75, 567-605. doi: 10.1146/annurev.biochem.73.011303.073908.

Groth AC, Calos MP. Phage integrases: biology and applications. J Mol Biol 2004;335:667–678.

Groth, A.C., Fish, M., Nusse, R., and Calos, M.P. (2004). Construction of transgenic Drosophila by 744 using the site-specific integrase from phage phiC31. Genetics 166(4), 1775-1782. doi: 745 10.1093/genetics/166.4.1775.

- Groth, A.C., Olivares, E.C., Thyagarajan, B., and Calos, M.P. (2000). A phage integrase directs 747 efficient site-specific integration in human cells. *Proc Natl Acad Sci U S A* 97(11), 5995-748 6000. doi: 10.1073/pnas.090527097.
- Gupta, K., Sharp, R., Yuan, J.B., Li, H., and Van Duyne, G.D. (2017). Coiled-coil interactions 753 mediate serine integrase directionality. *Nucleic Acids Res* 45(12), 7339-7353. doi: 754 10.1093/nar/gkx474.
- Hatfull, G.F., and Hendrix, R.W. (2011). Bacteriophages and their genomes. *Curr Opin Virol* 1(4), 756 298-303. doi: 10.1016/j.coviro.2011.06.009.
- Kasman, L.M., and Porter, L.D. (2023). "Bacteriophages," in StatPearls. (Treasure Island (FL) 760 ineligible companies. Disclosure: La Donna Porter declares no relevant financial relationships 761 with ineligible companies.: StatPearls Publishing
- Keravala, A., Groth, A.C., Jarrahian, S., Thyagarajan, B., Hoyt, J.J., Kirby, P.J., et al. (2006). A 764 diversity of serine phage integrases mediate site-specific recombination in mammalian cells. 765 *Mol Genet Genomics* 276(2), 135-146. doi: 10.1007/s00438-006-0129-5.
- Li, H., Sharp, R., Rutherford, K., Gupta, K., and Van Duyne, G.D. (2018). Serine Integrase attP 767 Binding and Specificity. *J Mol Biol* 430(21), 4401-4418. doi: 10.1016/j.jmb.2018.09.007.
- Mandali, S., and Johnson, R.C. (2021). Control of the Serine Integrase Reaction: Roles of the Coiled-788 Coil and Helix E Regions in DNA Site Synapsis and Recombination. *J Bacteriol* 203(16), 789 e0070320. doi: 10.1128/jb.00703-20.
- McEwan, A.R., Raab, A., Kelly, S.M., Feldmann, J., and Smith, M.C. (2011). Zinc is essential for high-affinity DNA binding and recombinase activity of Φ C31 integrase. *Nucleic Acids Res* 39(14), 6137-6147. doi: 10.1093/nar/gkr220.
- McEwan, A.R., Rowley, P.A., and Smith, M.C. (2009). DNA binding and synapsis by the large C-794 terminal domain of phiC31 integrase. *Nucleic Acids Res* 37(14), 4764-4773. doi: 795 10.1093/nar/gkp485.
- Merrick, C.A., Zhao, J., and Rosser, S.J. (2018). Serine Integrases: Advancing Synthetic Biology. 797 ACS Synth Biol 7(2), 299-310. doi: 10.1021/acssynbio.7b00308.

- Mills, S., Shanahan, F., Stanton, C., Hill, C., Coffey, A., and Ross, R.P. (2013). Movers and shakers: 799 influence of bacteriophages in shaping the mammalian gut microbiota. *Gut Microbes* 4(1), 4-800 16. doi: 10.4161/gmic.22371.
- Murovec J, Pirc Ž, Yang B. New variants of CRISPR RNA-guided genome editing enzymes. *Plant Biotechnol J* 2017;15:917–926. doi: 10.1111/pbi.12736.
- Nielsen AAK, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, Ross D, Densmore D, Voigt CA. Genetic circuit design automation. *Science* 2016;352(6281). doi: 10.1126/science.aac7341.
- Olorunniji FJ, Rosser SJ, Stark WM. Site-specific recombinases: molecular machines for the Genetic Revolution. *Biochemical Journal* 2016;473(6):673–684. doi:10.1042/bj20151112.
- Olorunniji FJ, McPherson AL, Rosser SJ, Smith MCM, Colloms SD, Stark WM. Control of serine integrase recombination directionality by fusion with the directionality factor. *Nucleic Acids Res* 2017;45(14):8635-8645. doi: 10.1093/nar/gkx567.
- Patel, P.H., Taylor, V.L., Zhang, C., Getz, L.J., Fitzpatrick, A.D., Davidson, A.R., et al. (2024). Anti-823 phage defence through inhibition of virion assembly. *Nat Commun* 15(1), 1644. doi: 824 10.1038/s41467-024-45892-x. 825.
- Piel, D., Bruto, M., Labreuche, Y., Blanquart, F., Goudenège, D., Barcia-Cruz, R., et al. (2022). 826 Phage-host coevolution in natural populations. *Nat Microbiol* 7(7), 1075-1086. doi: 827 10.1038/s41564-022-01157-1.
- Pokhilko A, Ebenhöh O, Stark WM, Colloms SD. Mathematical model of a serine integrase-controlled toggle switch with a single input. *J R Soc Interface* 2018;15. doi: 10.1098/rsif.2018.0160.
- Ross W, Landy A. Bacteriophage lambda int protein recognizes two classes of sequence in the phage att site: characterization of arm-type sites. *Proc Natl Acad Sci* 1982;79(24):7724772-8.
- Rostøl, J.T., Quiles-Puchalt, N., Iturbe-Sanz, P., Lasa, I., and Penadés, J.R. (2024). Bacteriophages 829 avoid autoimmunity from cognate immune systems as an

intrinsic part of their life cycles. Nat 830 Microbiol. doi: 10.1038/s41564-024-01661-6.

Rubtsova M, Kempe K, Gils A, Ismagul A, Weyen J, Gils M. Expression of active Streptomyces phage φC31 integrase in transgenic wheat plants. Plant Cell Rep 2008;27:1821–1831. doi: 10.1007/s00299-008-0604-z.

Rutherford, K., Yuan, P., Perry, K., Sharp, R., and Van Duyne, G.D. (2013). Attachment site 834 recognition and regulation of directionality by the serine integrases. Nucleic Acids Res 41(17), 835 8341-8356. doi: 10.1093/nar/gkt580.

Rutherford, K., and Van Duyne, G.D. (2014). The ins and outs of serine integrase site-specific 829 recombination. Curr Opin Struct Biol 24, 125-131. doi: 10.1016/j.sbi.2014.01.003.

Sabino, J., Hirten, R.P., and Colombel, J.F. (2020). Review article: bacteriophages in 837 gastroenterology-from biology to clinical applications. Aliment Pharmacol Ther 51(1), 53-63. 838 doi: 10.1111/apt.15557.

Siedentop, B., Rüegg, D., Bonhoeffer, S., and Chabas, H. (2024). My host's enemy is my enemy: 843 plasmids carrying CRISPR-Cas as a defence against phages. Proc Biol Sci 291(2015), 844 20232449. doi: 10.1098/rspb.2023.2449.

Simmonds, P., and Aiewsakun, P. (2018). Virus classification - where do you draw the line? Arch 846 Virol 163(8), 2037-2046. doi: 10.1007/s00705-018-3938-z.

Smith, M.C.M. (2015). Phage-encoded Serine Integrases and Other Large Serine Recombinases. 848 Microbiol Spectr 3(4). doi: 10.1128/microbiolspec.MDNA3-0059-2014.

Snoeck, N., De Mol, M.L., Van Herpe, D., Goormans, A., Maryns, I., Coussement, P., et al. (2019). 850 Serine integrase recombinational engineering (SIRE): A versatile toolbox for genome editing. 851 Biotechnol Bioeng 116(2), 364-374. doi: 10.1002/bit.26854.

Stark, W.M. (2017). Making serine integrases work for us. Curr Opin Microbiol 38, 130-136. doi: 858 10.1016/j.mib.2017.04.006.

- Suttle CA. Viruses in the sea. *Nature*. 2005 Sep 15;437(7057):356-61. doi: 10.1038/nature04160. PMID: 16163346.
- Suttle, C.A. (2007). Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* 5(10), 860 801-812. doi: 10.1038/nrmicro1750.
- Turner, D., Adriaenssens, E.M., Lehman, S.M., Moraru, C., and Kropinski, A.M. (2024). 864 Bacteriophage Taxonomy: A Continually Evolving Discipline. *Methods Mol Biol* 2734, 27-865 45. doi: 10.1007/978-1-0716-3523-0_3.
- Van Duyne, G.D., and Rutherford, K. (2013). Large serine recombinase domain structure and 867 attachment site binding. *Crit Rev Biochem Mol Biol* 48(5), 476-491. doi: 868 10.3109/10409238.2013.831807.
- Wang Y, Yau YY, Perkins-Balding D, Thomson JG. Recombinase technology: applications and possibilities. *Plant Cell Rep* 2011;30(3):267–285. doi:10.1007/s00299-010-0938-1.
- Wang, J., Chitsaz, F., Derbyshire, M.K., Gonzales, N.R., Gwadz, M., Lu, S., et al. (2023). The 874 conserved domain database in 2023. *Nucleic Acids Res* 51(D1), D384-d388. doi: 875 10.1093/nar/gkac1096.
- Wang, S., Yang, Y., and Jing, J. (2022). A Synthesis of Viral Contribution to Marine Nitrogen 877 Cycling. *Front Microbiol* 13, 834581. doi: 10.3389/fmicb.2022.834581.
- Warwick-Dugdale, J., Buchholz, H.H., Allen, M.J., and Temperton, B. (2019). Host-hijacking and 882 planktonic piracy: how phages command the microbial high seas. *Virol J* 16(1), 15. doi: 883 10.1186/s12985-019-1120-1.
- Weinbauer, M.G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiol Rev* 28(2), 127-181. doi: 888 10.1016/j.femsre.2003.08.001.
- Weinberg BH, Hang Pham NT, Caraballo LD, Lozanoski T, Engel A, Bhatia A, Wong WW. Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat Biotechnol* 2017;35:453–462.
- Yang L, Nielsen AA, Fernandez-Rodriguez J, McClune CJ, Laub MT, Lu TK, Voigt CA. Permanent genetic memory with > 1-byte capacity. *Nat Methods* 2014;11:1261–1266. doi: 10.1038/nmeth.3147.

Ye H, Fussenegger M. Synthetic therapeutic gene circuits in mammalian cells. FEBS Lett 2014;588(15):2537-2544. doi: 10.1016/j.febslet.2014.05.003.

Young, R. (2013). Phage lysis: do we have the hole story yet? Curr Opin Microbiol 16(6), 790-797. 910 doi: 10.1016/j.mib.2013.08.008.

Zhu, X., Tang, L., Wang, Z., Xie, F., Zhang, W., and Li, Y. (2024). A comparative analysis of phage 924 classification methods in light of the recent ICTV taxonomic revisions. Virology 594, 110016. 925 doi: 10.1016/j.virol.2024.110016.

Anexo I

Publicação de artigo - coautoria

Genetic switches designed for eukaryotic cells and controlled by serine integrases

Mayna S. Gomide, Thais T. Sales, Luciana R. C. Barros, Cintia G. Limia, Marco A. de Oliveira, Lilian H. Florentino, Leila M. G. Barros, Maria L. Robledo, Gustavo P. C. José, Mariana S. M. Almeida, Rayane N. Lima, Stevens K. Rehen, Cristiano Lacorte, Eduardo O. Melo, André M. Murad, Martín H. Bonamino, Cintia M. Coelho & Elibio Rech

Communications Biology (2020).

DOI: doi.org/10.1038/s42003-020-0971-8



COMMUNICATIONS BIOLOGY

ARTICLE

<https://doi.org/10.1038/s42003-020-0971-8>

OPEN



Genetic switches designed for eukaryotic cells and controlled by serine integrases

Mayna S. Gomide ^{1,2,3}, Thais T. Sales^{1,2}, Luciana R. C. Barros ⁴, Cintia G. Limia ⁴, Marco A. de Oliveira^{1,2}, Lilian H. Florentino¹, Leila M. G. Barros ¹, Maria L. Robledo⁴, Gustavo P. C. José¹, Mariana S. M. Almeida¹, Rayane N. Lima¹, Stevens K. Rehen ^{5,6}, Cristiano Lacorte¹, Eduardo O. Melo^{1,7}, André M. Murad¹, Martín H. Bonamino ^{4,8✉}, Cintia M. Coelho ^{9✉} & Elibio Rech ^{1✉}

Recently, new serine integrases have been identified, increasing the possibility of scaling up genomic modulation tools. Here, we describe the use of unidirectional genetic switches to evaluate the functionality of six serine integrases in different eukaryotic systems: the HEK 293T cell lineage, bovine fibroblasts and plant protoplasts. Moreover, integrase activity was also tested in human cell types of therapeutic interest: peripheral blood mononuclear cells (PBMCs), neural stem cells (NSCs) and undifferentiated embryonic stem (ES) cells. The switches were composed of plasmids designed to flip two different genetic parts driven by serine integrases. Cell-based assays were evaluated by measurement of EGFP fluorescence and by molecular analysis of *attL/attR* sites formation after integrase functionality. Our results demonstrate that all the integrases were capable of inverting the targeted DNA sequences, exhibiting distinct performances based on the cell type or the switchable genetic sequence. These results should support the development of tunable genetic circuits to regulate eukaryotic gene expression.

¹ Brazilian Agriculture Research Corporation – Embrapa – Genetic Resources and Biotechnology – CENARGEN, Brasília 70770917 DF, Brazil. ² Department of Cell Biology, Institute of Biological Science, University of Brasília, Brasília 70910900 DF, Brazil. ³ School of Medicine, Federal University of Juiz de Fora, Juiz de Fora 36036900 MG, Brazil. ⁴ Molecular Carcinogenesis Program, Research Coordination, National Cancer Institute (INCA), Rio de Janeiro 20231050 RJ, Brazil. ⁵ D’Or Institute for Research and Education (IDOR), Rio de Janeiro 22281100 RJ, Brazil. ⁶ Institute of Biomedical Sciences, Federal University of Rio de Janeiro, Rio de Janeiro 21941902 RJ, Brazil. ⁷ Graduation Program in Biotechnology, Federal University of Tocantins, Gurupi 77402970 TO, Brazil. ⁸ Vice-Presidency of Research and Biological Collections (VPPCB), FIOCRUZ – Oswaldo Cruz Foundation Institute, Rio de Janeiro 21040900 RJ, Brazil. ⁹ Department of Genetic and Morphology, Institute of Biological Science, University of Brasília, Brasília 70910900 DF, Brazil. ✉email: mbonamino@inca.gov.br; cintia.coelhom@unb.br; elibio.rech@embrapa.br

Recombinant DNA technology was a landmark for the development of cells that are able to perform several tasks in different areas of research, such as therapy¹, diagnosis², biosensing³, bioremediation⁴, and plant and animal genetic improvement^{5–7}. Over the last decade, the desirable traits were mostly monogenic, with transgene expression being mainly transcriptionally controlled by tissue/development-specific or chemically/physically inducible promoters, acting as activators or repressors separately or in loops^{8,9}. Currently, novel approaches in synthetic biology are being used for the design of new molecular entities that would lead to the creation and/or improvement of pathways, reprogramming organisms for new tasks^{10,11}. This scenario demands the development of tools to finely tune polygenic expression.

Large serine-type phage integrases have been described as useful genomic tools for gene manipulation^{12–17}. These proteins belong to a superfamily of site-specific serine recombinases, which differ in many aspects from their counterparts, the tyrosine recombinases. The latter are bidirectional proteins, which is a major drawback to their use to control gene expression due to the instability of the on/off end product¹⁸. Instead, the phage-encoded serine integrases (Ints) are capable of unidirectional recombination that leads to permanent DNA fragment inversion, which ultimately could be used to modulate gene expression¹⁹. To perform this function, Ints recognize specific attachment sites, named *attB* and *attP*²⁰. When both sites are placed in opposite directions flanking a genetic part (such as a promoter, coding sequence and/or terminator), Ints perform a 180° flip of these genetic parts, which can turn gene expression on/off. As a result of this recombination event, the attachment sites are converted into different sequences, called *attL* and *attR*²¹. This recombination process can only be reversed in the presence of a cognate protein, called recombination directionality factor (RDF)¹³.

Ints have been previously used to build logic gates based on Boolean algebra to regulate gene expression in prokaryotic organisms^{22,23}. In these reports, the authors utilized Ints to rotate promoters or terminators, hence controlling RNA polymerase flow in a way resembling an electronic transistor, now called a transcriptor²². Most importantly, GFP expression for every Int used was observed as predicted by the truth table of each designed logic gate that was evaluated^{22,23}. Another strategy used to regulate prokaryotic gene expression was recombinase-based state machines, a system in which several combinations of integrase inputs that inverted or excised genetic parts produced different outputs²⁴. However, in eukaryotic organisms, there is a scarcity of effective tools that allow broad and precise gene regulation, which is an essential requirement for multiplex gene control. To date, only a limited number of Ints have been successfully tested in eukaryotic cells, showing a restricted cell-type-dependent functionality^{25–28}. Nevertheless, most of these studies aimed to integrate or excise a DNA fragment from the genome of a particular organism, rather than using the Ints as regulators of gene expression^{28–35}. In a different approach, Weinberg et al.²⁶ used a combination of four tyrosine and two serine recombinases in mammalian cells to develop a six-input/one-output Boolean logic lookup table. However, the relatively small number of functional Ints characterized in eukaryotic cells restricts the scale-up of the use of these proteins to build genetic circuits.

Recently, Yang et al.¹⁸ identified 34 putative Ints, showing functionality for 11 of these Ints in prokaryotic cells. In an attempt to fill this existing gap in eukaryotic cells, we built unidirectional genetic switches to evaluate the functionality of six out of these 11 Ints that have already been tested in prokaryotic cells. To this end, we chose three different model systems: the human embryonic kidney cell lineage (HEK 293T), bovine primary fibroblasts, and *Arabidopsis thaliana* protoplasts. In addition, as a

medically important proof of concept, unidirectional genetic switches using these Ints were also evaluated in peripheral blood mononuclear cells (PBMCs), neural stem cells (NSCs) differentiated from induced pluripotent stem cells, and undifferentiated human embryonic stem cells (hES, BR-1 cell line).

Here we report that all tested Ints are functional genetic switch controllers, activating the coding sequence or the promoter switches designed to be turned on in the eukaryotic cells. The frequency of cells emitting the reporter fluorescence varies among the tested integrases and cell types, and, in some cases, the switch activation is proven by molecular tests. In addition, Ints show accuracy in their site recognition and recombination process, and are not cytotoxic for the cell models assayed. These data put the evaluated Ints as suitable candidates to regulate gene expression in wide synthetic genetic networks that now can be built for several eukaryotic organisms.

Results

Unidirectional genetic switches design. In our system, the HEK 293T lineage, bovine primary fibroblasts, and *A. thaliana* protoplasts were chosen as human, nonhuman mammal and plant models, respectively. Ints 2, 4, 5, 7, 9, and 13 evaluated in *Escherichia coli* by Yang et al.¹⁸ were selected to be tested in eukaryotic switch systems. The phiC31 and Bxb1 integrases were also evaluated because they were shown to be functional in genetic switches in human cells²⁶ and plants³² (only phiC31). We designed unidirectional genetic switches composed of two sets of synthesized plasmids. The first set contains either a human or plant-optimized (*A. thaliana*) coding sequence of an Int under the control of a strong species-specific constitutive promoter, named integrase expression vectors (set 1, pIE). The second set contains the reporter gene *egfp* in the reverse complement orientation flanked by the recognition sites *attB* and *attP* of the corresponding Int under distinct constitutive promoters, for either plant or animal systems. The resulting plasmids were called switch GFP vectors (set 2, pSG) (Fig. 1; Supplementary Fig. 1; Supplementary Methods). Therefore, eight plasmids were generated for each vector set of the mammalian systems, and the same number was generated for the plant system (Supplementary Table 1). One plasmid of set 1 and one plasmid of set 2 were used to transiently cotransfect mammalian cells and cotransform plant protoplasts (test condition). The negative control cells were cotransfected/cotransformed with only one of the two plasmids plus a mock plasmid to keep the DNA concentration constant. As a positive control (pGFP), the cells were cotransfected/cotransformed with a plasmid containing *egfp* in forward orientation under the same constitutive promoter as that in the plasmids from set 2 plus a mock plasmid. HEK 293T control cells were not cotransfected with the mock plasmid. In these transient assays many copies of both plasmids for each test or control conditions are inserted in the cells, according to the concentrations described in the Methods section.

The rationale for this system was that if a particular Int is functional, it would switch the *egfp* coding sequence to the forward orientation, leading to *egfp* expression and the formation of the *attL* and *attR* sites, referred to as the activated switch vector (Fig. 1; Supplementary Fig. 1). The six Ints were chosen because their *attL* sites formed after flipping did not lead to the formation of any ATG upstream of the *egfp* ORF. Although there was not ATG formation in its *attL*, Int 10 was not selected to be evaluated because its recognition sites were recognized by other integrases when evaluated in bacterial cells¹⁸. All assays were evaluated at 48 h for mammalian cells and 24 h for *A. thaliana* protoplasts, corresponding to the maximum EGFP-positive cell frequency.

No nuclear localization signal (NLS) was added to the Ints since previous studies showed contradictory results regarding the

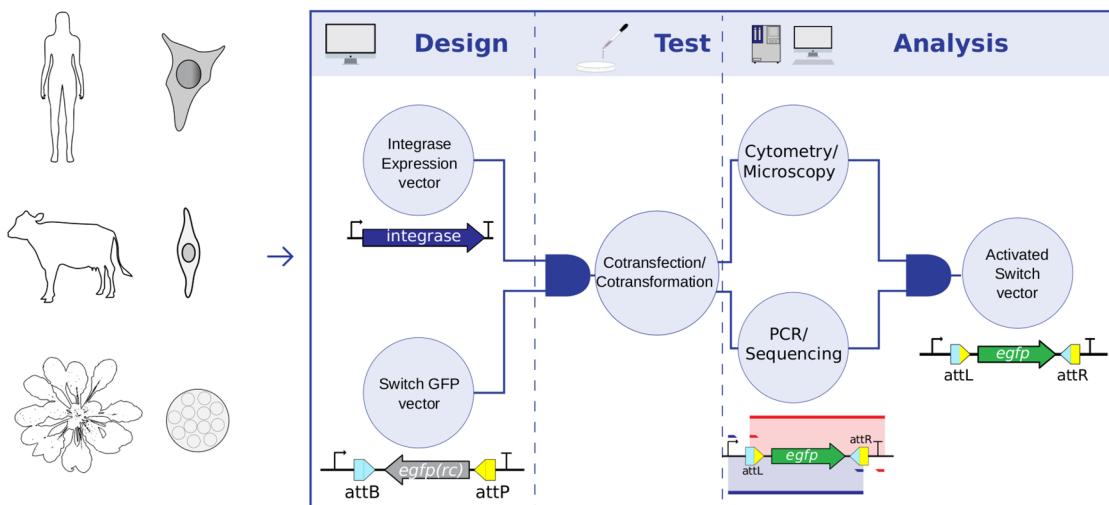


Fig. 1 Strategy overview of the eukaryotic genetic switch system. The human cell lineage HEK 293T, bovine fibroblasts and *A. thaliana* protoplasts were the selected model systems. The first step involved the design of two plasmid sets: the integrase expression vectors to express Ints 2, 4, 5, 7, 9, 13, phiC31, or Bxb1 and the switch GFP vectors with the *attB* and *attP* recognition sites of the respective Int flanking an *egfp* coding sequence in a reverse complement (rc) orientation. Acting as a schematic AND gate, combination of the corresponding plasmids of each of the vector sets results in the second step in the process, the test, accomplished by cotransfection or cotransformation assays of mammalian and plant cells, respectively. The third and last step led to the development of analytical methods that include the inputs of an additional schematic AND gate. Microscopy/flow cytometry analyses were used to detect EGFP fluorescence in cells resulting from the flipping action of the integrase. PCR/sequencing was used in the analysis of the *egfp* coding sequence rotated to the correct forward orientation flanked by the formed *attL* and *attR* sites. Both analytical inputs provide evidence of the activated switch vector output. The PCRs used one primer pair to amplify the complete *attL* site and the entire *egfp* coding sequence, now in the forward orientation (blue), and a second primer pair to amplify the complete *attR* site and the *egfp* sequence (red).

addition of an NLS at the N- and/or C-termini of these proteins^{17,25}. In addition, in silico analysis demonstrated the existence of potential cryptic NLSs in the Ints evaluated in this study with various scores, except for Int 5, for which no NLS was predicted (Supplementary Fig. 2).

Functional characterization of Ints as switch controllers. In the mammalian groups, EGFP-expressing cells were detected among HEK 293T cells cotransfected with pIE + pSG vectors corresponding to Int 13, phiC31, or Bxb1 (Fig. 2a, b; Supplementary Figs. 3 and 4) and in bovine fibroblasts cotransfected with pIE + pSG vectors corresponding to Int 9, 13, or Bxb1 (Fig. 3a, b; Supplementary Figs. 3 and 5). In addition, low levels of cells emitting EGFP fluorescence could also be observed in tests activated by Ints 2, 4, and 5 in HEK 293T cells, and by Ints 2 and phiC31 in bovine fibroblasts (Figs. 2b, 3b; Supplementary Figs. 4 and 5). In general, flow cytometry analysis showed that the abundance of EGFP-positive cells ranged from 0.03 to 16.02% for the tested switches in these cell groups, indicating different levels of Int functionality (Figs. 2b and 3b). In HEK 293T cells, Kruskal–Wallis statistics corroborated that the percentages of EGFP-positive cells cotransfected with pIE + pSG corresponding to Ints 13, phiC31, and Bxb1 resulted in significant differences compared with the negative controls ($p = 1.71 \times 10^{-5}$, 1.10×10^{-4} , and 1.50×10^{-4} , respectively; Fig. 2b; Supplementary Table 2). In bovine fibroblasts, in addition to Ints 9, 13, and Bxb1 ($p = 1.10 \times 10^{-6}$, 2.04×10^{-6} , and 9.18×10^{-7} , respectively), cells cotransfected with pIE + pSG corresponding to Ints 2, 5, and phiC31 were also significantly different from the negative controls ($p = 1.87 \times 10^{-6}$, 3.08×10^{-5} , and 1.61×10^{-6} , respectively; Fig. 3b; Supplementary Table 2). Analyzing only the herein tested integrases, Int 13 led to the highest number of EGFP-positive cells in both models evaluated. The positive cell frequencies in the Int 13 tests were 33% and 16% compared with the positive controls (pGFP) for HEK 293T cells and bovine fibroblasts, respectively.

To confirm Int functionality by targeted *egfp* sequence rotation and correct *attL/attR* sites formation, PCR and sequencing analysis were performed. To this end, two pairs of primers were used. One pair annealed to the promoter sequence (forward) and to the *attR* site (reverse) (Fig. 1, blue color), and the other pair annealed to the *attL* site (forward) and to a region next to the terminator sequence (reverse) (Fig. 1, red color). Thus, PCR amplifications are expected in only activated pSG vectors. Interestingly, although EGFP fluorescence was not detected through microscopy and flow cytometry analysis of cells tested with some of the Ints, the PCR and sequencing analysis of DNA extracted from tested cells confirmed the predicted *attL/attR* sites for Ints 2, 4, 5, 7, 9, 13, phiC31, and Bxb1 and that the *egfp* had flipped to the forward orientation (Figs. 2c, 3c; Supplementary Figs. 6–9). These results confirm that all the evaluated Ints were functional, however, their efficiency may set a threshold for detection of EGFP fluorescence by flow cytometry or microscopy analysis.

An important limiting factor to the use of Ints in genetic switches is their potential cytotoxicity. Thus we measured the viability of mammalian cells after 48 h of the integrase activity using MTT assay. Neither HEK 293T cells nor fibroblasts showed marked viability impairment comparing with control groups (Figs. 2d and 3d).

For *A. thaliana* protoplasts, EGFP-expressing cells were detected after cotransformation with the pIE + pSG vectors corresponding to Int 2, 4, 7, 9, 13, phiC31, or Bxb1 (Fig. 4a, b; Supplementary Figs. 3 and 10). All Int tests led to a statistically significant frequency of EGFP-positive cells compared with the respective negative controls, although Int 5 led to a very low percentage of EGFP-expressing cells (p -value for Ints 2, 4, 5, 7, 9, 13, phiC31, and Bxb1 groups were 2.57×10^{-11} , 2.22×10^{-7} , 2.28×10^{-6} , 2.45×10^{-9} , 7.24×10^{-7} , 2.01×10^{-8} , 1.28×10^{-6} , and 1.07×10^{-6} , respectively; Fig. 4b; Supplementary Table 2). As observed for HEK 293T cells and bovine fibroblasts, the flow cytometry data showed a variable number of cells expressing

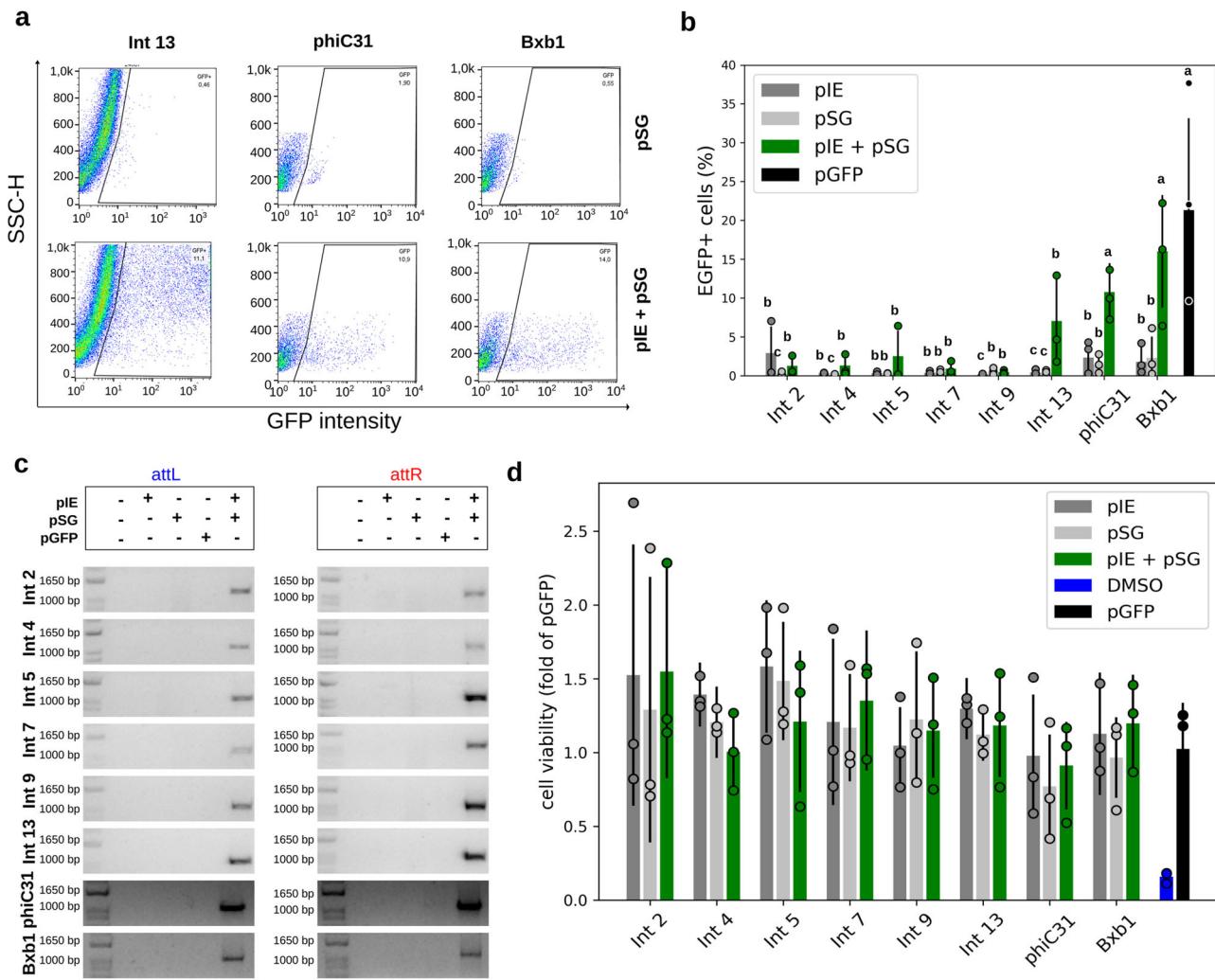


Fig. 2 Functional characterization of the genetic switches in human cells. **a** Flow cytometry distribution of HEK 293T cells at 48 h post transfection for Int 13, phiC31, and Bxb1, the integrases that led to the highest EGFP-expressing cell frequencies. The heat map indicates the scattering of high cell concentrations (warm colors) to low cell concentrations (cool colors). The gate encompasses the EGFP-expressing cell population. **b** Bar graph plots showing the total average percentage and standard deviation of a cell population expressing EGFP in biological repeat assays ($n = 3$) and circles showing the technical duplicate or triplicate average of each assay on the y axis. The x axis contains the different conditions. For each Int data group, different letters indicate significant differences ($p < 0.05$). **c** Amplicons obtained through PCR analysis using two specific primer sets, the first set to verify *attL* formation and the second set to verify *attR* (highlighted in Fig. 1). The expected amplicon sizes in the Int test groups varied from 1021 to 1104 bp for *attL* and from 1058 to 1084 bp for *attR*. **d** Bar graph plots showing the viable cells average (circles corresponding to technical replicates averages) and standard deviation normalized with pGFP of OD measurements obtained after MTT assays ($n = 3$). DMSO corresponds to the impairment negative control. Negative control cells were transfected with only one of the two vector sets, that is, integrase expression (pIE) or switch GFP (pSG) vectors. Positive control cells (pGFP) have an *egfp* sequence in the forward orientation under the control of the EF1 alpha promoter. All the data are representative of two or three technical and three biological replicates.

EGFP. The frequency of EGFP-positive cells ranged from 0.1% for Int 5 to 24.2% for Int 13 (Supplementary Table 2). This highest value showed Int 13 to be superior to phiC31 and Bxb1 in the induction of EGFP activation. Notably, overall, plant protoplasts resulted in higher numbers of EGFP-accumulating cells than the mammalian systems tested, indicating robustness in the vegetal system (Figs. 2b, 3b, and 4b; Supplementary Table 2). In this system, the pCaMV35S-GFP plasmid, with known EGFP expression efficiency, was used as a positive control³⁶. The cauliflower mosaic virus (CaMV) 35S promoter of this plasmid (alignment with GenBank V00140.1) has some single-nucleotide polymorphisms (SNPs) compared with the CaMV 35S promoter chosen for the plant switch vector syntheses (iGEM registry BBa_K1547006 and alignment with GenBank V00141.1). These

SNPs, however, did not result in statistically significant differences in the frequency of EGFP-expressing cells ($p = 0.1277$; Supplementary Fig. 11).

In addition, PCR and sequencing analysis, following the same strategy previously described for the mammalian cells, showed *attL/attR* sites formation and *egfp* flipping to the forward orientation for all Ints evaluated (Fig. 4c; Supplementary Figs. 12 and 13), even for the Int 5 test condition, in which EGFP-positive cells were detected at a very low level by flow cytometry analysis.

The cell viability in protoplasts after Ints activity was measured using fluorescein diacetate (FDA) hydrolysis assay. Also, as observed in mammalian cells, in protoplasts no marked cell impairment was observed after 24 h of the integrases transformation (Fig. 4d).

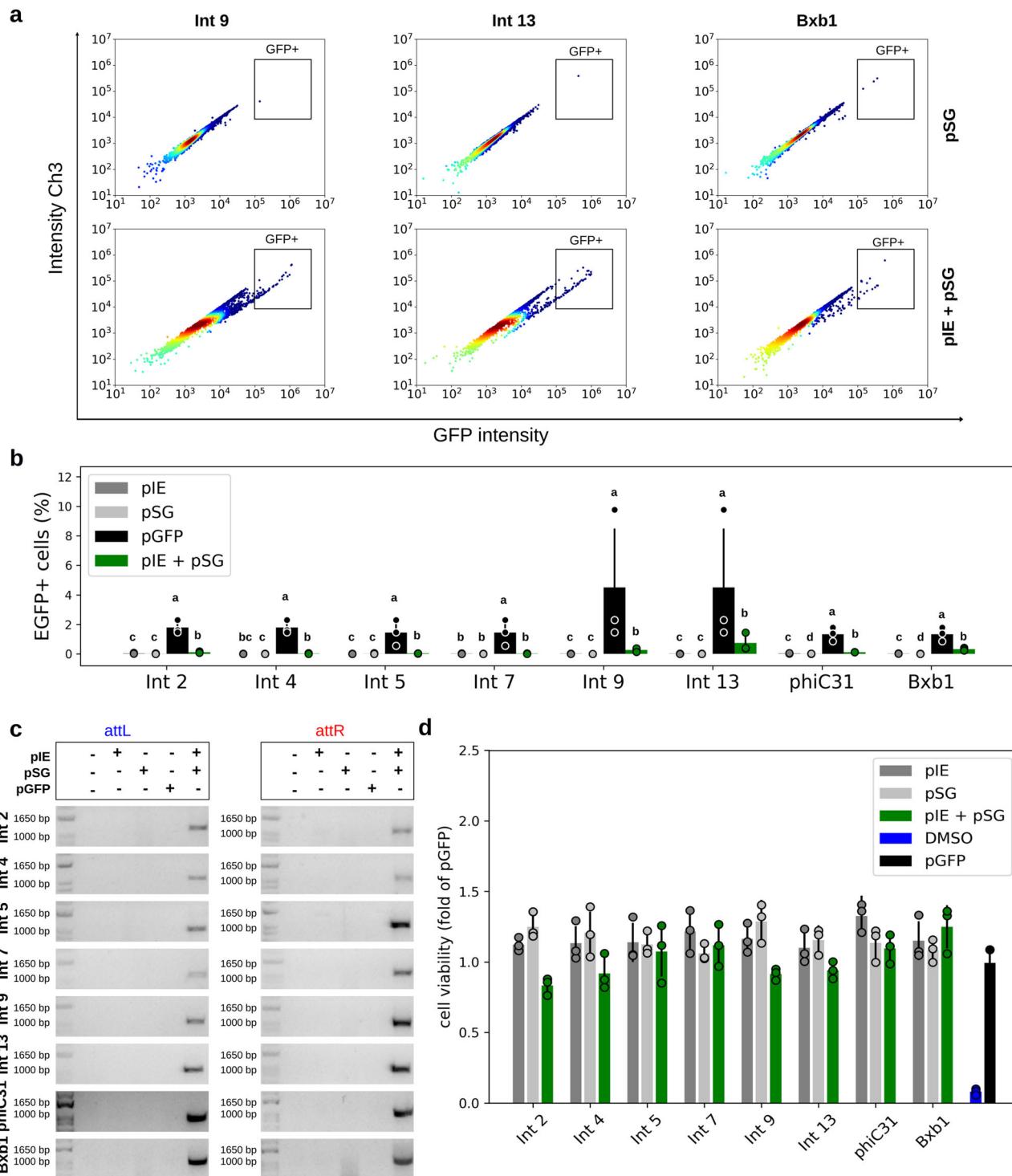


Fig. 3 Functional characterization of the genetic switches in bovine cells. **a** Flow cytometry distribution of bovine fibroblasts at 48 h post cotransfection for Ints 9, 13, and Bxb1, the integrases that led to the highest EGFP-expressing cell frequencies. The heat map indicates the scattering of high cell concentrations (warm colors) to low cell concentrations (cool colors). The gate encompasses the EGFP-expressing cell population. **b** Bar graph plots showing the total average percentage and standard deviation of a cell population expressing EGFP in biological repeat assays ($n=3$) and circles showing the technical triplicate average of each assay on the y axis. The x axis contains the different conditions. For each Int data group, different letters indicate significant differences ($p < 0.05$). **c** Amplicons obtained through PCR analysis using two specific primer sets, the first set to verify *attL* formation and the second set to verify *attR* (highlighted in Fig. 1). The expected amplicon sizes in the Int test groups varied from 1021 to 1104 bp for *attL* and from 1058 to 1084 bp for *attR*. **d** Bar graph plots showing the viable cells average (circles corresponding to technical replicates averages) and standard deviation normalized with pGFP of OD measurements obtained after MTT assays ($n=3$). DMSO corresponds to the impairment negative control. Negative control cells were transfected with only one of the two vector sets, that is, integrase expression (pIE) or switch GFP (pSG) vectors, plus a mock plasmid. Positive control cells (pGFP) have an *egfp* sequence in the forward orientation under the control of the EF1 alpha promoter. All the data are representative of three technical and three biological replicates.

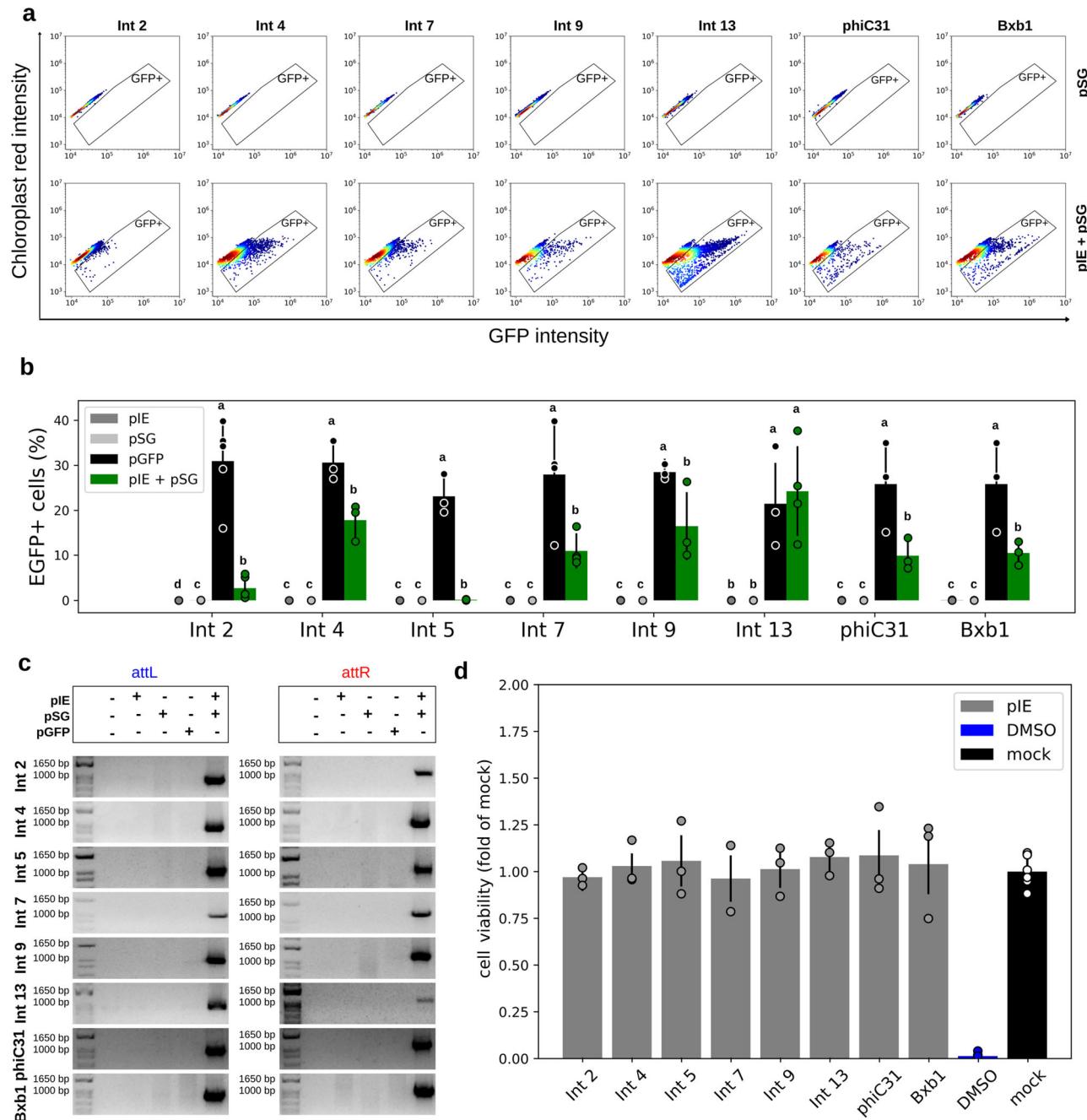


Fig. 4 Functional characterization of the genetic switches in plant protoplasts. a Flow cytometry distribution of protoplasts at 24 h post cotransformation for Ints 2, 4, 7, 9, 13, phiC31, and Bxb1, the integrases that led to the highest EGFP-expressing cell frequencies. The heat map indicates the scattering of high cell concentrations (warm colors) to low cell concentrations (cool colors). The gate encompasses the EGFP-expressing cell population. **b** Bar graph plots showing the total average percentage and standard deviation of a cell population expressing EGFP in biological repeat assays (n in Supplementary Table 2) and circles showing the technical triplicate average of each assay on the y axis. The x axis contains the different conditions. For each Int data group, different letters indicate significant differences ($p < 0.05$). **c** Amplicons obtained through PCR analysis using two specific primer sets, the first set to verify *attL* formation and the second set to verify *attR* (highlighted in Fig. 1). The expected amplicon sizes in the Int test groups varied from 948 to 983 bp for *attL* and from 1136 to 1132 bp for *attR*. Negative control cells were transfected with only one of the two vector sets, that is, integrase expression (pIE) or switch GFP (pSG) vectors, plus a mock plasmid. Positive control cells (pGFP) have an *egfp* sequence in the forward orientation under the control of the CaMV 35S promoter. All the data are representative of three technical and three or more biological replicates. **d** Bar graph plots showing the viable cells average (circles corresponding to technical replicates averages) and standard deviation normalized with mock plasmid positive cells obtained after FDA assays. DMSO corresponds to the impairment negative control. The cells were transfected with pIE vectors in technical triplicates and three biological replicates (except for Int 7, for which there was three technical and two biological replicates).

Another important concern related to the use of Ints is the accuracy with which the Ints mediate the site-specific recombination. Therefore, the individual sequence clone reads, obtained from PCR amplicons, were aligned with the expected flipping plasmid sequence, trimmed, and analyzed for all three cell model systems evaluated. In addition to confirming Int functionality, DNA sequencing demonstrated minimal errors in the recombinant *attL/attR* sites, indicating that these proteins are not error prone. For only Int 4 in the bovine fibroblast system and Int 7 in the protoplast system, one covered SNP was observed at the *attR* sites. However, mutations were also observed in the *egfp* sequence for most clones sequenced, suggesting that the observed mutations may be due to PCR errors instead of imprecise Int recombination (Supplementary Table 3; Supplementary Data 1, 2, 3).

Sequencing analyses also evidenced a one-base difference for the Int 9 *attL/attR* attachment sites compared with the original described by Yang et al.¹⁸ (Supplementary Fig. 14). These data indicate that the adenine after the previously marked *attB* core and a thymine after the *attP* should be part of the Int 9 crossover site. However, this arrangement would lead to a one-base difference between the *attB/attP* cores, making it impossible for dinucleotide overhang recombination to occur after integrase subunit rotation³⁷. Thus, we found a possible arrangement with the next CT dinucleotide as the core for the Int 9 attachment sites.

Promoter as switchable genetic part and Ints orthogonality. We wanted to evaluate whether a different genetic component could be used as a switchable part and if these proteins were orthogonal. To this end, *A. thaliana* protoplast was chosen as a model organism, since we observed a high number of cells expressing EGFP in this system, indicating robust Int activity. Three Ints with variable levels of efficiency were selected: Ints 2, 4, and 5. Although Int 13 led to the highest *egfp* activation in protoplasts, it was not selected because there is an ATG start codon in the *attR* site formed after recombination, what would remove the *egfp* gene sequence frame, and because Yang et al.¹⁸ reported a constitutive promoter activity for Int 13 *attP* site.

In this proposal, the switch GFP construct (set 2 plasmid) was redesigned, placing the constitutive promoter, CaMV 35S, in the reverse complement orientation flanked by the *attB* and *attP* sites of Ints 2, 4, and 5 in tandem, followed by the *egfp* sequence in the forward orientation; this construct was named the switch promoter vector (pSP) (Fig. 5a). Flow cytometry data showed 32.2%, 38.6%, and 12.5% EGFP-positive cells in the protoplast population cotransformed with the pIE Int vectors 2, 4, and 5 + pSP, respectively (Fig. 5b, c, Supplementary Figs. 3 and 15). Residual EGFP accumulation on the pSP negative control cells was observed; however, all three Int tests exhibited statistically significant differences compared with the negative control groups ($p = 1.51 \times 10^{-11}$; Supplementary Table 4). Notably, a higher EGFP-positive cell frequency was observed for the Int 2 test than for the flipping strategy with the switch GFP vector.

PCR and sequencing analysis corroborated these results, showing *attL/attR* sites formation and CaMV 35S flipping, driving the RNA polymerase to transcribe the *egfp* gene (Fig. 5d; Supplementary Fig. 16). Because we used primer sets flanking the recognition sites of Ints 2, 4, and 5 (Fig. 5a, activated switch promoter vector), we could observe that each Int only recognized its own *attB/attP* sites (Supplementary Fig. 17), with just one covered deletion observed at the formed *attL* site of Int 5 (Supplementary Table 5, Supplementary Data 4). These results indicate that Ints 2, 4, and 5 are accurate and orthogonal. In conclusion, the promoter sequence can also be used to build genetic switches.

Ints activity in primary human T lymphocytes and stem cells. As a proof of concept, unidirectional genetic switches using integrase expression (pIE) and switch GFP (pSG) vectors were tested in human peripheral blood mononuclear cells (PBMCs), the main source of T lymphocytes. These cells were chosen due to their significance in several different areas of health research, such as vaccine development, hematological malignancies, high-throughput screening, and immunology. Recently, T lymphocytes have gained importance in cancer immunotherapies^{38,39} through their clinical use with or without genetic manipulation. Considering the results obtained for HEK 293T cells, we selected Int 13, an integrase with highly detectable functionality, and Int 4, which showed a low frequency of EGFP-expressing cells, to be tested with PBMCs; phiC31 and Bxb1 were also evaluated. Flow cytometry analysis showed that the Int 13 and Bxb1 tests resulted in the same EGFP-expressing cell frequency (7%) in PBMCs extracted from three healthy independent donors (Fig. 6a, b; Supplementary Fig. 18). These results indicate that Int 13 and Bxb1 were able to promote the inversion of the *egfp* coding sequence to the forward orientation, leading to that high EGFP-expressing cells frequency, equivalent to 87.5% of the result observed in the positive control population (pGFP).

PCR and sequencing analysis performed using the same strategy previously described for the other switch GFP assays corroborated these data and expanded the data, showing that Ints 4 and phiC31 were also capable of flipping the *egfp* coding sequence (Fig. 6c; Supplementary Figs. 19 and 20). The DNA sequence reads obtained were also aligned with the activated plasmid expected sequence, and the covered mutations were counted. Only a few SNPs were observed at the formed *attL/attR* sites (Supplementary Table 6, Supplementary Data 5).

In addition, this study aimed to investigate Int functions in another human cell type used as an efficient in vitro model in studies on several diseases and embryogenesis. Switch systems using integrase expression vectors (pIE Ints 2, 9, 13, phiC31, and Bxb1) and the respective switch GFP (pSG) vectors were thus evaluated in neural stem cells (NSCs), and undifferentiated human embryonic stem (hES) cells. The plasticity of these cells has made them the focus of basic developmental research and also of the challenging regenerative medicine field. As an example, NSCs exhibit promise in the treatment of neurodegenerative diseases⁴⁰. The hES-BR-1 cell line was the first hES cell line established from the Brazilian population and is a relevant model for stem cell assays⁴¹. The flow cytometry measurements were within a narrow range, with 2.0–3.9% EGFP-positive cells observed in the Int 2, 9, and 13 tests (Fig. 7), indicating that these Ints are functional in these relevant disease modeling systems. PhiC31 and Bxb1 tests exhibited the highest values, showing activity in between 9.6 and 24.8% of EGFP-positive cells (Fig. 7).

Discussion

Although tools based on activators and repressors, including recently modified CRISPR-Cas9 systems⁴², are used to regulate eukaryotic gene expression, there remains a need for new technologies that are scalable and precise for multiplex gene regulation control. In our study, through eukaryotic cell-based assays, we demonstrated the widespread use of six Ints 2, 4, 5, 7, 9, and 13 as genetic switchers in mammalian and plant cells. All the Ints tested were able to (i) recognize their respective predicted *attB/attP* sites, (ii) generate the predicted *attL/attR* sites, and (iii) invert the *egfp* or promoter sequences. Furthermore, our results demonstrated that these six Ints have different degrees of functionality depending on the eukaryotic cell type or the genetic part to be flipped, potentially leading to a tunable system. Of the six Ints evaluated, the flow cytometry analysis showed that Int 13 led

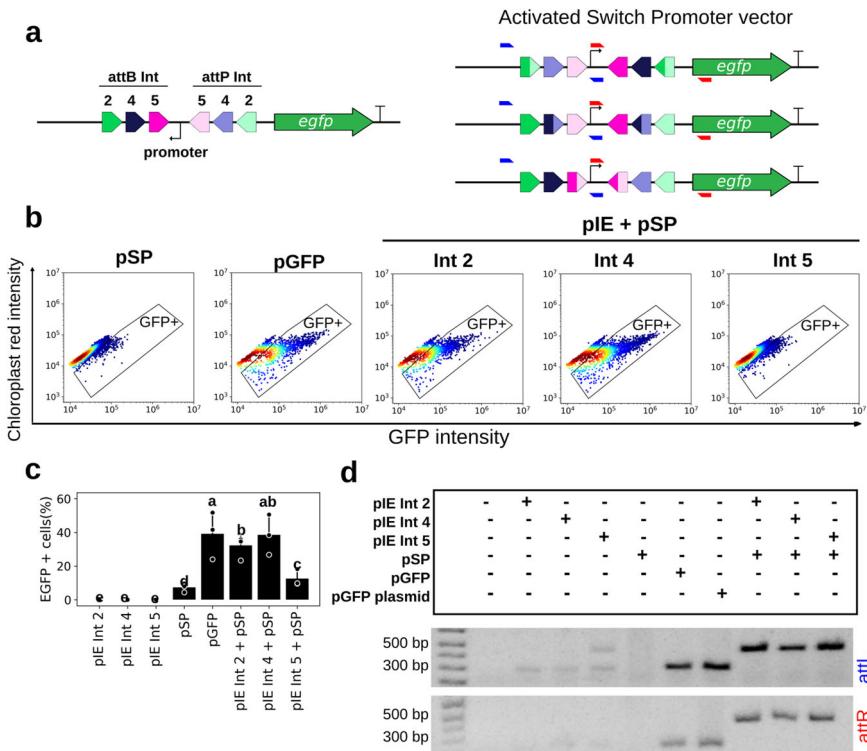


Fig. 5 Promoter as a switchable genetic part in plant protoplasts. **a** Schematic representation of the CaMV 35S promoter as a switchable genetic part. In this case, the switch vector was redesigned to contain the promoter in the reverse complement sequence orientation flanked by the *attB* and *attP* sites of three different Ints (2, 4, and 5) in tandem; this construct was named the switch promoter vector (pSP). The integrase expression vectors (pIE) for each Int were cotransformed separately with the pSP, and each resulting activated promoter vector is shown. **b** Flow cytometry distribution of protoplasts at 24 h post cotransformation. The heat map indicates the scattering of high cell concentrations (warm colors) to low cell concentrations (cool colors). The gate encompasses the EGFP-expressing cell population. **c** Bar graph plots showing the total average percentage and standard deviation of a cell population expressing EGFP in biological repeat assays ($n = 3$) and circles showing the technical triplicate average of each assay on the y axis. The x axis contains the different conditions. Different letters indicate significant differences ($p < 0.05$). **d** Amplicons obtained through PCR analysis using two specific primer sets. The first primer set (blue) was complementary to the pSP vector backbone sequence (forward) and promoter sequence (reverse). Expected amplicon size: 429 bp. The second primer set (red) was complementary to the promoter sequence (forward) and *egfp* coding sequence (reverse). Expected amplicon size: 438 bp (primers are colored and marked in letter a). Negative control cells were cotransformed with one of the integrase expression vectors (pIE) plus a mock plasmid or with the switch promoter vector (pSP) plus a mock plasmid. Positive control cells (pGFP) have a plasmid containing the *egfp* sequence under the control of the CaMV 35S promoter in the forward orientation plus a mock plasmid. Expected amplicon size for pGFP: *attL* gel: 285 bp; *attR* gel: 227 bp. These amplicons were smaller than those obtained under the test conditions due to the absence of the Int attachment sites. All the data are representative of three technical and three biological replicates.

to the highest proportion of EGFP-expressing cells in all three eukaryotic model systems. Int 13 (NCBI No. WP_012095429.1) is a protein with 55.5 kDa in a range of 54.0–67.1 kDa among the eight integrases evaluated, and it was identified from a prophage inside *Bacillus cytotoxicus* NVH 391-98^{18,43}. Int 9 for bovine fibroblasts and Ints 4 and 9 for plant protoplasts were set at the second position. PhiC31 and Bxb1, characterized elsewhere, yielded the highest EGFP percentage values only for HEK 293T cells, slightly overcoming Int 13 activity. On the other hand, for bovine fibroblasts, Bxb1 exhibited activity close to that of Int 9, and phiC31 showed low activity. In protoplasts, these integrases promoted an intermediate effect, similar to that of Int 7. Ints 9, 4, and 13 yielded higher percentages of EGFP-positive cells than phiC31 and Bxb1 in this plant model. Moreover, interestingly, in the study conducted by Yang et al.¹⁸ in bacteria, Ints 2, 4, 5, 7, and 13 yielded 100% GFP-positive cell populations and Int 9 yielded ~80%. In eukaryotic model systems, different inherent factors of these highly complex organisms can be hypothesized to interfere with integrase functionality, leading to different overall results. However, this variability of Ints potency in eukaryotic systems can be used to design genetic circuits with distinct functionalities and modulation proprieties.

Despite none of these Ints were previously evaluated as regulators of gene expression in eukaryotic cells, two studies evaluated different Ints performing other functions in yeast (*Saccharomyces cerevisiae*) and in two mammalian cell types^{25,29}. These authors showed that 10 out of 14 integrases were active in *S. cerevisiae* through recombinase exchange reactions and that 7 out of 15 Ints promoted site-specific deletions in both lineages of mammalian cells. They also showed that seven Ints were cytotoxic in yeast and that all 15 were cytotoxic in mammalian cells to some degree, and some of the Ints were error prone in the mammalian systems. Here, the *attL*/*attR* sites were amplified and sequenced, and no mutations in these sites were found for Ints 2 and 9 in any of the models tested. Regarding the other Ints, only a few mutations were observed in some models. In addition, our cell viability assays indicated no limiting toxicities related to the integrases activity, suggesting that the systems reported herein can be used in stable expression-based experiments.

Taking into account the CaMV 35S promoter-based switch construction for protoplasts, despite a residual leak observed in the switch promoter vector negative control, Ints 4 and 5 led to a result that was compatible with the switch GFP system. Int 2, however, exhibited a much higher number of EGFP-positive cells

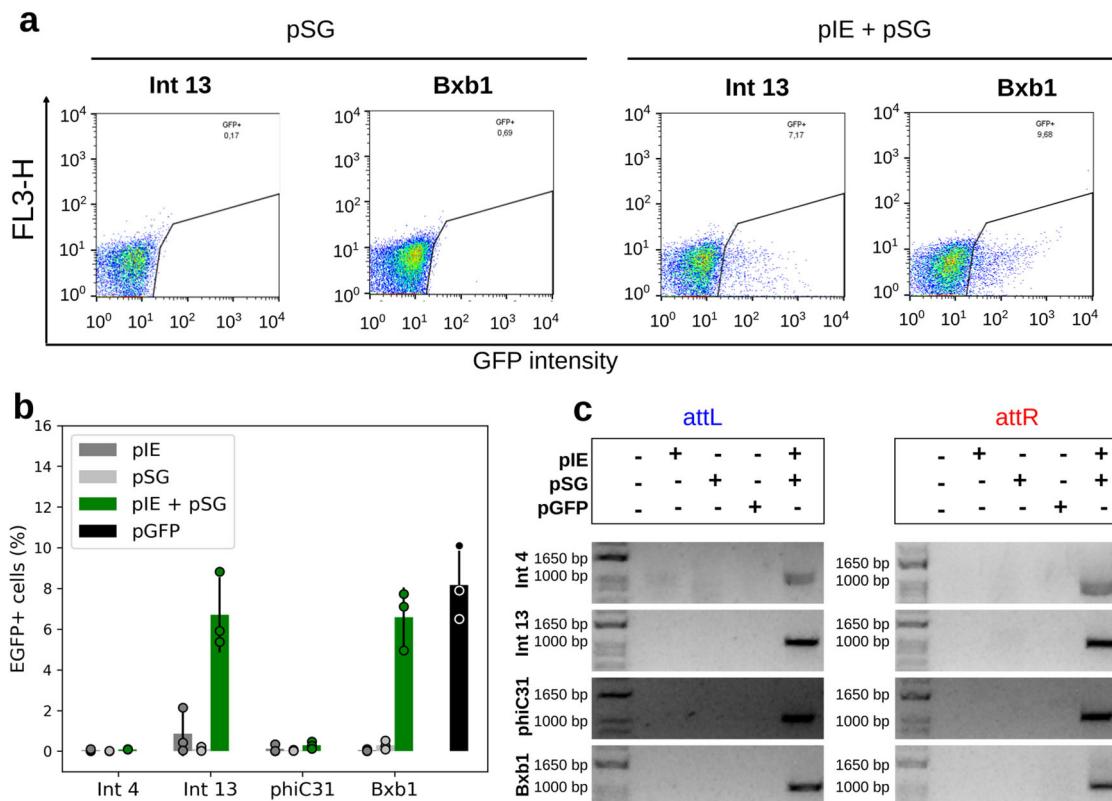


Fig. 6 Int activity in primary human T lymphocytes from PBMCs isolated from three independent donors. **a** Flow cytometry distribution of PBMCs at 48 h post electroporation for GFP switch tests with Int 13 and Bxb1, the integrases that led to the highest EGFP-expressing cell frequencies. The heat map indicates the scattering of high cell concentrations (warm colors) to low cell concentrations (cool colors). The gate encompasses the EGFP-expressing cell population. **b** Bar graph plots showing the total average percentage and standard deviation of a cell population expressing EGFP in biological repeat assays ($n = 3$) and circles corresponding to single data points from each donor material. The x axis contains the different conditions. In the PBMCs assays, Ints 2, 13, phiC31, and Bxb1 were evaluated. **c** Amplicons obtained through PCR analysis using two specific primer sets, the first set to verify *attL* formation and the second set to verify *attR* (highlighted in Fig. 1). The expected amplicon sizes in the Int test groups varied from 1021 to 1104 bp for *attL* and from 1058 to 1084 bp for *attR*. Negative control cells were electroporated with only one of the two vector sets, that is, integrase expression (pIE) or switch GFP (pSG) vectors. Positive control cells (pGFP) have an *egfp* sequence in the forward orientation under the control of the EF1 alpha promoter. All the data were representative of three donors, corresponding to biological triplicates with single measurements.

when the promoter was flipped than when the *egfp* sequence was. Moreover, this system suggested that these three Ints are orthogonal, as previously demonstrated in bacteria¹⁸.

Therefore, these proteins can be multiplexed in different combinations of genetic switches or logic gates based on Boolean algebra to build genetic circuits. Such applications would facilitate differential control of metabolic routes or synthetic transgenic systems in livestock animals or crop plants, allowing precise and efficient regulation of gene expression regarding, for example, abiotic/biotic conditions or growth timing responses to activate or deactivate resistance, defense, or nutritional improvement genes.

An important challenge to be considered in studies that aim to develop biotechnological tools with wide applicability is proof of concept. Here, in addition to the evaluation of Ints in model systems, unidirectional genetic switches were evaluated in cells of great clinical relevance, such as primary T lymphocytes, and stem cells^{39,44,45}. The therapies based on T lymphocytes and stem cells represent a field of research with a demand for transcription systems with fine-tuning capabilities⁴⁶, such as the Int-based systems described in this study. As for the model systems HEK 293T cells, bovine fibroblasts, and plant protoplasts, our results in PBMCs, NSCs, and hES cells showed that the evaluated Ints were able to recognize the *attB/attP* sites, precisely forming the *attL/attR* sites and performing the 180° rotation (flip) of the *egfp* coding sequence. Accordingly, not all the Ints tested led to high

frequencies of EGFP detection by flow cytometry and microscopy analysis, indicating differential Int activity in these cells. These results are very relevant because, first, we were able to show that these phage proteins are active in eukaryotic cells used in a variety of biologically important studies, reinforcing the robustness of the Int platform⁴⁷. Second, we can foresee the construction of genetic circuits to improve the already successful cancer immunotherapy strategies and for a wide range of potential applications of Ints in disease modeling in vitro and in therapeutic-based approaches in human stem cells. One important contribution in this context can be using Ints to refine specific temporal/space gene activation/deactivation, minimizing potential undesired side effects.

The use of Ints can also be expanded to the study of essential genes in eukaryotic organisms. Recently, Cre/Lox tyrosine integrases were utilized in a SCRaMbLE analysis to identify non-essential genes from chromosome III of *S. cerevisiae*⁴⁸. However, due to the use of only one integrase, all genes called nonessential were extracted from the genome of this organism at the same time, resulting in immediate loss of cell viability due to unknown genome redundancy⁴⁸. Our work led to a considerable increase in the number of functionally characterized Ints that were available for use in eukaryotic cells. Now, these Ints can be multiplexed to flank several endogenous genes in random combinations, making it possible to knockout separate groups of genes. Ultimately, this multiplex strategy can allow the investigation of functional gene

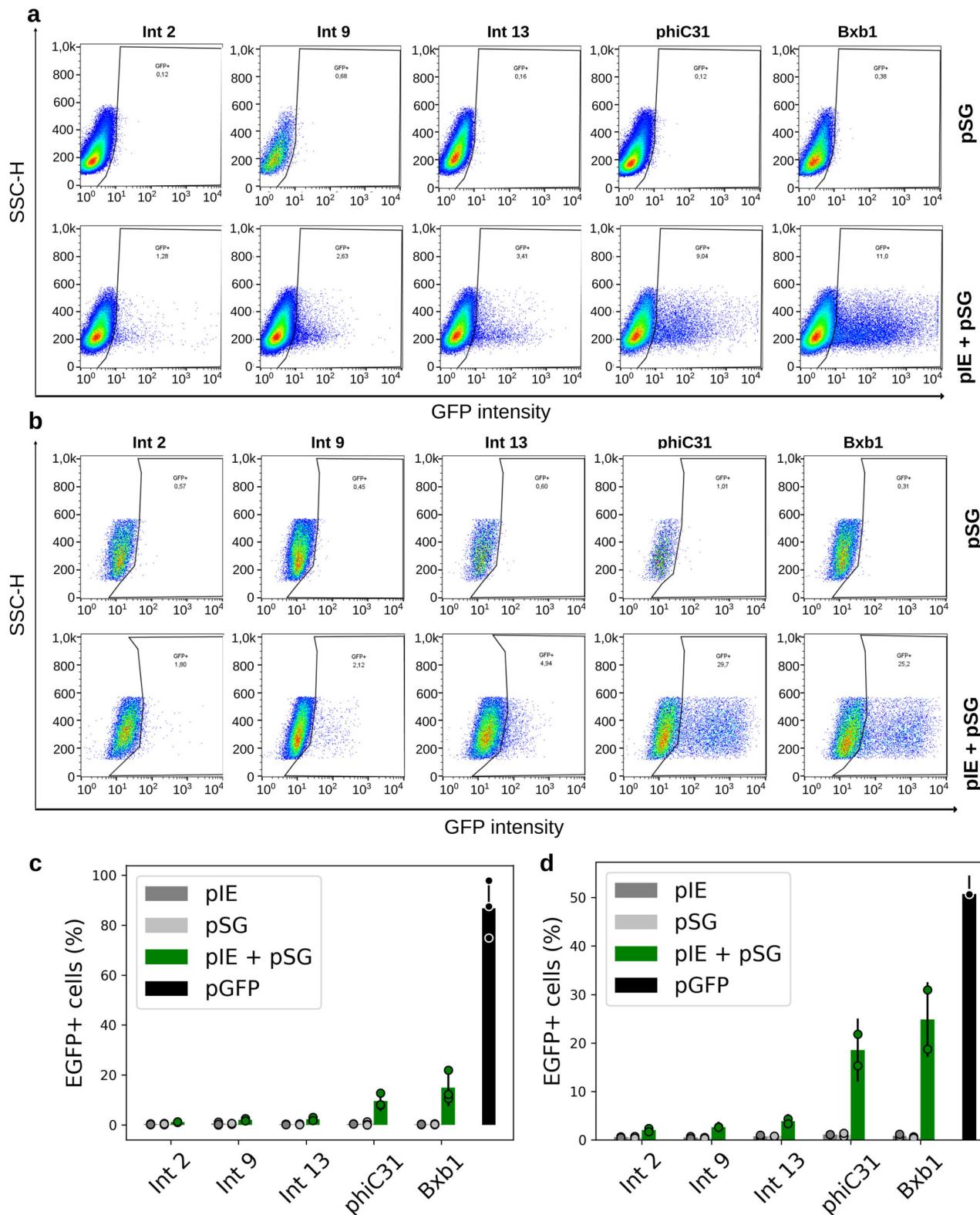


Fig. 7 Int activity in stem cells. Flow cytometry distribution of NSCs (**a**) and hES cells (**b**) at 48 h post electroporation for GFP switch tests with Ints 2, 9, 13, phiC31, and Bxb1. The heat map indicates the scattering of high cell concentrations (warm colors) to low cell concentrations (cool colors). The gate encompasses the EGFP-expressing cell population. Bar graph plots showing the total average percentage and standard deviation of NSCs (**c**) and hES cell (**d**) populations expressing EGFP in biological repeat assays ($n=3$ for NSCs; $n=2$ for hES cells) and circles showing the technical triplicate average of each assay on the y axis. The x axis contains the different conditions. In the stem cell assays, Ints 2, 9, 13, phiC31, and Bxb1 were evaluated. Negative control cells were transfected with only one of the two vector sets, that is, integrase expression (pIE) or switch GFP (pSG) vectors. Positive control cells (pGFP) have an *egfp* sequence in the forward orientation. All the data are representative of three technical and three (NSCs) or two (hES cells) biological replicates.

redundancy on the genome-scale. Furthermore, as different Ints have different degrees of functionality, these proteins can be used in studies of vital multifamily genes, allowing modulation of gene expression, or even to evaluate extrachromosomal toxic protein-coding genes. Due to their accuracy, Ints can also be used to investigate the roles of specific domains of selected genes or gene families, flanking predicted functional domains with their *attB*/*attP* sites and triggering the flipping of these sequences to the nonfunctional reverse orientation in a specific tissue or developmental stage. Finally, these proteins can also be used as DNA barcodes to identify cell lineages in developmental and evolutionary studies⁴⁹.

Importantly, in this study, six Ints were evaluated, so the number of functional Ints in eukaryotic cells can be substantially increased by taking into account the pool of more than 4000 integrases observed by Yang et al.¹⁸ in sequence databases with predicted recognition sites for 34 of these proteins.

Furthermore, studies regarding nuclear localization and protein accumulation could be performed, leading to improvement in Int functionality with the addition of NLSs or removal/modification of putative degradation signals. Last, the results presented in this work indicate that Ints 2, 4, 5, 7, 9, and 13 can further be used for a myriad of biotechnological applications.

Methods

Integrases and plasmids. The serine integrases 2, 4, 5, 7, 9, 13¹⁸, phiC31⁵⁰, and Bxb1²² were codon optimized by an online codon optimization tool (IDT software) for expression compatibility in the respective eukaryotic system: for mammalian cells, the proteins were codon optimized for expression in *Homo sapiens*, and for protoplasts, the proteins were codon optimized for expression in *A. thaliana*. Two plasmid sets were synthetically constructed (Epoch Life Science Inc.) to compose the genetic switch systems. The first plasmid set was constructed to express the Int gene (integrase expression vectors (pIE)). The second plasmid set, the switch vectors, carried the reporter *egfp* gene (switch GFP vectors (pSG)) or its promoter (switch promoter vector (pSP)) in reverse complement orientation, flanked by the *attB* site and the reverse complement sequence of the *attP* site of the Int expressed by the first plasmid. Addgene accession numbers of all vectors used in this study are described in Supplementary Table 1. Vector's part sequences are also provided in Supplementary Methods.

Mammalian system plasmids. For the integrase expression vectors, Ints 2, 4, 5, 7, 9, 13, phiC31, and Bxb1 were placed under the control of the ubiquitin promoter and β-globin poly(A) signal terminator. The coding sequences of these integrases were cloned into the pUB-GFP plasmid (Addgene, 11155), replacing the GFP-coding sequence, which resulted in a set of integrase-expressing vectors called pUB-HspINTX (X = 2, 4, 5, 7, 9, 13, phiC31, or Bxb1). For the switch GFP vectors, the *egfp* coding sequence (Addgene, 11154) was cloned in reverse complement orientation, flanked by the *attB* site and the reverse complement sequence of the *attP* site of the individual Ints 2, 4, 5, 7, 9, 13, phiC31, and Bxb1. These cassettes were cloned into the pEF-GFP plasmid (Addgene, 11154), replacing the original *egfp* coding sequence under the regulation of the EF1 alpha promoter and β-globin poly(A) signal terminator. The resulting switch GFP vectors were named pEF-GFP (rc)X (X = 2, 4, 5, 7, 9, 13, phiC31, or Bxb1) (Supplementary Fig. 1a). The GFP expression-positive control pT3-Neo-EF1α-GFP (Addgene, 69134) was used for HEK 293T and PBMCs assays; the pEF-GFP plasmid (Addgene, 11154) was used for bovine fibroblast assays; and the pT2-GFP (previously kindly provided⁵¹) was used for NSCs and hES cell assays.

Plant system plasmids. In this system, the Ints 2, 4, 5, 7, 9, 13, phiC31, and Bxb1 CDS, under the control of the actin2 gene promoter⁵² and NOS terminator (pBI426 plasmid⁵³), were cloned into the pUC57, pSB3K3 or pBluescript II SK(–) vectors by Epoch Life Science Inc. These plasmids resulted in a set of integrase expression vectors, individually called pAct-AtINTX (X = 2, 4, 5, 7, 9, 13, phiC31, or Bxb1). The *egfp*³⁶ coding sequence in a reverse complement orientation, flanked by *attB* and the reverse complement sequence of the *attP* attachment site of the Ints 2, 4, 5, 7, 9, 13, phiC31, and Bxb1, was placed under the control of the CaMV 35S promoter (iGEM registry BBa_K1547006) and the NOS terminator, constituting a set of switch GFP vectors for plants. These cassettes were inserted into the plasmids pUC18 or pBluescript SK(–) (Epoch Life Science Inc.), resulting in p35S-GFP(rc)X (X = 2, 4, 5, 7, 9, 13, phiC31, or Bxb1) (Supplementary Fig. 1b). For the plant system, a second switch vector was constructed. This vector consisted of the reverse complement of the CaMV 35S promoter flanked by the *attB* and the reverse complement sequence of the *attP* sites of the Ints 2, 4 and 5, sequentially positioned together (Fig. 5a). This promoter cassette was synthesized (Epoch Life Science Inc.)

and cloned, replacing the CaMV 35S sequence in the positive control vector pCaMV35S-GFP³⁶. The final vector was called p35S(rc)2_4_5-GFP and was used as the switch promoter vector (pSP).

HEK 293T maintenance and PBMCs isolation. The human embryonic kidney cell lineage HEK 293T (a gift from Dr. Elio Vanin of St. Jude Children's Research Hospital) was cultivated in 75-cm² tissue culture flasks with 15 ml of Dulbecco's modified Eagle's medium (DMEM; Gibco), 10% fetal bovine serum (FBS; HyClone), and penicillin-streptomycin (10 U/ml; Gibco). The cells were detached and seeded every 2–3 days at below 80% confluence. For PBMCs, white blood cells from healthy blood donors were collected using a leukocyte reduction filter (RS; Haemonetics) and washed with phosphate-buffered saline (PBS). To isolate PBMCs, a density gradient centrifugation was performed using Ficoll-Hypaque®-1077 (GE HealthCare) (deceleration off; centrifugation for 20 min at 800 × g) followed by three washes with PBS. The use of PBMCs from healthy donors was approved by an Institutional Review Board (the Brazilian National Cancer Institute (INCA) Ethics Committee), and donors signed review board-approved informed consent forms.

Human stem cell culture. Neural stem cells (NSCs) differentiated from induced pluripotent stem cells (iPSCs)⁵⁴ and pluripotent human embryonic stem cells (hES, BR-1 cell line)⁴¹ were used. All human stem cell experiments were approved by the ethics committee of Copo D'Or Hospital (CAAE number 60944916.5.0000.5249, approval number 1.791.182). The cells were cultured/maintained in neural advanced DMEM/F12 and neurobasal medium (50% v/v) plus neural induction supplement (NIS) medium (all from Thermo Fisher Scientific), called NEM (neural expansion medium), over Geltrex (Thermo Fisher Scientific) at 37 °C in 5% CO₂ as previously described^{54,55}. hES cells were cultured in mTeSR medium (Stem Cell Technologies) over Geltrex.

HEK 293T transfection. A total of 4 × 10⁶ HEK 293T cells were plated in 75-cm² flasks with 15 ml of DMEM (Gibco), 10% FBS (HyClone), and penicillin-streptomycin (10 U/ml; Gibco). After 24 h, the medium was removed, and 10 ml of fresh DMEM/FBS was added. In the next step, 5 µg of each integrase expression vector pUB-IntX (X = 2, 4, 5, 7, 9, 13, phiC31, or Bxb1) and each switch GFP vector pEF-GFP(rc)X carrying the respective integrase site were added to 500 µl of 2X CaCl₂ at 250 nM. Next, 500 µl of HBS (pH 7.1) was slowly added while the solution was vortexed at 10,000 rpm. Bubbles were produced in the solution with a Pasteur pipette and mixed. The solution was incubated for 10 min at room temperature and then dripped with the Pasteur pipette throughout the flask. The medium was changed after 16 h. The results were analyzed 48 h after the transient transfection.

PBMC, NSC, and hES electroporation. Cells (1 × 10⁷ PBMCs, 1 × 10⁶ NSCs or hES cells) were transferred to a sterile 0.2-cm cuvette (Mirus Biotech®) and electroporated as previously described⁵¹. Briefly, PBMCs and hES cells were resuspended in 100 µl of 1SM buffer, and NSCs were resuspended in 100 µl of 1S buffer. PBMCs were electroporated with 5 µg of each integrase expression (pIE) and switch GFP (pSG) plasmids using the U-14 program of the Lonza® Nucleofector® II electroporation system. NSCs and hES cells were electroporated with 12 µg of pIE plasmids and 8 µg of pSG plasmids using A-33 and A-23, respectively, from the Lonza® Nucleofector® II electroporation system. The mock control was electroporated with 100 µl of 1SM (PBMCs and hES cells) or 1S (NSCs) buffer without plasmid (used to set flow cytometry gates). After transfection, PBMCs were gently resuspended in 1 ml of prewarmed RPMI medium supplemented with 2 mM L-Glu and 20% fetal calf serum (FCS; Gibco). NEM was used for NSCs, and mTeSR was used for hES cells. All cells were cultured for 48 h after electroporation for transient transfection, and then, the analyses were performed.

Bovine fibroblast isolation. Fibroblasts were isolated according to the protocol described by Freshney⁵⁶ with some modifications. The cells were removed from 14-month-old Nelore (*Bos indicus*) bull oxtail by biopsies and washed three times in 0.05% trypsin (Gibco). The cells were then transferred to 25-cm² cell culture flasks and incubated in DMEM (Gibco) supplemented with 10% FBS (Gibco) and penicillin-streptomycin at 37 °C in a 5% CO₂ atmosphere. After three passages, or when the fibroblast cultures showed homogeneity, the cells were ready for transfection. Cell cultures with 60–70% confluence were picked. The use of the bovine cells was approved by the Ethics Committee on the Use of Animals (CEUA) of Embrapa Genetic Resources and Biotechnology in March 2013 under the reference number 001/2013.

Bovine fibroblast cell transfection. After growth, the cells were enzymatically dissociated with a trypsin solution (0.5% trypsin, 0.2% EDTA), and after 10 min, the reaction was inactivated using DMEM (Gibco). The cells were counted in a Neubauer chamber, transferred (10⁵ cells/well) to 24-well culture dishes and grown for 24 h, when they reached >70% confluence. Primary bovine fibroblasts were cotransfected with 350 ng of each of the two plasmid sets utilizing Lipofectamine LTX & Plus Reagent (Invitrogen) and cultured in Opti-MEM (Invitrogen)

according to the manufacturer's instructions. The results were analyzed 48 h after transient transfection.

Protoplast isolation. The protoplasts were obtained following the protocol described by Yoo et al.³⁷ with some modifications. *A. thaliana* ecotype Columbia was grown under a 12-h light/12-h dark cycle at 22 °C. Four to five weeks after seeding, ~20 young leaves were collected, transferred to a plate with W5 solution (154 mM NaCl, 125 mM CaCl₂, 5 mM KCl, 2 mM MES, pH 5.7), and scalped by using a blade. Then, the leaves were placed on a digestion plate containing 5 ml of enzyme solution [0.5 M mannitol, 20 mM KCl, 20 mM MES (pH 5.7), 0.2% pectolyase (Sigma-Aldrich), 0.5% driselase (Kyowa Hakko Bio Co., Ltd.), 1.5% cellulase (Sigma-Aldrich), 10 mM CaCl₂, 1 mg/ml BSA]. A 15–20 pol Hg vacuum was applied three times for ~5 s, and the plate was incubated at room temperature in a platform shaker at 40 rpm for 3 h. The digested sample was filtered through a 44-μm mesh, and the W5 solution was added to wash the obtained protoplasts, followed by centrifugation at 100 × g for 2 min. After two additional washing steps followed by centrifugation with 10 ml of W5, the protoplasts were resuspended in 1 ml of MMg solution (0.4 M mannitol, 15 mM MgCl₂, 4 mM MES, pH 5.7), and the concentration was adjusted to 4–5 × 10⁵ protoplasts/ml.

Protoplast transformation. Cotransformation was performed in a 15-ml Corex tube using 100 μl of 4–5 × 10⁵ protoplasts/ml, 10 μg of each desired plasmid DNA of the two sets of vectors and 110 μl of 40% PEG solution [PEG 4000 (Sigma), 0.2 M mannitol, 100 mM CaCl₂] for each reaction (scaling up to 6 reactions per tube). After 15 min, the reaction was stopped with two volumes of W5 solution, centrifuged at 100 × g for 2 min, resuspended in 500 μl of W1 solution (0.5 M mannitol, 20 mM KCl, 4 mM MES, pH 5.7) (each reaction) and plated on a 12-well plate. The plates were incubated at room temperature in a platform shaker at 40 rpm, and the results were analyzed after 24 h transient transformation.

Flow cytometry. HEK 293T cells, NSCs, hES cells, bovine fibroblasts, and *A. thaliana* protoplasts were analyzed by flow cytometry in technical triplicates. PBMCs were analyzed in one sample for each of the three donors. HEK 293T cells, PBMCs, NSCs, and hES cells were detached/resuspended from the culture flasks with cold PBS and washed again with PBS. Signal acquisition was performed with a FACS Calibur (BD Biosciences), and the cells with FSC/SSC patterns compatible with viable cells were gated. Then, a second selection was done for 7AAD negative population. Analyses were performed using FlowJo software, version 10 (Tree Star). Bovine fibroblasts were trypsinized from 24-well plates, where 200 μl of DMEM was added per well. The contents were transferred to a microtube and centrifuged at 687 × g for 5 min, and the supernatant was removed, leaving ~50 μl for analysis. Protoplasts from each well of 12-well plates were transferred to a microtube and centrifuged at 100 × g for 1 min, and the supernatant was removed, leaving a volume of ~50 μl. The entire contents of each tube were analyzed. Bovine and protoplast cells were analyzed on an imaging flow cytometer (Amnis FlowSight) under 488-nm laser excitation and a power of 60 and 10 mW, respectively. Signal acquisition in bovine cells was determined by channel 3 intensity (filters at 566–635) versus channel 2 intensity (green EGFP reporter emission, 532–555 nm). Signal acquisition in the protoplast population was determined by channel 4 intensity (chloroplast autofluorescence red emission, 610/30 nm) versus channel 2 intensity. Before the first reading, a no-gating population was acquired under intensity channels and the imaging system from the Amnis FlowSight cytometer allowed visually gating the viable cells population. Once the gate was set up, the samples were acquired using this gate. The results from bovine and protoplast cells were processed and visualized using IDEAS software. For mammalian systems, at least 10,000 single cells were analyzed, and for protoplasts, at least 1000 single cells were analyzed.

DNA extraction, amplification, and sequencing. For HEK 293T cells, PBMCs, and bovine fibroblasts, total DNA was extracted from the pool of technical triplicates using DNeasy Blood & Tissue Kits (Qiagen) after a 48-h assay. Plant protoplasts were also pooled after a 24-h assay, and the DNA was extracted using the DNeasy Plant Mini Kit (Qiagen). For all samples, the PCRs were carried out using appropriate primer pairs (Supplementary Table 7) to amplify the *attL* and *attR* site-containing regions from the reporter switch vectors (pSG and pSP) using Platinum Taq DNA polymerase (Invitrogen). The Wizard® SV Gel and PCR Clean-Up System (Promega) was used to clean the expected amplicon from the agarose gel. The PCR products were cloned into the pGEM-T Easy vector (Promega) according to the manufacturer's protocol and transformed into *E. coli* DH10B or XL1-blue chemically competent cells by heat shock. The plasmid DNA was extracted by the Wizard® Plus SV Miniprep DNA Purification System (Promega) and sequenced by Macrogen, always using the M13F and M13R or SP6 and T7 primer pair for coverage sequencing. The obtained sequences were aligned with the expected activated plasmid sequences and trimmed using Geneious software (version 7.0.6).

Cell viability assays. HEK 293T cells and bovine fibroblasts were plated in 96-well plates at a density of 1 × 10⁵ cells/well in triplicates and grown for 24 h at 37 °C in a 5% CO₂ atmosphere. The cells were cotransfected with pIE and pSG vectors or only

with one of these vectors plus a mock plasmid as previously described proportionally to a 96-well plate assay. For the impairment negative control were used 20 μL dimethyl sulfoxide (DMSO; Sigma-Aldrich) for a final volume of 200 μL per well. After 48 h transient cotransfection, the cells were incubated with 15 μl of 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT; Thermo Fisher Scientific) (5 mg/mL) for 4 h at 37 °C. Subsequently, the medium was removed and 150 μl DMSO (Sigma-Aldrich) was added to each well to dissolve formazan crystals. A microplate reader (Sunrise reader, Tecan) calibrated to read absorbance at 595 nm was used to quantify the formazan product. Protoplasts were transformed with pIE vector of each integrase or with a mock plasmid in triplicates as previously described. The impairment negative control was obtained with 50 μL DMSO (Sigma-Aldrich) added per well to complete the final incubation volume of 500 μL. The fluorescein diacetate (FDA; Sigma-Aldrich) assay was performed following Lin et al.³⁸ protocol with some modifications. After 24 h transient transformation, each sample was incubated with 3 μL FDA work solution [10 μL stock solution (5 mg/mL) in 2.5 mL W1 solution] by 8 min and analyzed by flow cytometry as previously described. The work solution was remade every 2 h. Channel 2 intensity was used to acquire FDA positive cells population in an adequate gate corresponding to viable cells.

Statistics and reproducibility. The dataset for the genetically activated cell proportion (EGFP+) obtained by flow cytometry was analyzed using R software (version 3.6.0). The nonparametric Kruskal-Wallis test was used to determine significant differences between controls and test conditions of each Int group at the 5% statistical probability level.

Data availability

All plasmids constructed for this study are available in Addgene repository and the accession numbers are listed in Supplementary Table 1. Also, the genetic sequence parts used are listed in Supplementary Information. The complete sequence alignment dataset is available in Supplementary Data. The full dataset of positive GFP cell proportion obtained by flow cytometry analysis in all cell experiments and the OD acquisitions of MTT assays were deposited in Dryad Digital Repository (<https://doi.org/10.5061/dryad.dr7sqv9tv>)⁵⁹. Any other data are available from the corresponding authors upon request.

Received: 7 October 2019; Accepted: 28 April 2020;

Published online: 22 May 2020

References

1. Wirth, T., Parker, N. & Ylä-herttula, S. History of gene therapy. *Gene* **525**, 162–169 (2013).
2. Ueki, H., Watanabe, M., Kaku, H., Huang, P. & Li, S. A novel gene expression system for detecting viable bladder cancer cells. *International J. Oncol.* **41**, 135–140 (2012).
3. Meer, J. R. VanDer & Belkin, S. Where microbiology meets microengineering: design and applications of reporter bacteria. *Nat. Rev. Microbiol.* **8**, 511–522 (2010).
4. Fasani, E., Manara, A., Martini, F., Furini, A. & Dalcorso, G. The potential of genetic engineering of plants for the remediation of soils contaminated with heavy metals. *Plant. Cell Environ.* **41**, 1201–1232 (2018).
5. Christou, P. Plant genetic engineering and agricultural biotechnology 1983–2013. *Trends Biotechnol.* **31**, 125–127 (2013).
6. Ahmad, N. & Mukhtar, Z. Genetic manipulations in crops: challenges and opportunities. *Genomics* **109**, 494–505 (2017).
7. Eenennaam, A. L. Van Genetic modification of food animals. *Curr. Opin. Biotechnol.* **44**, 27–34 (2017).
8. Lewandoski, M. Conditional control of gene expression in the mouse. *Nat. Rev. Genet.* **2**, 743–755 (2001).
9. Hernandez-garcia, C. M. & Finer, J. J. Identification and validation of promoters and cis -acting regulatory elements. *Plant Sci.* **217–218**, 109–119 (2014).
10. Purnick, P. E. M. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* **10**, 410–422 (2009).
11. Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nat. Publ. Gr.* **11**, 367–379 (2010).
12. Arber, W. Host-controlled modification of bacteriophage. *Annu. Rev. Microbiol.* **19**, 365–378 (1965).
13. Merrick, C. A., Zhao, J. & Rosser, S. J. Serine integrases: advancing synthetic biology. *ACS Synth. Biol.* **7**, 299–310 (2018).
14. Stark, W. M. Making serine integrases work for us. *Curr. Opin. Microbiol.* **38**, 130–136 (2017).
15. Smith, M. C. M. in *Mobile DNA III* 253–272 (American Society of Microbiology, 2014).
16. Rutherford, K. & Van Duyne, G. D. The ins and outs of serine integrase site-specific recombination. *Curr. Opin. Struct. Biol.* **24**, 125–131 (2014).

17. Brown, W. R. A., Lee, N. C. O., Xu, Z. & Smith, M. C. M. Serine recombinases as tools for genome engineering. *Methods* **53**, 372–379 (2011).
18. Yang, L. et al. Permanent genetic memory with >1-byte capacity. *Nat. Methods* **11**, 1261–1266 (2014).
19. Olorunniyi, F. J., Rosser, S. J. & Stark, W. M. Site-specific recombinases: molecular machines for the Genetic Revolution. *Biochem. J.* **473**, 673–684 (2016).
20. Nash, H. A. Integration and excision of bacteriophage λ : the mechanism of conservative site specific recombination. *Annu. Rev. Genet.* **15**, 143–167 (1981).
21. Rutherford, K., Yuan, P., Perry, K., Sharp, R. & Van Duyne, G. D. Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res.* **41**, 8341–8356 (2013).
22. Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P. & Endy, D. Amplifying genetic logic gates. *Science* **340**, 599–603 (2013).
23. Siuti, P., Yazbek, J. & Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* **31**, 448–452 (2013).
24. Roquet, N., Soleimany, A. P., Ferris, A. C., Aaronson, S. & Lu, T. K. Synthetic recombinase-based state machines in living cells. *Science* **353**, aad8559 (2016).
25. Xu, Z. & Brown, W. R. A. Comparison and optimization of ten phage encoded serine integrases for genome engineering in *Saccharomyces cerevisiae*. *BMC Biotechnol.* **16**, 13 (2016).
26. Weinberg, B. H. et al. Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat. Biotechnol.* **35**, 453–462 (2017).
27. Wang, X. et al. Bxb1 integrase serves as a highly efficient DNA recombinase in rapid metabolite pathway assembly. *Acta Biochim. Biophys. Sin.* **49**, 44–50 (2017).
28. Kapusi, E., Kempe, K., Rubtsova, M., Kumlehn, J. & Gils, M. phiC31 integrase-mediated site-specific recombination in barley. *PLoS ONE* **7**, e45353 (2012).
29. Xu, Z. et al. Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnol.* **13**, 87 (2013).
30. Thomson, J. G. et al. The Bxb1 recombination system demonstrates heritable transmission of site-specific excision in *Arabidopsis*. *BMC Biotechnol.* **12**, 9 (2012).
31. Fogg, P. C. M., Haley, J. A., Stark, W. M. & Smith, M. C. M. Genome integration and excision by a new *Streptomyces* bacteriophage, ϕ Joe. *Appl. Environ. Microbiol.* **83**, 1–16 (2017).
32. Rubtsova, M. et al. Expression of active *Streptomyces* phage phiC31 integrase in transgenic wheat plants. *Plant Cell Rep.* **27**, 1821–1831 (2008).
33. Kempe, K. et al. Transgene excision from wheat chromosomes by phage phiC31 integrase. *Plant Mol. Biol.* **72**, 673–687 (2010).
34. Yu, Y. et al. Improved site-specific recombinase-based method to produce selectable marker- and vector-backbone-free transgenic cells. *Sci. Rep.* **4**, 4240 (2014).
35. Thomson, J. G., Chan, R., Thilmony, R., Yau, Y.-Y. & Ow, D. W. PhiC31 recombination system demonstrates heritable germinal transmission of site-specific excision from the *Arabidopsis* genome. *BMC Biotechnol.* **10**, 17 (2010).
36. Chiu, W. et al. Engineered GFP as a vital reporter in plants. *Curr. Biol.* **6**, 325–330 (1996).
37. Van Duyne, G. D. & Rutherford, K. Large serine recombinase domain structure and attachment site binding. *Crit. Rev. Biochem. Mol. Biol.* **48**, 476–491 (2013).
38. Barros, L. et al. Immunological-based approaches for cancer therapy. *Clinics* **73**, 1–11 (2018).
39. Chicaybam, L., Laino Sodré, A. & Bonamino, M. Chimeric antigen receptors in cancer immuno-gene therapy: current status and future directions. *Int. Rev. Immunol.* **30**, 294–311 (2011).
40. Tang, Y., Yu, P. & Cheng, L. Current progress in the derivation and therapeutic application of neural stem cells. *Nat. Publ. Gr.* **8**, e3108 (2017).
41. Fraga, A. M. et al. Establishment of a Brazilian line of human embryonic stem cells in defined medium: implications for cell therapy in an ethnically diverse population. *Cell Transplant.* **20**, 431–440 (2011).
42. Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. *Nat. Publ. Gr.* **17**, 5–15 (2015).
43. Guinebretiere, M.-H. et al. *Bacillus cytotoxicus* sp. nov. is a novel thermotolerant species of the *Bacillus cereus* Group occasionally associated with food poisoning. *Int. J. Syst. Evol. Microbiol.* **63**, 31–40 (2013).
44. Vargas, J. E. et al. Retroviral vectors and transposons for stable gene therapy: advances, current challenges and perspectives. *J. Transl. Med.* **14**, 1–15 (2016).
45. Dakic, V. et al. Harmine stimulates proliferation of human neural progenitors. *PeerJ* **4**, e2727 (2016).
46. Chicaybam, L. & Bonamino, M. H. Moving receptor redirected adoptive cell therapy toward fine tuning of antitumor responses. *Int. Rev. Immunol.* **33**, 402–416 (2014).
47. Zhu, F. et al. DICE, an efficient system for iterative genomic editing in human pluripotent stem cells. *Nucleic Acids Res.* **42**, e34 (2014).
48. Annaluru, N. et al. Total synthesis of a functional designer eukaryotic chromosome. *Science* **344**, 55–58 (2014).
49. Kalhor, R. et al. Developmental barcoding of whole mouse via homing CRISPR. *Science* **361**, eaat9804 (2018).
50. Grandchamp, N. et al. Hybrid lentivirus-phiC31-int-NLS vector allows site-specific recombination in murine and human cells but induces DNA damage. *PLoS ONE* **9**, e99649 (2014).
51. Chicaybam, L. et al. An efficient electroporation protocol for the genetic modification of mammalian cells. *Front. Bioeng. Biotechnol.* **4**, 1–13 (2017).
52. Aragão, F. J. L., Vianna, G. R., Carvalheira, S. B. R. C. & Rech, E. L. Germ line genetic transformation in cotton (*Gossypium hirsutum* L.) by selection of transgenic meristematic cells with a herbicide molecule. *Plant Sci.* **168**, 1227–1233 (2005).
53. Datla, R. S. S., Hammerlindl, J. K., Pelcher, L. E., Crosby, W. L. & Selvaraj, G. A bifunctional fusion between β -glucuronidase and neomycin phosphotransferase: a broad-spectrum marker enzyme for plants. *Gene* **101**, 239–246 (1991).
54. Casas, B. S. et al. hiPSC-derived neural stem cells from patients with schizophrenia induce an impaired angiogenesis. *Transl. Psychiatry* **8**, 48 (2018).
55. Garcez, P. P. et al. Zika virus impairs growth in human neurospheres and brain organoids. *Science* **352**, 816–818 (2016).
56. Freshney, R. I. *Culture of Animal Cells: a Manual of Basic Technique*. (Wiley-Liss, Inc, New York, 1994).
57. Yoo, S. D., Cho, Y. H. & Sheen, J. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* **2**, 1565–1572 (2007).
58. Lin, H.-Y., Chen, J.-C. & Fang, S.-C. A protoplast transient expression system to enable molecular, cellular, and functional studies in *Phalaenopsis* orchids. *Front. Plant Sci.* **9**, 1–13 (2018).
59. Gomide, M. et al. Genetic switches designed for eukaryotic cells and controlled by serine integrases. Dryad data sets. <https://doi.org/10.5061/dryad.dr7sq9tv> (2020).

Acknowledgements

We thank Joseane Padilha da Silva, MSc, from Embrapa Genetic Resources & Biotechnology for providing assistance with the statistical analyses; Dr Miguel Andrade, for figures graphic design; Adriano Caiado A. F. de Almeida, MSc, for contributions to the p35S(rc)2_4_5-GFP (pSP) cloning process; Lara Rodrigues Nesrala, MSc, who took part in the adaptation of the protocols for protoplast extraction and transformation; Dr Andrielle Thainar Mendes Cunha, for assistance with the flow cytometry methodology; and INCA's blood bank and hemotherapy service for the preparation of donor samples. This research was funded by INCT BioSyn (National Institute of Science and Technology in Synthetic Biology), CNPq (National Council for Scientific and Technological Development), CAPES (Coordination for the Improvement of Higher Education Personnel), Brazilian Ministry of Health, and FAPDF (Research Support Foundation of the Federal District), Brazil.

Author contributions

Conceived and designed the experiments: E.R., C.M.C., M.H.B., and A.M.M. Performed vector design and cloning: C.M.C., M.S.G., G.P.C.J., M.S.M.A., L.M.G.B., and R.N.L. Performed the *Arabidopsis thaliana* protoplast experiments: M.S.G., L.M.G.B., L.H.F., and C.L. Performed the bovine fibroblast experiments: T.T.S., M.A.O., E.O.M., G.P.C.J., and R.N.L. Performed the HEK 293T cell, PBMCs, NSCs, and hES cell experiments: L.R. C.B., C.G.L., M.L.R., L.H.F., M.A.O., and S.K.R. Wrote the paper: E.R., C.M.C., M.S.G., T. T.S., L.M.G.B., R.N.L., and M.H.B. All the authors analyzed, interpreted, and discussed the results and contributed to the final paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s42003-020-0971-8>.

Correspondence and requests for materials should be addressed to M.H.B., C.M.C. or E.R.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020

Anexo II

Publicação de artigo - coautoria

Protocol for the establishment of a serine integrase-based platform for functional validation of genetic switch controllers in eukaryotic cells

Marco A de Oliveira^{1 2}, Lilian H Florentino^{1 2 3}, Thais T Sales^{1 2 3}, Rayane N Lima^{2 3}, Luciana R C Barros⁴, Cintia G Limia⁵, Mariana S M Almeida^{2 3}, Maria L Robledo⁵, Leila M G Barros^{2 3}, Eduardo O Melo^{2 3}, Daniela M Bittencourt^{2 3}, Stevens K Rehen⁶⁷, Martín H Bonamino^{8 9}, Elibio Rech^{2 3}

PLoS One (2024)

DOI: 10.1371/journal.pone.0303999

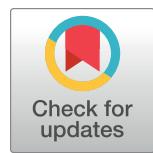
LAB PROTOCOL

Protocol for the establishment of a serine integrase-based platform for functional validation of genetic switch controllers in eukaryotic cells

Marco A. de Oliveira^{1,2}, Lilian H. Florentino^{1,2,3}, Thais T. Sales^{1,2,3}, Rayane N. Lima^{2,3}, Luciana R. C. Barros⁴, Cintia G. Limia⁵, Mariana S. M. Almeida^{2,3}, Maria L. Robledo⁵, Leila M. G. Barros^{2,3}, Eduardo O. Melo^{2,3}, Daniela M. Bittencourt^{2,3}, Stevens K. Rehen^{6,7}, Martín H. Bonamino^{3,9}, Elíbio Rech^{1,2,3*}

1 Department of Cell Biology, Institute of Biological Science, University of Brasília, Brasília, Distrito Federal, Brazil, **2** National Institute of Science and Technology in Synthetic Biology (INCT BioSyn), Brasília, Distrito Federal, Brazil, **3** Embrapa Genetic Resources and Biotechnology, Brasília, Distrito Federal, Brazil, **4** Center for Translational Research in Oncology, Instituto do Câncer do Estado de São Paulo, Hospital das Clínicas da Faculdade de Medicina de Universidade de São Paulo, São Paulo, Brazil, **5** Molecular Carcinogenesis Program, Research Coordination, National Cancer Institute (INCA), Rio de Janeiro, Brazil, **6** D'Or Institute for Research and Education (IDOR), Rio de Janeiro, Brazil, **7** Institute of Biomedical Sciences, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, **8** Cell and Gene Therapy Program, Research Coordination, National Cancer Institute (INCA), Rio de Janeiro, Brazil, **9** Vice-Presidency of Research and Biological Collections (VPPCB), FIOCRUZ – Oswaldo Cruz Foundation Institute, Rio de Janeiro, Brazil

* elibio.rech@embrapa.br



OPEN ACCESS

Citation: de Oliveira MA, Florentino LH, Sales TT, Lima RN, Barros LRC, Limia CG, et al. (2024) Protocol for the establishment of a serine integrase-based platform for functional validation of genetic switch controllers in eukaryotic cells. PLoS ONE 19(5): e0303999. <https://doi.org/10.1371/journal.pone.0303999>

Editor: Chen Ling, Fudan University, CHINA

Received: September 6, 2023

Accepted: May 4, 2024

Published: May 23, 2024

Copyright: © 2024 de Oliveira et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All results and dataset generated in the development and use of the present protocol are available in previous publication and supplementary information (DOI <https://doi.org/10.13039/5018693/28203482>). Plasmid maps and sequences are available from Addgene database (<https://www.addgene.org/browse/article/28203482/>).

Funding: This research was funded by Embrapa Genetic Resources and Biotechnology/National Institute of Science and Technology in Synthetic

Abstract

Serine integrases (Ints) are a family of site-specific recombinases (SSRs) encoded by some bacteriophages to integrate their genetic material into the genome of a host. Their ability to rearrange DNA sequences in different ways including inversion, excision, or insertion with no help from endogenous molecular machinery, confers important biotechnological value as genetic editing tools with high host plasticity. Despite advances in their use in prokaryotic cells, only a few Ints are currently used as gene editors in eukaryotes, partly due to the functional loss and cytotoxicity presented by some candidates in more complex organisms. To help expand the number of Ints available for the assembly of more complex multifunctional circuits in eukaryotic cells, this protocol describes a platform for the assembly and functional screening of serine-integrase-based genetic switches designed to control gene expression by directional inversions of DNA sequence orientation. The system consists of two sets of plasmids, an effector module and a reporter module, both sets assembled with regulatory components (as promoter and terminator regions) appropriate for expression in mammals, including humans, and plants. The complete method involves plasmid design, DNA delivery, testing and both molecular and phenotypical assessment of results. This platform presents a suitable workflow for the identification and functional validation of new tools for the genetic regulation and programming of organisms with importance in different fields, from medical applications to crop enhancement, as shown by the initial results obtained. This protocol can be

Biology, National Council for Scientific and Technological Development (465603/2014-9) and Research Support Foundation of the Federal District (0193.001.262/2017).

Competing interests: The authors have declared that no competing interests exist.

completed in 4 weeks for mammalian cells or up to 8 weeks for plant cells, considering cell culture or plant growth time.

Introduction

The ability to regulate gene expression in response to external cues is one of the central mechanisms of the differentiation and maintenance of life in nature, as well as one of the main goals of scientists in efforts to control and reprogram organisms. Therefore, the availability of molecular tools that allow genetic manipulation is crucial for advances in synthetic biology, especially in creating intricate genetic circuits and activation cascades to work as synthetic regulatory networks. A prominent group of effectors used to that end are integrases, a superfamily of site-specific recombinases (SSRs) capable of directed and controlled rearrangement of DNA sequences [1–4]. Although tyrosine integrases such as Cre/LoxP [5–8] and λ Int [5] systems have historically been predominantly used, recently, another family of SSRs known as serine integrases has received attention, in great part because of the advantages they present over their counterparts. The main advantages for application in synthetic biology include the shorter length of their attachment sites, unidirectional recombination and nondependence on auxiliary effectors to work [9–14]. Despite the advantages, only a very limited number of functional serine integrases have been available for use in eukaryotic organisms. We have recently used 6 out of the 13 newly described functional serine integrases previously identified and characterized in *Escherichia coli* by Yang and collaborators [15] to assemble and test genetic switches capable of modulating gene expression upon Int activation in eukaryotic systems, including plant-, bovine- and human-derived cells [16]. While in the original work in *E. coli* all integrases showed similar efficiency, their efficiency in plant and mammal cells varied considerably, yet successful recombination was detected for all integrases used.

Originally present in nature as a mechanism used by some bacteriophages to integrate their DNA into the genome of a prokaryotic host, the serine-integrase mechanism of action involves recognition and physical interaction with a specific pair of attachment sites present in both phage and bacterial host DNA known as *attP* and *attB* sites, respectively [17, 18]. Upon binding to DNA, conformational alterations of the complex help to expose the central part of the *att* sites, known as the core sequence, to the catalytic site responsible for the double break in the DNA strands. Subsequent conformational changes lead to the recombination of cut ends from complementary *att* sites and final ligation to form the newly recombined *attL* and *attR* sites in a permanent, unidirectional way [19–21]. Since this whole process occurs without the need for any additional endogenous effector, serine integrases exhibit valuable host plasticity when considered as a gene editing tool. However, it is the rational manipulation of *att* site combination and orientation that truly broadens the use possibilities of these recombinases. When cognate *att* sites are present in different DNA molecules, the natural integration process occurs. Notwithstanding, if *att* sites are synthetically designed to be present at the same molecule, the recombination of these sites can have varying outcomes according to their orientation, namely, excision of a portion of the DNA molecule, a 180° inversion of a DNA sequence flanked by the *att* sites [15] and recombinase-mediated cassette exchange (RMCE) [22, 23]. Diagrams to better illustrate these possible outcomes are presented in Fig 1.

Genetic switches designed to invert a given sequence are especially interesting for expression control. In this type of system, a gene of interest or the promoter regulating its expression is assembled in an opposing orientation relative to the other and flanked by a pair of *att* sites. Upon integrase activation, *att* sites are recombined with consequent inversion of the DNA

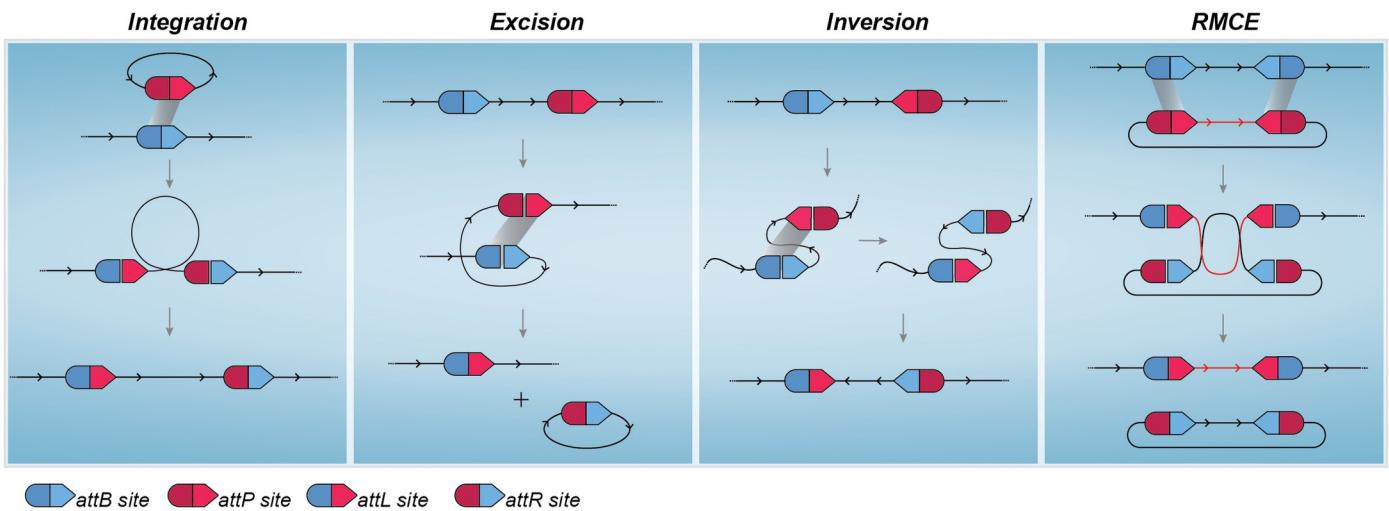


Fig 1. Serine integrase mediated DNA rearrangement. (A) Although the original function of this class of recombinases in nature is the integration of bacteriophage DNA into a host genome upon recombination of *attB* and *attP* sites, synthetic manipulation of *att* sites can lead to different outcomes. Recombination of *att* sites located in the same molecule will result in (B) Excision of flanked DNA region if both sites have the same orientation or (C) Inversion when sites are present in opposite orientations. (D) Recombinase Mediated Cassette Exchange (RMCE) allows the exchange of a large sequence of DNA when a molecule harbors two copies of an *attP* site and a second molecule harbors two copies of the cognate *attB* site.

<https://doi.org/10.1371/journal.pone.0303999.g001>

sequence in between, therefore allowing proper gene transcription [15, 16]. Recombination by serine-integrase is unidirectional and permanent unless a cognate recombination directionality factor (RDF; an accessory protein encoded in the same phage genome that contains the one integrase with which it can interact) [24] or bidirectional activating mutations [25] are present.

Although the most prominent applications of recombinases to date involve the use of tyrosine-integrases, serine-integrases were first identified and used to deliver DNA in *Streptomyces spp.* over 30 years ago [26]. Since then, many works have taken advantage of their recombination capabilities in prokaryotic models [15], unicellular eukaryotic organisms [23], and in a few cases, complex multicellular organisms, including mouse and human cells [16, 27], and they have even been used as molecular tools for in vitro DNA plasmid assembly [14, 28]. More recently, the serine-integrase ability to excise or flip terminators and genes has been more widely applied in the assembly of genetic circuits designed to either modulate gene expression or compute events by permanently rearranging DNA parts based on binary logic [15, 29–32]. The higher intricacy of activation cascades and genetic circuits with multiple integrases as effectors intensifies the demand for new functional integrases and corresponding *att* sites, encouraging efforts to describe new members and to develop systematic models for the identification and testing of candidates, such as bioinformatic tools to identify potential serine-integrases and *att* sites in genome databases [15, 27] and *in silico* platforms to help with the design and assembly of complex genetic circuits [33–35]. That need is even more accentuated when considering the use of genetic circuits based on integrases in eukaryotic organisms, since the higher complexity and compartmentalization create a functional bottleneck in which many integrases that function in bacteria do not work or present cytotoxic effects in eukaryotic systems [16, 23]. Xu et al. [36] showed that some integrases able to rearrange sequences on plasmid DNA in human cells lost their ability to do so when the target DNA was integrated into the genome, while Andreas et al. [37] showed that alterations such as adding a nuclear localization signal (NLS) at the Φ C31 C-terminus region enhanced its editing capabilities in mouse

cells. As mentioned, we also showed that integrases with similar efficiency levels in bacteria showed varying degrees of efficiency when tested in plant and animal cells. Moreover, despite the increasing levels of interest and publications using serine-integrases in eukaryotic organisms, very few new members have been validated in these systems, with many groups always working with the same limited pool of candidates, mainly φ C31, BxB1, and less frequently, φ R4 and TP901. Considering this scenario, we propose in this protocol a platform to systematically test and identify functional serine integrases for the assembly of genetic circuits in different plant and mammalian organisms.

The platform can be seen as a workflow comprising six main stages: I) *in silico* design and synthesis of plasmid constituents of each effector and reporter module; II) cell acquisition and culture maintenance, which can be subdivided into in vitro culture of established mammalian cell lines, patient cell isolation and plant growth for protoplast isolation according to the model selected; III) DNA delivery; IV) cytotoxicity evaluation; V) molecular analysis, including DNA preparation, primer design, PCR setup and sequencing; and VI) phenotypic analyses by fluorescence microscopy and flow cytometry. Fig 2 shows a schematic overview of the workflow proposed.

Materials and methods

The protocol described in this peer-reviewed article is published on [protocols.io](#) [dx.doi.org/10.17504/protocols.io.rm7vzx945gx1/v1] and is included for printing purposes as [S1 File](#). Sub-protocols specific for each organism group used in the present work are included for printing purposes as [S2–S5 Files](#).

Ethical statement

All work involving cells derived from mammalian organisms must comply with and be performed under strict ethical guidelines. Experiments with human materials must conform to all

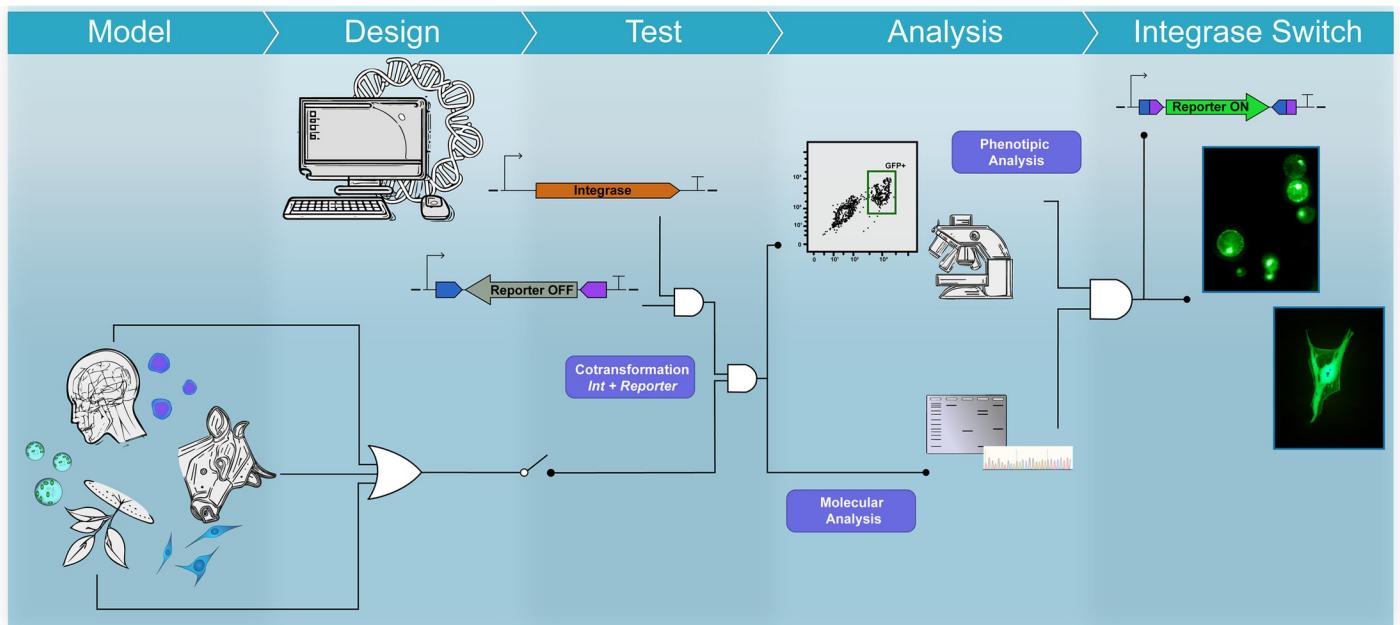


Fig 2. Strategic overview of the serine integrase-based platform for the functional characterization of genetic switch controllers in eukaryotic cells.

<https://doi.org/10.1371/journal.pone.0303999.g002>

relevant institutional and governmental ethics regulations, and appropriate informed consent must be obtained for the use of human blood or patient-derived materials. For the development of the present protocol, Peripheral Blood Mononuclear Cells from healthy donors were collected and used upon approval by the Brazilian National Cancer Institute (INCA) Ethics Committee and signing of Review board-approved informed consent forms by the donors. Experiments involving the use of human-derived stem cells were approved by the ethics committee of Copa D'Or Hospital (CAAE number 60944916.5.0000.5249, approval number 1.791.182). Regarding the acquisition and use of bovine primary fibroblasts, all experimentation performed was approved by the Ethics Committee on the Use of Animals (CEUA) of Embrapa Genetic Resources and Biotechnology (Brasilia, Brazil) in March 2013 (approval reference no. 001/2013).

Experimental design

In silico design of plasmids. These initial steps are crucial for a successful experiment since they include the definition of all the regulatory parts to be used, as well as the rational design of the switch to be tested. The modular aspect of the construction allows the easy and fast assembly of numerous plasmid sets carrying different integrases and their respective reporter cassettes. As indicated in the schematics presented in Fig 3, the basic expression unit for the effector plasmids consists of the promoter sequence most appropriate to the organism in which the test will be performed, followed by the integrase gene and a terminator sequence downstream. For better expression rates, the integrase gene sequence may be codon optimized before synthesis to match the codon usage patterns of the organism or group of interest. This

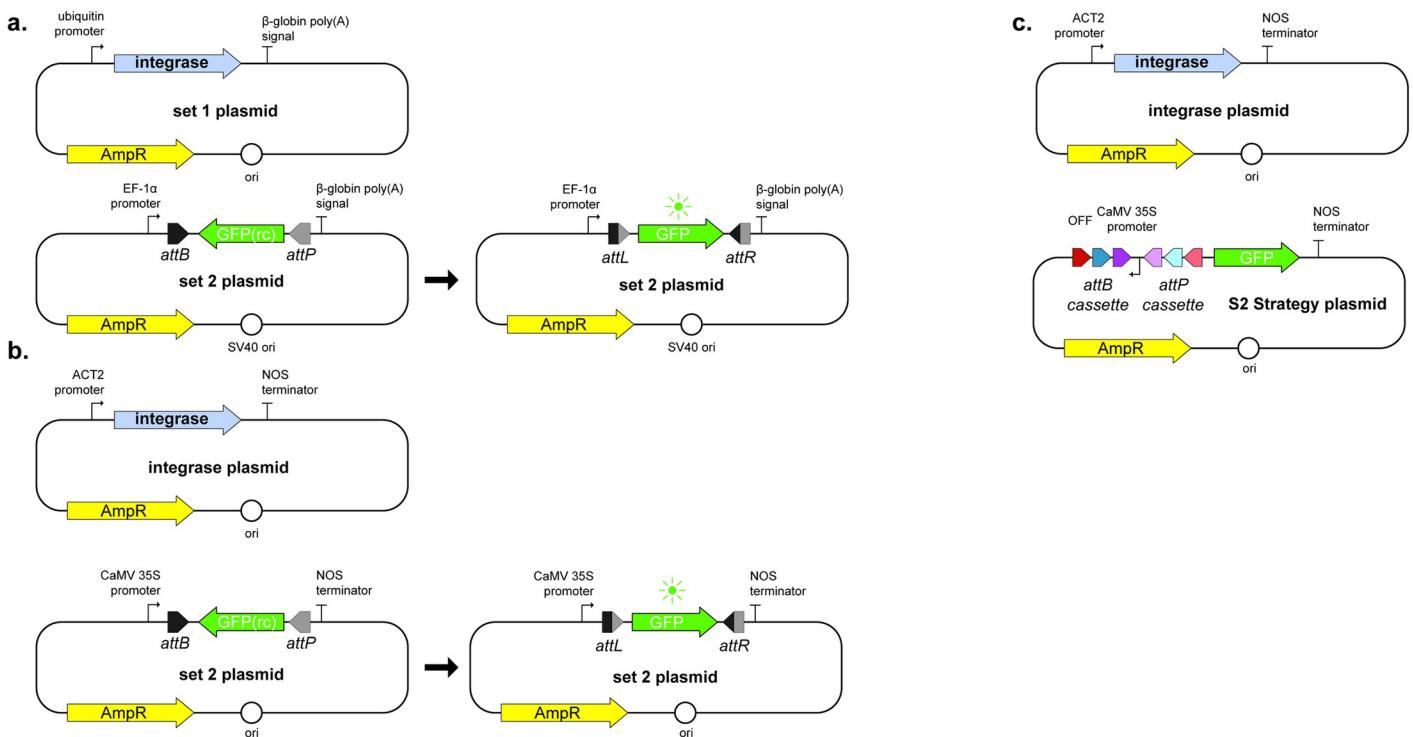


Fig 3. Designed unidirectional genetic switches composed of two sets of synthesized plasmids. (a) and (b) Plasmid sets for S1 strategy in mammalian and plant models, respectively. (c) Plasmid sets for S2 strategy applied for plant models. Here, the promoter sequence, not the *gfp* gene, is under the control of integrase-mediated inversion and activation. Figure adaptation from Gomide et. al., 2020 [16].

<https://doi.org/10.1371/journal.pone.0303999.g003>

expression unit must be inserted into an appropriate backbone that allows cloning and transformant selection for the preparation of DNA in bacteria before assays and proper behavior once delivered into the eukaryotic model. The promoter sequence must render constitutive expression, and weak promoters, i.e. sequences known to promote low transcription rates—therefore affecting mRNA accumulation—, are avoided to reduce the chance of falsely identifying an integrase as nonfunctional in the system when in reality it is due to low expression levels. Additional elements, like RBS and enhancer sequences can be screened and used to fine-tune transcription and increase integrase expression. In this protocol, we selected the promoters pUbC [38] and pAct [39] for use with mammalian and plant cells, respectively.

The reporter unit has a more complex design but still follows the modular aspect of the effector plasmids with the classic downstream sequence of promoter, gene of interest and terminator sequence. The promoter and terminator sequences selected for this protocol in the reporter module were pEF1a [38] and B-globin poly(A) signal for use with mammalian cells and pCaMV35S and NOS terminator for use with plant cells. *gfp* was used as the reporter gene here, although the choice of reporter must take into consideration background fluorescence wavelength and possible emissions from cell constituents such as chloroplasts in the plant cells to avoid emission overlap. Two reporter strategies were designed and named S1 and S2 Strategy. S1 strategy (Fig 3A and 3B) is designed to assess the integrase's capacity for rearranging a coding DNA sequence orientation, for which the reporter gene sequence must be inserted into the construct in an anti-sense position in relation to the promoter, thus in a silenced state—here referred to as the OFF STATE—and flanked by the integrase recombination sites. It is crucial that recombination sites are inserted in opposite directions to each other to ensure inversion of the DNA sequence between them given the recombination dynamics discussed above. The selection of integrases and their cognate *att* site pair should involve considering the eventual presence of starting codons in the *att* sites or recombined *attL* and *attR* site sequences that can interfere with gene expression and reading frames. In the occurrence of such codons, swapping *attB* and *attP* positions can be considered, as long as the final design contains the most downstream site in its reverse complement form.

The genetic switch can be assembled to direct the reorientation of genetic parts other than a gene of interest. In the S2 strategy designed for this platform, the promoter is initially inactive due to its opposite orientation in relation to the reporter gene (Fig 3C). Another improvement in this strategy is the use of a cassette of *att* sites in tandem, which allows the use of the same reporter plasmid to test the activity of different integrases or sequential induction events with different integrases to invert the flanked sequence back to its initial position. Despite broadening the applications of the genetic switch, this strategy demands prior knowledge of the integrases selected or a more thorough evaluation and proper use of controls to ensure their orthogonality.

All plasmids used in the development of this protocol are listed in Table 1 below, with Addgene database accession number for more information including annotated sequence maps. The genetic parts in the constructions can be changed to better suit different organisms.

Cell acquisition and culture maintenance. Another important aspect of the presented protocol is the plasticity in regard to eukaryotic cell models with which it can be performed. Primary bovine fibroblasts were obtained from 14-month-old *Bos indicus* bull oxtail by biopsies following the protocol defined by Freshney [40] with modifications. Briefly, recovered pieces of tissue are washed with 0.05% trypsin to detach cells, which are then cultured in Dulbecco's Modified Eagle Medium supplemented with fetal bovine serum 10% at 37°C and 5% CO₂ atmosphere in culture flasks for at least three passages or until a homogeneous culture of attached fibroblasts can be observed. HEK293T cells can be purchased from cell line culture collections and stored, although some culture passages are recommended prior to cell

Table 1. Plasmids used in the assembly of S1 and S2 strategy sets.

Plasmid Set	Name	Addgene ID
pSG	INCTbiosyn-pEF-gfp(rc)2	127504
	INCTbiosyn-pEF-gfp(rc)4	127505
	INCTbiosyn-pEF-gfp(rc)5	127506
	INCTbiosyn-pEF-gfp(rc)7	127507
	INCTbiosyn-pEF-gfp(rc)9	127508
	INCTbiosyn-pEF-gfp(rc)13	127509
	INCTbiosyn-pEF-gfp(rc)phiC31	127510
	INCTbiosyn-pEF-gfp(rc)Bxb1	127511
pIE	INCTbiosyn-pUB-HspINT2	127512
	INCTbiosyn-pUB-HspINT4	127513
	INCTbiosyn-pUB-HspINT5	127514
	INCTbiosyn-pUB-HspINT7	127515
	INCTbiosyn-pUB-HspINT9	127516
	INCTbiosyn-pUB-HspINT13	127517
	INCTbiosyn-pUB-HspINTphiC31	127518
	INCTbiosyn-pUB-HspINTBxb1	127519
pSP	INCTbiosyn-p35S(rc)2_4_5-gfp	127520
pSG	INCTbiosyn-p35S-gfp(rc)2	127521
	INCTbiosyn-p35S-gfp(rc)4	127522
	INCTbiosyn-p35S-gfp(rc)5	127523
	INCTbiosyn-p35S-gfp(rc)7	127524
	INCTbiosyn-p35S-gfp(rc)9	127525
	INCTbiosyn-p35S-gfp(rc)13	127526
	INCTbiosyn-p35S-gfp(rc)phiC31	127527
	INCTbiosyn-p35S-gfp(rc)Bxb1	127528
pIE	INCTbiosyn-pAct-AtINT2	127529
	INCTbiosyn-pAct-AtINT4	127530
	INCTbiosyn-pAct-AtINT5	127531
	INCTbiosyn-pAct-AtINT7	127532
	INCTbiosyn-pAct-AtINT9	127533
	INCTbiosyn-pAct-AtINT13	127534
	INCTbiosyn-pAct-AtINTphiC31	127535
	INCTbiosyn-pAct-AtINTBxb1	127536

<https://doi.org/10.1371/journal.pone.0303999.t001>

transfection to ensure that metabolically fit cultures are being used. For human donor primary or derived cells used to pursue potential medical applications of the genetic switches, PBMCs were isolated from donor blood via Ficoll density gradient centrifugation, while neural stem cells and human embryonic stem cells from the BR-1 cell line were cultured in neural induction medium (NEM, Advanced DMEM/F12 and neurobasal medium (1:1) with neural induction supplement). All animal and human donor-derived cells and experimentation performed with them must undergo appropriate ethics committee evaluation and approval.

Plant cells derived from *A. thaliana*. Briefly, protoplasts were isolated from young leaves of 5-week-old plants cultivated at 22°C under a 12 h light/12 h dark regimen according to the protocol described by Yoo et al. [41] with modifications. The collected leaves were prepared by making shallow cuts on the adaxial face with a scalpel to enhance internal exposure to the enzymatic solution. In this work, *A. thaliana* leaf cells were disaggregated using an enzymatic

cocktail of 0.2% pectolyase, 0.5% driselase and 1.5% cellulase, but other combinations and adaptations are possible to better digest the cellular walls of cells from other species. The age and overall morphology and health of leaves can have a great impact on protoplast isolation; therefore, the use of young specimens of up to 60 days is recommended.

DNA delivery and controls. Given the particularities of each model, different plasmid delivery strategies are applied depending on the cell line. In general, nonintegrative transfection resulting in transient expression of the effector integrase and reporter gene is sufficient for the functional screening proposed in this protocol. All strategies applied here have in common the concurrent transfection of effector and corresponding reporter plasmids in equimolar ratio and a period of at least 24 h post-transfection before assessing the outcome responses. Different delivery strategies were used depending on the model. Transfection of HEK293T cells and primary bovine fibroblasts was performed by lipid-mediated delivery using Lipofectamine LTX and Plus reagent (Invitrogen) according to the manufacturer's recommendations. Despite being widely used and considered a gold standard method for DNA delivery [42], lipofectamine transformation can result in low transformation efficiency and high cytotoxicity for some cell lines, including human primary cells and non-adherent cultures. Hence, PBMC, NCS and hES cell transformation was carried out by electroporation using an optimized adaptation of the nucleofection method established by Chicaybam et al. [43]. This method yields higher transformation efficiency for some human cells, including primary and stem cells, with low cytotoxic effects at a lower price than the classical nucleoporation method. For plant protoplasts, CaCl₂-PEG chemical transfection [44, 45] was used to carefully deliver the plasmid sets. Although other methods exist for DNA delivery in protoplasts, including agrobacteria-mediated transformation [46] and electroporation [47], PEG-mediated transfection is the most used delivery system due to its easy manipulation, low cost and no need for specific equipment. Another important advantage of this method is that it allows for more gentle manipulation of cells, an important aspect when handling fragile protoplasts following cell wall digestion.

Although the cotransformation of an effector and corresponding reporter plasmid pair is the central aspect of the functional evaluation of an integrase, many control groups containing different plasmid combinations are necessary to validate the obtained results. The first set of controls will be the samples to which only one member from the plasmid pair is delivered to eliminate any background signal or endogenous machinery activity interference that could lead to false-positive results. With these samples, we will have the opportunity to assess fluorescence levels coming from a non-activated reporter in the absence of the effector integrase, for instance, which could be indicative of a leaky system or unwanted promoter activity from the downstream *att* site present. One important example of such occurrence is the observed promoter activity of Int13 *attP* site described by Yang et al. in *E. coli* [15]. An unlikely, although not impossible, fluorescence emission by the integrase itself could also be ruled out in this context. These control groups are also important to identify potential cytotoxic effects from any of the components of the system, especially from the integrase being tested, which may cause DNA damage or integration at endogenous pseudosites [36, 48, 49], disrupting the proper expression of important unrelated genes. It is important to note that for these controls, the plasmid absent from the original pair must be replaced by the same amount of a mock plasmid, normally the empty backbone used in the constructions, to maintain the DNA load concentration and molar ratio. Another set of controls is especially necessary in the case of our S2 strategy or when future applications of the integrases being tested involve the simultaneous presence of more than one in a genetic switch or cascade: in this case, combinations of an effector plasmid with one integrase should be cotransformed with the reporter plasmid

carrying recombination sites of a different integrase to assess orthogonality, i.e., one integrase is not capable of interacting with or rearranging *att* sites other than their cognate site pair.

Positive controls included plasmids carrying unaltered reporter gene sequences under constitutive expression and the design and synthesis of the expected rearranged cassette carrying the reporter gene sequence in its forward orientation flanked by the expected recombined *attL* and *attR* sites. The first group is a classic positive control to ensure cell behaviour and allow equipment calibration. As for the constructions with *attL* and *attR* sites, it is an important group to help evaluate if an observed lowered level or lack of signal from the reporter following integrase activation indicates a negative result or inhibitory transcriptional interference due to the presence of the post-recombination sites. Although serine-integrases are not able to recombine *attL/R* sites, Chao et al. demonstrated that integrase interaction with these sites can alter target expression [50].

As mentioned, the use of the first set of control groups is suitable for evaluating cytotoxicity. Given that one of the steps in the recombination of *att* sites by an integrase involves double-strand breaks in the DNA, the occurrence of off-target activity at pseud-sites causing unintended DNA damage can be a concerning source of toxicity and limit the use of such integrase in the assembly of a genetic switch. Cell integrity markers and colorimetric assays designed to measure cell viability are an easy way to check for deleterious effects. PBMCs and stem cells used were stained with 7-AAD, a live cell impermeable fluorophore that allows the quantification of stained dead cells in the same flow cytometry run used to assess reporter expression. A similar strategy was applied to protoplasts by staining cells with fluorescein diacetate (FDA), although in this case, live cells were stained. For HEK293T cells and primary bovine fibroblasts, an enzymatic MTT assay was performed.

Molecular analyses

To confirm integrase activity, both molecular and phenotypical analysis methodologies should be applied. As we observed previously [16], when screening for integrases capable of rearranging DNA in eukaryotic cells, sometimes the confirmation of sequence inversion by amplification and sequencing was not accompanied by a positive reporter signal. DNA extraction and amplification should follow standard practices established for Sanger sequencing, but primer design must be carefully performed. In addition to confirming DNA inversion, amplification and sequencing are means to confirm proper recombination of *att* sites by directed editing of the integrase not causing point mutations or DNA damage at the core sequencing, gene of interest or surrounding regions. A good strategy is to amplify the 5' end of the gene and upstream formed *attL* site as one fragment and the 3' end of the gene and the downstream *attR* formed site as another fragment, keeping an overlap region between them to ensure good coverage of the whole cassette, as shown by the diagram in Fig 4. It is important to note that when opting to use primers annealing to post recombination *attL* and *attR* sites to increase specificity, the primers must have a high Tm point and PCR setup performed with higher annealing temperatures since part of the primer will inevitably anneal to respective *attB* and *attP* sites to a certain extent.

Obtained amplicons must be sequenced to ensure amplification specificity and correct formation of *attL* and *attR* sites with target coding sequence inversion. For most standard model organisms, especially cell cultures with high transformation efficiency, amplification should be no problem and Sanger sequencing will provide good results whilst still easily accessible in facilities or sequencing companies at low costs. A recommendation is to have fragments sequenced with forward and reverse primers to allow better coverage of both ends of the region of interest. Overlap of these sequences can also help identify mutations from

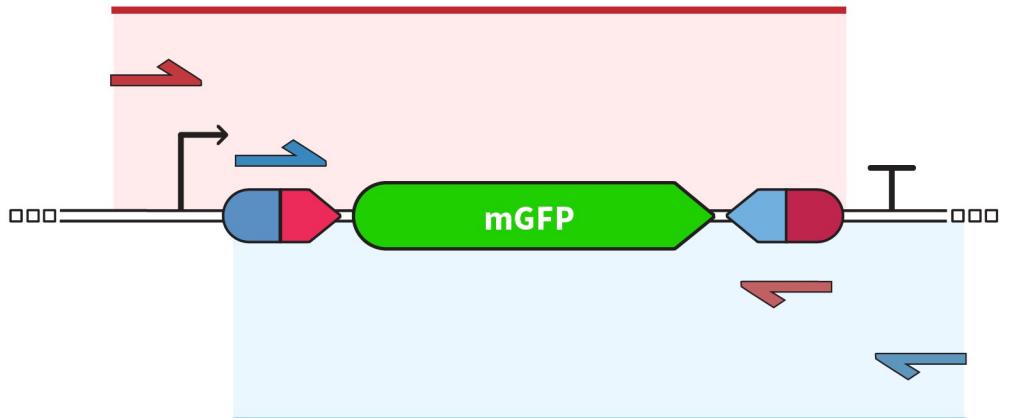


Fig 4. Oligonucleotide pairs and expected amplification regions in the two-amplicon strategy. Figure adaptation from Gomide et. al., 2020 [16].

<https://doi.org/10.1371/journal.pone.0303999.g004>

sequencing error instead of integrase-caused DNA damage. However, sequencing method selection can be affected by other factors. When working with cells of low transformation efficiency that could impair DNA delivery, tissues instead of cell culture that can result in mixed population with more non-transformed than transformed cells, or screening integrases with very low recombination efficiency, a more robust and sensitive sequencing method may be used. Nanopore DNA sequencing is a good alternative in these scenarios [51], more easily detecting low copy numbers of edited sequences in a mixed population. This methodology presents other advantages when compared to Sanger sequencing, including the possibility to better evaluate the occurrence and rate of integrase-related mutations and DNA damage due to its high accuracy and possibility of reads quantification for integrated stoichiometry analysis.

Phenotypical analyses

Reporter activation can be assessed by fluorescence visualization or quantification. Fluorescence microscopy is a fast way to observe reporter expression activation but will result only in qualitative data. For accurate quantification to allow efficiency ranking between all integrases tested flow cytometry must be performed. Multiple time points should be evaluated to identify the best incubation period for each model. For the models used in our work, 48h was defined as standard incubation time, with eGFP signal decrease after 72h post-incubation in mammalian cells. Due to the short survival in culture conditions, flow cytometry of treated protoplasts was performed after 24 hours following transformation. High event count must be used, and all control groups must be considered to ensure a proper definition of gates to differentiate between eGFP positive and negative cell populations.

Assays must be done with at least three independent biological replicates; technical replicates should be included to account for measurement and equipment variations. Statistical method choice must be done considering the specificities of each new model and experimental design selected. For our platform, quantitative data were analyzed using the nonparametric Kruskal-Wallis test, with Dunn's test as post hoc for pairwise comparisons. Although One-way ANOVA is typically used for multiple comparisons, nonparametric tests do not rely on assumptions regarding normal distribution and variance homogeneity, which can be hard to confirm when working with small sample sets.

Expected results

We first developed our platform to assess the functionality of 6 integrases previously characterized only in *E. coli*, named Int2, Int4, Int5, Int7, Int9 and Int13 [15]. Original bacterial host and sequence of *attB/attP* sites for each Int are presented at Table 2. However, the plasticity of our workflow makes it suitable for the initial screening in eukaryotic cells of any other new serine integrases identified in silico from genome databases or with ones previously tested only in prokaryotic cells with very little adaptation needed.

In our work [16], we were able to confirm the successful recombination of recognition sites and formation of correct *attL* and *attR* sites with consequent inversion of the *egfp* gene used as a reporter for all the integrases studied. Despite the positive results at the molecular level, *mgfp* expression levels after switch activation varied depending on the integrase and organism used when cells were analyzed by flow cytometry or fluorescence microscopy. Overall, for the S1 experimental design Int13 switches resulted in higher levels of fluorescence emission for all the organisms used, followed by Int9 and Int4 in bovine fibroblasts and protoplasts. Int5 switch yielded the lowest levels of signal despite molecular confirmation of DNA inversion, possibly an indication of low frequencies of recombination in the systems (Fig 5). Integrases phiC31 and BxB1, used for comparison given their well established applications in mammal and plant systems, also resulted in varying signal levels depending on the cell used, although overall results were similar to the ones obtained when using Int13. Interestingly, the S2 promoter switch construction led to a result compatible with the S1 GFP switch in plant protoplasts. However, Int2 exhibited a much higher number of EGFP-positive cells when the promoter was flipped than when the targeted gene was [16].

These analyses have proven the efficiency of our Int-based platform for the functional characterization of these enzymes, as well as its robustness for the further investigation of Ints as genetic switch controllers in eukaryotic cells. Considering that all six integrases showed similar efficiency compared to each other when first used in *E. coli*, the broader range of efficiency obtained in eukaryotic cells can be advantageous when considering the assembly of multicomponent circuits, where variations on potency can help regulate flow rates at defined points in a cascade, allowing a fine-tuning of the system.

Although not further investigated in the present work, the variations in *gfp* expression levels observed both amongst integrases or between mammalian and plant models were partially unexpected. Some works studying integrase applications on eukaryotic models concluded that such variations could occur not only between distant groups like mammalian cells and plants or yeast but also when comparing more closely related organisms, as reported by Xu et al. (2013) [36] when comparing the use of more than ten Ints in human and murine cells. One example can be easily observed with Int TP901: In the mentioned work, this Int was found functional in human cells but incapable of rearranging DNA in mouse stem cells. The same group reported its functionality in *Saccharomyces cerevisiae* with considerably higher efficiency [23], while Guiziou et al. (2023) [52] found no TP901 activity in *Arabidopsis*. Despite reporting and discussing these variations, none of the works presented a deeper investigation into their biological causes. Possible clues can be found in the work published by Chao and collaborators (2021) [50]. They measured recombination kinetic parameters for six Ints, including TP901, to propose a prediction model for systematically reporting Int recombination activity. One important finding in their work is that different Ints can present varying expression levels under the same promoter, and Int concentration can also impact the recombination rate. For instance, phiC31 showed higher recombination rates when under transcriptional control of a weaker Ub promoter than when controlled by the strong CMV promoter.

Table 2. Bacterial host and sequence of attachment sites for the six integrases used by Gomide and collaborators (2020) [16].

Adaptation from Yang et al (2014) [15].

<https://doi.org/10.1371/journal.pone.0303999.t002>

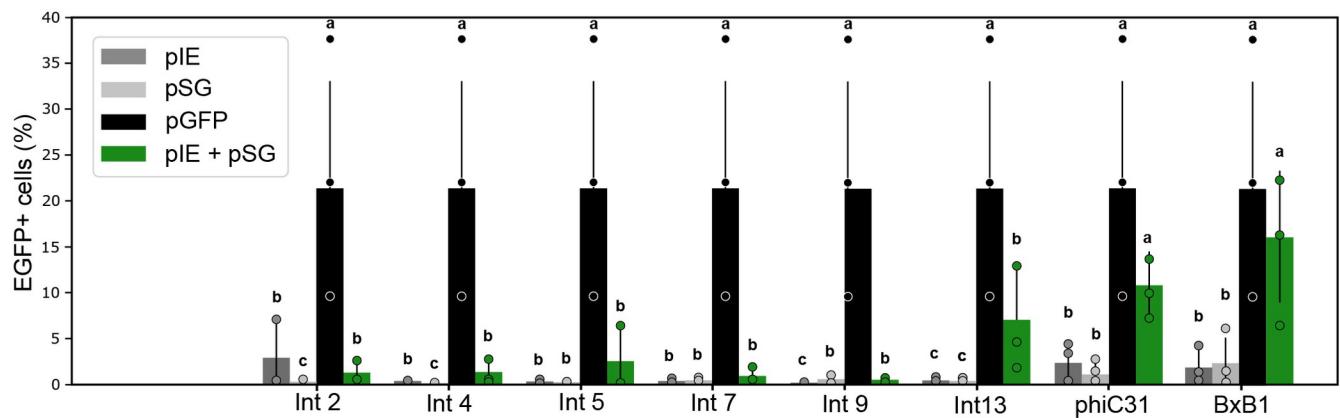
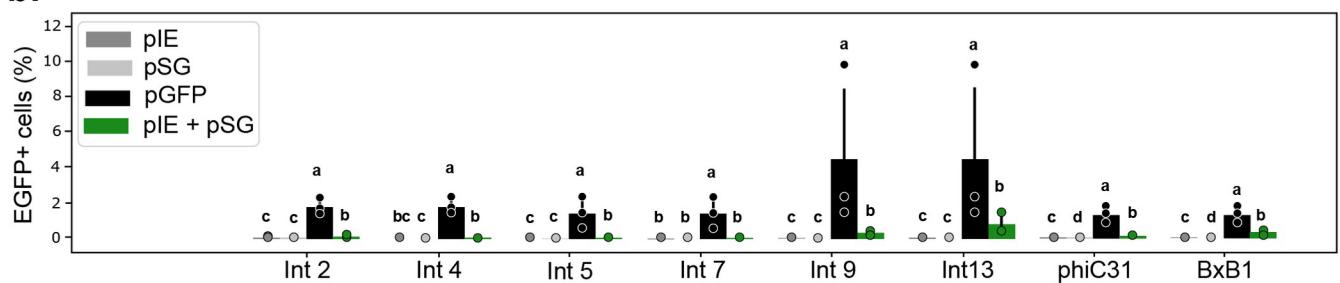
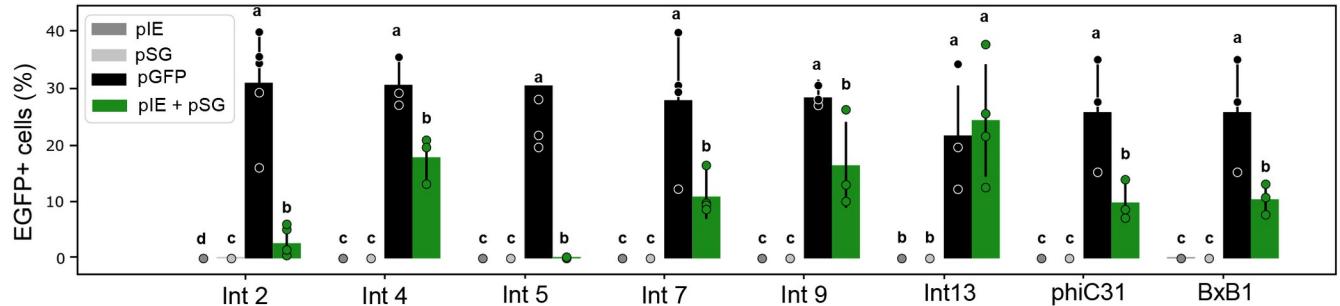
a.**b.****c.**

Fig 5. EGFP fluorescent cell ratio between Int S1 treatments in the different eukaryotic systems analyzed (a, human cell; b, bovine fibroblasts cells; and c, *A. thaliana* protoplasts). pIE: effector Integrase Expression plasmid; pSG: reporter Switch GFP plasmid; pGFP: positive control with constitutive expression of *gfp*. Figure adaptation from Gomide et al., 2020 [16].

<https://doi.org/10.1371/journal.pone.0303999.g005>

These results highlight the complexity of Int efficiency control and how organism-dependent molecular context can affect Int behavior.

Limitations

Although the Int-based platform proved efficient and robust for the functional validation of integrases in eukaryotic cells, some limitations exist. Studies have revealed that Ints are typically very specific to their cognate *att* sites. However, Int shows low levels of off-target activity at pseudo-site sequences similar to their natural *att* sites [17, 53–55]. Additionally, Ints can

cause residual DNA damage or mutations when acting in cells that are not their natural hosts. Various types of damage have been found, including mutations and deletions of the *att* sites, which can make them refractory to later reactions [56]. Cell cytotoxicity is another limiting factor observed when using Ints [23, 36].

As for technical limitations in our platform, the main one we should consider is the discovery of new Ints to be tested since a complete system involves not only the integrase itself but also the identification of the cognate *attB/attP* site pair, which can be more complex to identify than the recombinase gene [15]. The use of RDFs to allow controlled bidirectional recombination is another factor that can broaden Int applications and should be considered. However, limitations including the need of one specific RDF protein for each new serine-integrase, the availability of only a few Int-RDF pairs identified at the moment and laborious validation and optimization required to establish new tools present significant challenges for their use. Regarding analytics, implementing quantitative methods for molecular detection of DNA rearrangement like qPCR assays can bring another valuable parameter for Int efficiency classification. The use of a more robust method for sequencing, like nanopore NGS, can be an alternative in some situations, especially when dealing with low recombination efficiency or a mixed population of edited and not edited targets if an inadequate representation of edited molecules could be preventing identification of DNA rearrangement.

Supporting information

S1 File. Step-by-step protocol collection.

(PDF)

S2 File. Step-by-step protocol for assembly of a serine integrase-based platform for functional validation of genetic switch controllers in eukaryotic cells-human.

(PDF)

S3 File. Step-by-step protocol for assembly of a serine integrase-based platform for functional validation of genetic switch controllers in eukaryotic cells-animal.

(PDF)

S4 File. Step-by-step protocol for assembly of a serine integrase-based platform for functional validation of genetic switch controllers in eukaryotic cells-plant.

(PDF)

S5 File. Step-by-step protocol molecular analyses.

(PDF)

Author Contributions

Conceptualization: Martín H. Bonamino, Elibio Rech.

Funding acquisition: Elibio Rech.

Investigation: Marco A. de Oliveira, Lilian H. Florentino, Thais T. Sales, Rayane N. Lima, Luciana R. C. Barros, Cintia G. Limia, Mariana S. M. Almeida, Maria L. Robledo, Leila M. G. Barros, Eduardo O. Melo, Stevens K. Rehen.

Resources: Eduardo O. Melo, Stevens K. Rehen, Martín H. Bonamino, Elibio Rech.

Writing – original draft: Marco A. de Oliveira, Lilian H. Florentino, Thais T. Sales, Rayane N. Lima, Luciana R. C. Barros, Cintia G. Limia, Eduardo O. Melo, Daniela M. Bittencourt, Stevens K. Rehen, Martín H. Bonamino, Elibio Rech.

Writing – review & editing: Marco A. de Oliveira, Rayane N. Lima, Luciana R. C. Barros, Daniela M. Bittencourt, Martín H. Bonamino, Elibio Rech.

References

1. Arber W. Host-controlled modification of bacteriophage. *Annu Rev Microbiol.* 1965; 19: 365–378. <https://doi.org/10.1146/annurev.mi.19.100165.002053> PMID: 5318444
2. Fogg PCM, Colloms S, Rosser S, Stark M, Smith MCM. New applications for phage integrases. *J Mol Biol.* 2014; 426: 2703–2716. <https://doi.org/10.1016/j.jmb.2014.05.014> PMID: 24857859
3. Sheth RU, Wang HH. DNA-based memory devices for recording cellular events. *Nat Rev Genet.* 2018; 19: 718–732. <https://doi.org/10.1038/s41576-018-0052-8> PMID: 30237447
4. Tian X, Zhou B. Strategies for site-specific recombination with high efficiency and precise spatiotemporal resolution. *J Biol Chem.* 2021; 296: 100509. <https://doi.org/10.1016/j.jbc.2021.100509> PMID: 33676891
5. Meinke G, Bohm A, Hauber J, Pisabarro MT, Buchholz F. Cre recombinase and other tyrosine recombinases. *Chem Rev.* 2016; 116: 12785–12820. <https://doi.org/10.1021/acs.chemrev.6b00077> PMID: 27163859
6. Kim H, Kim M, Im S-K, Fang S. Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes. *Lab Anim Res.* 2018; 34: 147–159. <https://doi.org/10.5625/lar.2018.34.4.147> PMID: 30671100
7. Abremski K, Hoess R, Sternberg N. Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. *Cell.* 1983; 32: 1301–1311. [https://doi.org/10.1016/0092-8674\(83\)90311-2](https://doi.org/10.1016/0092-8674(83)90311-2) PMID: 6220808
8. Nakayama M. VCre/VloxP and SCre/SloxP as reliable site-specific recombination systems for genome engineering. *Methods Mol Biol Clifton NJ.* 2023; 2637: 161–180. https://doi.org/10.1007/978-1-0716-3016-7_13 PMID: 36773146
9. Merrick CA, Zhao J, Rosser SJ. Serine integrases: advancing synthetic biology. *ACS Synth Biol.* 2018; 7: 299–310. <https://doi.org/10.1021/acssynbio.7b00308> PMID: 29316791
10. Stark WM. Making serine integrases work for us. *Curr Opin Microbiol.* 2017; 38: 130–136. <https://doi.org/10.1016/j.mib.2017.04.006> PMID: 28599144
11. Brown WRA, Lee NCO, Xu Z, Smith MCM. Serine recombinases as tools for genome engineering. *Methods.* 2011; 53: 372–379. <https://doi.org/10.1016/jymeth.2010.12.031> PMID: 21195181
12. Snoeck N, De Mol ML, Van Herpe D, Goormans A, Maryns I, Coussement P, et al. Serine integrase recombinational engineering (SIRE): A versatile toolbox for genome editing. *Biotechnol Bioeng.* 2019; 116: 364–374. <https://doi.org/10.1002/bit.26854> PMID: 30345503
13. Zhao J, Pokhilko A, Ebenhö O, Rosser SJ, Colloms SD. A single-input binary counting module based on serine integrase site-specific recombination. *Nucleic Acids Res.* 2019; 47: 4896–4909. <https://doi.org/10.1093/nar/gkz245> PMID: 30957849
14. Abioye J, Lawson-Williams M, Lecanda A, Calhoon B, McQue AL, Colloms SD, et al. High fidelity one-pot DNA assembly using orthogonal serine integrases. *Biotechnol J.* 2023; 18: e2200411. <https://doi.org/10.1002/biot.202200411> PMID: 36504358
15. Yang L, Nielsen AAK, Fernandez-Rodriguez J, McClune CJ, Laub MT, Lu TK, et al. Permanent genetic memory with >1-byte capacity. *Nat Methods.* 2014; 11: 1261–1266. <https://doi.org/10.1038/nmeth.3147> PMID: 25344638
16. Gomide MS, Sales TT, Barros LRC, Limia CG, de Oliveira MA, Florentino LH, et al. Genetic switches designed for eukaryotic cells and controlled by serine integrases. *Commun Biol.* 2020; 3: 255. <https://doi.org/10.1038/s42003-020-0971-8> PMID: 32444777
17. Gupta K, Sharp R, Yuan JB, Li H, Van Duyne GD. Coiled-coil interactions mediate serine integrase directionality. *Nucleic Acids Res.* 2017; 45: 7339–7353. <https://doi.org/10.1093/nar/gkx474> PMID: 28549184
18. Li H, Sharp R, Rutherford K, Gupta K, Van Duyne GD. Serine integrase attP binding and specificity. *J Mol Biol.* 2018; 430: 4401–4418. <https://doi.org/10.1016/j.jmb.2018.09.007> PMID: 30227134
19. Mandali S, Johnson RC. Control of the serine integrase reaction: roles of the coiled-coil and helix e regions in DNA site synapsis and recombination. *J Bacteriol.* 2021; 203: e00703–20. <https://doi.org/10.1128/JB.00703-20> PMID: 34060907
20. Yuan P, Gupta K, Van Duyne GD. Tetrameric structure of a serine integrase catalytic domain. *Struct Lond Engl* 1993. 2008; 16: 1275–1286. <https://doi.org/10.1016/j.str.2008.04.018> PMID: 18682229

21. Fan H-F, Hsieh T-S, Ma C-H, Jayaram M. Single-molecule analysis of ϕ C31 integrase-mediated site-specific recombination by tethered particle motion. *Nucleic Acids Res.* 2016; 44: 10804–10823. <https://doi.org/10.1093/nar/gkw861> PMID: 27986956
22. Turan S, Zehe C, Kuehle J, Qiao J, Bode J. Recombinase-mediated cassette exchange (RMCE)—a rapidly-expanding toolbox for targeted genomic modifications. *Gene.* 2013; 515: 1–27. <https://doi.org/10.1016/j.gene.2012.11.016> PMID: 23201421
23. Z X, Wr B. Comparison and optimization of ten phage encoded serine integrases for genome engineering in *Saccharomyces cerevisiae*. *BMC Biotechnol.* 2016; 16. <https://doi.org/10.1186/s12896-016-0241-5> PMID: 26860416
24. Fogg PCM, Younger E, Fernando BD, Khaleel T, Stark WM, Smith MCM. Recombination directionality factor gp3 binds ϕ C31 integrase via the zinc domain, potentially affecting the trajectory of the coiled-coil motif. *Nucleic Acids Res.* 2018; 46: 1308–1320. <https://doi.org/10.1093/nar/gkx1233> PMID: 29228292
25. Fan H-F, Su B-Y, Ma C-H, Rowley PA, Jayaram M. A bipartite thermodynamic-kinetic contribution by an activating mutation to RDF-independent excision by a phage serine integrase. *Nucleic Acids Res.* 2020; 48: 6413–6430. <https://doi.org/10.1093/nar/gkaa401> PMID: 32479633
26. Kuhstoss S, Richardson MA, Rao RN. Plasmid cloning vectors that integrate site-specifically in *Streptomyces* spp. *Gene.* 1991; 97: 143–146. [https://doi.org/10.1016/0378-1119\(91\)90022-4](https://doi.org/10.1016/0378-1119(91)90022-4) PMID: 1995427
27. Durrant MG, Fanton A, Tycko J, Hinks M, Chandrasekaran SS, Perry NT, et al. Systematic discovery of recombinases for efficient integration of large DNA sequences into the human genome. *Nat Biotechnol.* 2022. <https://doi.org/10.1038/s41587-022-01494-w> PMID: 36217031
28. Ba F, Liu Y, Liu W-Q, Tian X, Li J. SYMBIOSIS: synthetic manipulable biobricks via orthogonal serine integrase systems. *Nucleic Acids Res.* 2022; 50: 2973–2985. <https://doi.org/10.1093/nar/gkac124> PMID: 35191490
29. Brophy JAN, Voigt CA. Principles of genetic circuit design. *Nat Methods.* 2014; 11: 508–520. <https://doi.org/10.1038/nmeth.2926> PMID: 24781324
30. Bonnet J, Yin P, Ortiz ME, Subsoontorn P, Endy D. Amplifying genetic logic gates. *Science.* 2013; 340: 599–603. <https://doi.org/10.1126/science.1232758> PMID: 23539178
31. Roquet N, Soleimany AP, Ferris AC, Aaronson S, Lu TK. Synthetic recombinase-based state machines in living cells. *Science.* 2016; 353: aad8559. <https://doi.org/10.1126/science.aad8559> PMID: 27463678
32. Courbet A, Endy D, Renard E, Molina F, Bonnet J. Detection of pathological biomarkers in human clinical samples via amplifying genetic switches and logic gates. *Sci Transl Med.* 2015; 7: 289ra83. <https://doi.org/10.1126/scitranslmed.aaa3601> PMID: 26019219
33. Nielsen AAK, Der BS, Shin J, Vaidyanathan P, Paralanov V, Strychalski EA, et al. Genetic circuit design automation. *Science.* 2016; 352: aac7341. <https://doi.org/10.1126/science.aac7341> PMID: 27034378
34. Jones TS, Oliveira SMD, Myers CJ, Voigt CA, Densmore D. Genetic circuit design automation with Cello 2.0. *Nat Protoc.* 2022; 17: 1097–1113. <https://doi.org/10.1038/s41596-021-00675-2> PMID: 35197606
35. Park Y, Espah Borujeni A, Gorochowski TE, Shin J, Voigt CA. Precision design of stable genetic circuits carried in highly-insulated *E. coli* genomic landing pads. *Mol Syst Biol.* 2020; 16: e9584. <https://doi.org/10.1525/msb.20209584> PMID: 32812710
36. Xu Z, Thomas L, Davies B, Chalmers R, Smith M, Brown W. Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnol.* 2013; 13: 87. <https://doi.org/10.1186/1472-6750-13-87> PMID: 24139482
37. Andreas S, Schwenk F, Küter-Luks B, Faust N, Kühn R. Enhanced efficiency through nuclear localization signal fusion on phage PhiC31-integrase: activity comparison with Cre and FLPe recombinase in mammalian cells. *Nucleic Acids Res.* 2002; 30: 2299–2306. <https://doi.org/10.1093/nar/30.11.2299> PMID: 12034816
38. Matsuda T, Cepko CL. Electroporation and RNA interference in the rodent retina in vivo and in vitro. *Proc Natl Acad Sci U S A.* 2004; 101: 16–22. <https://doi.org/10.1073/pnas.2235688100> PMID: 14603031
39. Rech EL, Vianna GR, Aragão FJL. High-efficiency transformation by biolistics of soybean, common bean and cotton transgenic plants. *Nat Protoc.* 2008; 3: 410–418. <https://doi.org/10.1038/nprot.2008.9> PMID: 18323812
40. Wickramasinghe SN. Culture of animal cells. A manual of basic technique, 3rd edn. R. Ian Freshney, Wiley-Liss, Inc: New York. xxiv + 486 pages (1994). *Cell Biochem Funct.* 1996; 14: 75–76.
41. Yoo S-D, Cho Y-H, Sheen J. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat Protoc.* 2007; 2: 1565–1572. <https://doi.org/10.1038/nprot.2007.199> PMID: 17585298

42. Cardarelli F, Digiacomo L, Marchini C, Amici A, Salomone F, Fiume G, et al. The intracellular trafficking mechanism of Lipofectamine-based transfection reagents and its implication for gene delivery. *Sci Rep.* 2016; 6: 25879. <https://doi.org/10.1038/srep25879> PMID: 27165510
43. Chicaybam L, Barcelos C, Peixoto B, Carneiro M, Limia CG, Redondo P, et al. An efficient electroporation protocol for the genetic modification of mammalian cells. *Front Bioeng Biotechnol.* 2016; 4: 99. <https://doi.org/10.3389/fbioe.2016.00099> PMID: 28168187
44. Krens FA, Molendijk L, Wullems GJ, Schilperoort RA. In vitro transformation of plant protoplasts with Ti-plasmid DNA. *Nature.* 1982; 296: 72–74. <https://doi.org/10.1038/296072a0>
45. Paszkowski J, Shillito RD, Saul M, Mandák V, Hohn T, Hohn B, et al. Direct gene transfer to plants. *EMBO J.* 1984; 3: 2717–2722. <https://doi.org/10.1002/j.1460-2075.1984.tb02201.x> PMID: 16453573
46. De Buck S, Jacobs A, Van Montagu M, Depicker A. Agrobacterium tumefaciens transformation and cotransformation frequencies of *Arabidopsis thaliana* root explants and tobacco protoplasts. *Mol Plant-Microbe Interact MPMI.* 1998; 11: 449–457. <https://doi.org/10.1094/MPMI.1998.11.6.449> PMID: 9612943
47. Hauptmann RM, Ozias-Akins P, Vasil V, Tabaeizadeh Z, Rogers SG, Horsch RB, et al. Transient expression of electroporated DNA in monocotyledonous and dicotyledonous species. *Plant Cell Rep.* 1987; 6: 265–270. <https://doi.org/10.1007/BF00271995> PMID: 24248756
48. Malla S, Dafnis-Calas F, Brookfield JFY, Smith MCM, Brown WRA. Rearranging the centromere of the human Y chromosome with phiC31 integrase. *Nucleic Acids Res.* 2005; 33: 6101–6113. <https://doi.org/10.1093/nar/gki922> PMID: 16246911
49. Sangiorgi E, Shuhua Z, Capecchi MR. In vivo evaluation of PhiC31 recombinase activity using a self-excision cassette. *Nucleic Acids Res.* 2008; 36: e134. <https://doi.org/10.1093/nar/gkn627> PMID: 18829714
50. Chao G, Travis C, Church G. Measurement of large serine integrase enzymatic characteristics in HEK293 cells reveals variability and influence on downstream reporter expression. *FEBS J.* 2021; 288: 6410–6427. <https://doi.org/10.1111/febs.16037> PMID: 34043859
51. Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol.* 2021; 39: 1348–1365. <https://doi.org/10.1038/s41587-021-01108-x> PMID: 34750572
52. Guizou S, Maranas CJ, Chu JC, Nemhauser JL. An integrase toolbox to record gene-expression during plant development. *Nat Commun.* 2023; 14: 1844. <https://doi.org/10.1038/s41467-023-37607-5> PMID: 37012288
53. Hu Z, Chen L, Jia C, Zhu H, Wang W, Zhong J. Screening of potential pseudo att sites of *Streptomyces* phage ΦC31 integrase in the human genome. *Acta Pharmacol Sin.* 2013; 34: 561–569. <https://doi.org/10.1038/aps.2012.173> PMID: 23416928
54. Herisse M, Porter JL, Guerillot R, Tomita T, Goncalves Da Silva A, Seemann T, et al. The ΦBT1 large serine recombinase catalyzes DNA integration at pseudo-attB sites in the genus Nocardia. *PeerJ.* 2018; 6: e4784. <https://doi.org/10.7717/peerj.4784> PMID: 29740520
55. Bessen JL, Afeyan LK, Dančík V, Koblan LW, Thompson DB, Leichner C, et al. High-resolution specificity profiling and off-target prediction for site-specific DNA recombinases. *Nat Commun.* 2019; 10: 1937. <https://doi.org/10.1038/s41467-019-09987-0> PMID: 31028261
56. Calos MP. Phage integrases for genome editing. In: Cathomen T, Hirsch M, Porteus M, editors. *Genome editing: the next step in gene therapy.* New York, NY: Springer; 2016. pp. 81–91.