12804–12816 *Nucleic Acids Research, 2020, Vol. 48, No. 22* *Published online 3 December 2020*
doi: 10.1093/nar/gkaa1140

HK022 bacteriophage Integrase mediated RMCE as a potential tool for human gene therapy

Amer Elias[1], Hala Kassis[1], Suha Abd Elkader[1], Natasha Gritsenko[1], Alessio Nahmad[1], Hodaya Shir[1], Liana Younis[1], Atheer Shannan[1], Hideki Aihara[2], Gali Prag[1], Ezra Yagil[1] and Mikhail Kolot[1,*]

[1]Department of Biochemistry and Molecular Biology, School of Neurobiology, Biochemistry &Biophysics, Tel-Aviv University, Tel-Aviv, Israel and [2]Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota TwinCities, Minneapolis, MN, 55455, USA

Received August 13, 2020; Revised October 31, 2020; Editorial Decision November 04, 2020; Accepted November 08, 2020

ABSTRACT

HK022 coliphage site-specific recombinase Integrase (Int) can catalyze integrative site-specific recombination and recombinase-mediated cassette exchange (RMCE) reactions in mammalian cell cultures. Owing to the promiscuity of the 7 bp overlap sequence in its *att* sites, active '*attB'* sites flanking human deleterious mutations were previously identified that may serve as substrates for RMCE reactions for future potential gene therapy. However, the wild type Int proved inefficient in catalyzing such RMCE reactions. To address this low efficiency, variants of Int were constructed and examined by integrative site-specific recombination and RMCE assays in human cells using native '*attB'* sites. As a proof of concept, various Int derivatives have demonstrated successful RMCE reactions using a pair of native '*attB'* sites that were inserted as a substrate into the human genome. Moreover, successful RMCE reactions were demonstrated in native locations of the human *CTNS* and *DMD* genes whose mutations are responsible for Cystinosis and Duchene Muscular Dystrophy diseases, respectively. This work provides a steppingstone for potential downstream therapeutic applications.

INTRODUCTION

Gene therapy and genome editing are promising approaches used for gene and cell-engineering modifications in biomedical research, regenerative medicine, and synthetic biology (1–3). Targeted genome editing is carried out by sequence-specific endonucleases systems such as Zinc Finger Nucleases (ZFN), Transcription Activator-Like Effector Nucleases (TALENs), and Clustered Regularly Interspaced Short Palindromic Repeats associated protein-9 nuclease (CRISPR-Cas9) (4–7). Whereas, random genome integration has been demonstrated by the viral and transposon-based systems (8–13). A genome editing event performed by any of these systems is subject to various safety concerns such as insertional mutagenesis, transgene silencing, off-target editing, unrestrained genomic double-strand breaks and oncogene activation. These concerns especially pertain to cells with high proliferative potential, such as human pluripotent stem cells (hPSCs) (14–20). Additionally, the efficiency of these systems is limited by the insert size. Strategies that involve the insertion of linear donor DNA, and in particular those involving over 5 kb multigene constructs, often lead to illegitimate genome integration and/or incomplete integrants (21–24).

An alternative class of tools, site-specific recombinases (SSRs), can address safety concerns and insert size limitation. SSRs catalyze recombination reactions between two specific short DNA recombination sites (RSs). The location and relative orientation of the RSs defines whether a recombination event will result in an integration, excision, inversion, or translocation reaction. The use of SSRs for targeted genome manipulation relies on their efficiency and specificity for recombining only at their respective RSs. For these reasons, an extensive repertoire of SSRs including the most used Cre, FLP and phiC31, are widely harnessed in genome manipulations, gene therapy, developmental and synthetic biology (25–28). SSRs are non-viral and do not rely on host cell machinery to achieve transgenesis, hence, providing attractive genome editing tools for use in human cells. Although, for some of the SSRs, the reverse excision reaction is kinetically favored over their integration reaction (27), which leads to poor integration efficiency. This limitation is overcome by the recombinase-mediated cassette exchange (RMCE) method (29,30).

The RMCE method is based on a reciprocal exchange of DNA fragments that consists of two crossover events within two incompatible RS pairs that are localized on both sides of the cassette of interest. This strategy has the capa-

*To whom correspondence should be addressed. Tel.: +972 3 6409441; Fax: +972 3 6406834; Email: kolott@tauex.tau.ac.il

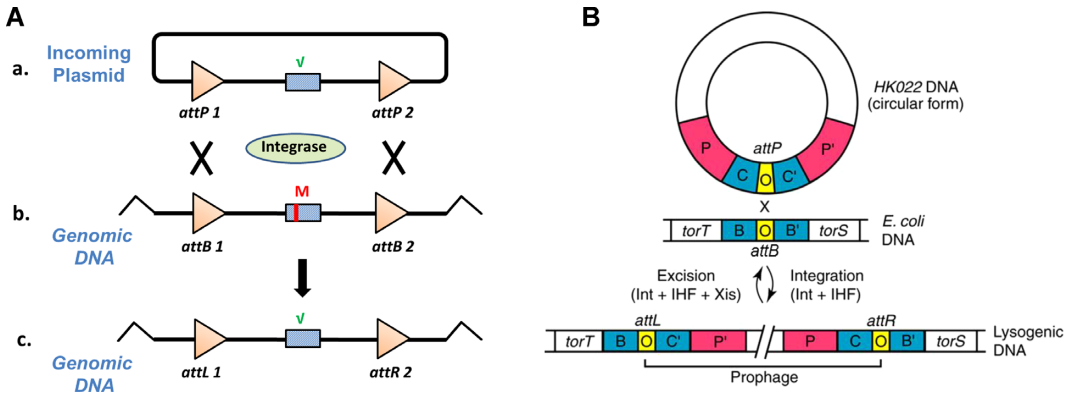

Figure 1. (**A**) Scheme of an Int-based RMCE reaction in which a mutated genomic sequence is exchanged with a cassette of interest. a. Incoming plasmid. b. Genomic mutated (M) sequence. c. Corrected (√) RMCE genome. Triangles – relevant *att* sites. (**B**) Phage HK022 infected *E. coli*, the infected and circularized phage integrates into the host genome via an Int-catalyzed *attP* x *attB* recombination reaction forming a lysogenic host, in which the inserted prophage is flanked by the recombinant *attL* and *attR* sites. O is the overlap, P, B and C are Int binding sites.

bility to swap large transgene cassettes (>100 kb) (31). To achieve reciprocal DNA cassette exchange, cells with the target DNA sequence, flanked by two hetero-specific RSs, are transfected with two plasmids. The replacement cassette of interest, flanked by the compatible RSs, is introduced by a donor plasmid, while another plasmid expresses the cognate SSR (Figure 1A) (32). RMCE can be used as a gene therapy tool by replacing a harmfully mutated genomic sequence flanked by two incompatible RSs (Figure 1A, b) with a plasmid-borne sequence of interest flanked by matching RSs (Figure 1A, a). A molar excess of the donor plasmid over the genomic target sequence will drive the reaction towards replacement by the sequence of interest (27). A successful RMCE reaction results in a 'clean' correction of the mutation without the use of selection markers or other undesired sequences (Figure 1A, c) (32–34). Furthermore, In recent years, RMCE has been used substantially in various research areas including the generation of induced pluripotent stem cells (35), production of therapeutic monoclonal antibodies (36), and in combination with other genome editing systems, such as ZFN, TALENs (37) and CRISPR/Cas (27,38).

In this work, we focused on the integrase (Int) site-specific recombinase of *Escherichia coli* HK022 bacteriophage that belongs to the tyrosine family of SSRs. It naturally catalyzes phage genome integration and excision in a mechanism very similar to that of the well-known lambda coliphage integrase (Figure 1B). In the lysogenic cycle, a recombination reaction between the *attP* site of the injected and circularized phage genome and the bacterial recombination site *attB* results in the formation of the integrated prophage (39). In the case of phage HK022, *attB* (or BOB') is only 21 bp long, comprising a 7 bp central overlap (O) sequence which is the site of the recombination event. O is flanked by two palindromic 7 bp sites (B and B', see also Table 1) that are binding sites of Int. The longer *attP* (or POP') RS of the phage carries a core (COC') similar to *attB* and is flanked by two longer arms (P and P') that are essential for integration in *E. coli.* The 7 bp O sequences of the recombining *att* pairs, *attB* and *attP*, are identical (Figure 1B) (40). We have previously shown that HK022's wild type (*w.t.*) Int, as well as a codon-optimized for human cells

Table 1. List of *attB* sites

***attB* sites**	**B O B`**	**Reference**
w.t.	GCACTTTaggtgaaAAAGGTT	(46)
consensus	NNACTTTnnnnnnnAAAGGNN G A C T T	(46)
ATM4	TTTCTTTgactcagAAAGGCG	(46)
HEXA3	ACACTTTaccaatgAAAGTGA	(46)
CTNS1	TCACTTTggtacagAAAGGTA	This work
CTNS4	ATACTTAtgagtgaAAAGTAT	(46)
DMD2	TTGCTTAatggagaAAAGGTA	This work
DMD3	GTGCTTTaaaaagaAAAGGGG	This work
CF10	CTACTTTtaaaaacAAAGTCT	This work
CF12	ACGCTTTccccttcAAAGGTG	This work

Overlap sequence is presented in lowercase letters. Bases that different from the wild type are in red. N refers to any base.

Int (oInt), are active in human cells. The activity of these Int does not require any of the prokaryotic accessory proteins that are essential in *E. coli* (41–43). More importantly, it has also been shown that, along with the Flp site-specific recombinase, HK022 Int can catalyze an RMCE reaction in mammalian cells (44).

In previous work, we have identified active *attB*-like sites in the human genome that flank deleterious disease-causing mutations. These active *attB*-like sites with different overlaps and reduced B and B' sites were designated as *'attB'*s. Each *att* site has 7 bps in the O region that can be arbitrary if they are identical between the recombining pair. Additionally, permissible modifications in the previously assigned 7 bp in each *B* and *B*' Int binding sites (40) have led to their reduction to 5 bp requirement (Table 1) (45). These charac-

teristics provide flexibility in the RS's and increase the possibility to identify such *'attB'* sites in the human genome. The *'attB'* sites are key to the implementation of an Int-catalyzed RMCE reaction in human cells that can cure such mutations. (45). However, neither *w.t.* Int nor oInt exhibit sufficient RMCE activity on the mammalian genome (44).

The aim of this study is to identify optimized Int recombinase with enhanced activity that can perform successful RMCE reaction within the human genome. The present study describes the construction and examination of several Int variants to perform RMCE reactions using native *'attB'* sites pairs. First, HK022 Int was optimized in the context of transient *in trans* recombination in human cells. Then, a proof of concept successful genomic RMCE was achieved using a pair of native '*attB'* sites that have been inserted as a substrate into the human genome. Finally, *'attB'* sites in the human *CTNS* and *DMD* genes were targeted to demonstrate successful RMCE in their native location. Our results show that Int-based RMCE can become a safety, non-viral and sequence-specific gene therapy tool that addresses gene manipulations with large DNA fragments.

MATERIALS AND METHODS

Cells, growth conditions

The bacterial hosts used were *E. coli* K12 strains: TAP114 (*lacZ*) deltaM15 (46) and S17–1 lambda *pir* (47). *Escherichia coli* cells were grown and plated on Luria-Bertani rich medium with the appropriate antibiotics. Plasmid transformations were performed by electroporation (48).

Human embryonic kidney cells HEK293 (ATCC® CRL1573™), HEK293T (ATCC® CRL-3216™), and Flp-In™-293 (Invitrogen) were cultured in Dulbecco's modified Eagle's medium (DMEM). For transient transfections HEK293T cells ($\sim 6 \times 10^5$) were plated in a six-well plate and 24 h later treated with 3 μg of the proper plasmid DNA using PureFection Transfection Reagent (System Biosciences, Mountain View, CA, USA). For the human native '*attB'* sites chromosomal assay transfection, Flp-In™-293 cells ($\sim 6 \times 10^5$) were plated in a six-well plate and 24 h later treated with 5.5 μg of the proper plasmid DNA using Mirus Transfection Reagent (Mirus, WI, USA). For the *CTNS* and *DMD* chromosomal assay transfection, HEK293 cells ($\sim 6 \times 10^5$) were plated in a six-well plate and 24 h later treated with 3 μg of the proper plasmid DNA using PureFection Transfection Reagent.

Plasmid construction

Plasmids and primers used in the work are presented in Supplementary Tables S1 and S2. The details of the cloning procedure are presented in the Supplementary Materials and Methods section. In short, the two plasmids that were used as substrates in the transient human recombination assay were constructed as described (45).

pNA1756 and pNA1757 used in the off-target assay were constructed as described (45). pNG1924 (*HEXA3*) and pNG1926 (*ATM2*) *'attP'* plasmids used in the off-target assay, were constructed by RF cloning (49). The *'attP'* PCR fragment was obtained using the primers 944 and 945 on pNA1285 and pNA1608 plasmids respectively as templates (Supplementary Tables S1 and S2), then cloned into the pSSK10 vector. These plasmids were propagated in S17–1 lambda *pir* as host.

The *w.t.* Int-expressing plasmid pMK22 was constructed by cloning NcoI-Int-HindIII PCR fragment obtained using primers 1074+173 on pNR69 (40) plasmid as a template into the same sites in pKK322–2. E174K Int-expressing plasmid for *E. coli* was constructed based on pKH70 (50) as a template using two-step PCRs procedure: In the first step, two PCR reactions were performed: the first was with primers 513+144 and the second was with 143+203. In the second step, the two PCRs from the first step were assembled by PCR with primers 513+203, producing the desired E174K Int mutated fragment. Next, this fragment cut with NdeI and HindIII was ligated into the same sites in pKH70.

All plasmid constructs were verified by DNA sequencing.

Off-target integration assays in *E. coli*

Cells of *E. coli* strain TAP114 that carried an Ap^R-plasmid expressing *w.t.* Int or the E174K mutant (pMK22 or pMK174, respectively) were transformed with the relevant *attP* plasmid constructed on the basis of pSSK10 (Km^R) (*w.t. attP* or human *'attP'*s: *HEXA*3 and *7* or *ATM*2 and *4*). pSSK10 plasmid cannot replicate in the test strain, so Km^R expression occurs only in integration products. Transformed cells were plated on LB rich medium supplemented with Km and Ap. To avoid the interference of possible false-positive colonies, the presence of the Km^R gene in the Ap+Km resistant colonies was tested by PCR analysis using primers 1069+1070 (Supplementary Figure S8A, Step 1). In three independent experiments, the average growth of the cells transformed with the plasmid that carried the *w.t. attP* was between 35–60 colonies. When the presence of the Km^R gene in 100 resistant colonies was examined by PCR analysis using primers 1069+1070, only 40 colonies yielded the positive expected PCR product but none in the absence of Int (Supplementary Figure S8C and D). The remaining 60 Ap^R+Km^R colonies were PCR negative, so they were considered false-positive colonies. The 40 Km^R positive PCR clones were used for the Int-catalyzed integration activity analysis (Supplementary Figure S8A, Step 2). Site-specific integration of the *w.t. attP* Km^R plasmid into the native *attB* was confirmed by colony PCR analysis using primers 958+1080 (for *attL*) and 788+1069 (for *attR*) followed by sequencing. Plasmids that carried different *HEXA* and *ATM* *'attP'* sites yielded 5–40 Ap^R+Km^R colonies in the repeated independent experiments regardless of the Int plasmid presence. Km^R gene PCRs of 30 such colonies (Supplementary Figure S8B, C and D) with the same primers used for *w.t. attP* plasmid was all negative indicating the false-positive phenotype of these colonies.

Construction of Int mutants

All Int single mutants were constructed based on the oInt expression plasmid pNA979 as a template using the two-step PCRs procedure. In the first step, two PCR reactions were needed: the first PCR was performed using 204 as a forward primer + the reverse primer that contained the desired mutation. The second PCR was performed using a for-

ward primer that contained the desired mutation (complementary to the mutated reverse primer of the first reaction) + 469 as a reverse primer. In the second step, the two PCRs from the first step were assembled by PCR using primers 204+469, producing the desired Int variant fragment. Next, this fragment cut with EcoRI and HindIII was ligated into the same sites in the pCDNA3 vector. Int double mutants were constructed by using pNG1862 as a template, by the same two-step procedure described above. The triple mutant was constructed based on pAE2029 as a template, by the same two-step procedure described above.

Fluorescent-activated cell sorting (FACS) analysis

$\sim 2 \times 10^6$ cells from one well of a six-well plate were collected following trypsin treatment of which 10^4 cells were selected by the FACS sorter (Becton Dickinson Instrument) for fluorescent measurements. Dot plot of side scatter (SSC) versus forward scatter (FSC) was used to gate live cells in order to separate them from aggregated and dead cells. For gated cells, the fluorescence statistics was performed using the Kaluza Analysis Software (Beckman Coulter) and represented as a quadrant analysis that clearly showed the percentage of GFP expressing/ GFP-mCherry co-expressing cells.

DNA and RNA manipulations

Plasmid DNA was prepared from *E. coli* using a DNA Spin Plasmid DNA purification Kit (Intron Biotechnology, Korea) or a NucleoBond™ Xtra Midi kit (Macherey-Nagel, Germany). Genomic DNA was prepared using the Genomic DNA Mini Kit (Geneaid Biotech Ltd, Taiwan). cDNA was synthesized using the Verso cDNA kit (Applied Biosystems, Foster City, CA). mRNA was prepared using Blood/Cell Total RNA Mini Kit (Geneaid Biotech Ltd, Taiwan). Gibson reaction was performed using the NEBuilder HiFi DNA assembly master mix (NEB, MA, USA). General genetic engineering experiments were performed as described by Sambrook and Russell (48).

Identification of *'attB'* sites in the human genome

A computer search for potentially active human *'attB'* sites was performed as described (45). Genomic coordinates of the relevant *'attB'* sites were obtained from the UCSC Genome Browser. The *'attB'* distribution in the human genome was performed by NCBI Nucleotide Blast browser using human (taxid: 9606) RefSeq Genome databases.

Statistical analysis

Data were presented as the mean ± SD of at least three independent experiments, each with three repeats. *P*-values of <0.05 were considered statistically significant by Student's two-tailed *t*-test assuming equal variance.

RESULTS

Optimization of HK022 Integrase activity in human cells

The sequence of HK022 Int shares 70% identity with that of lambda's Int and both consist of three domains (Figure 2A). These domains facilitate assembly and function of a higher order tetrameric complex with the DNA *attP* and *attB* substrates known as the intasome (51,52). The N-terminal DNA binding domain (AB) (residues 1–63) recognizes 'arm-type' DNA sequences that flank the *attP* core-site. The activity of lambda integrase within the tetramer is tightly regulated so that only two diagonally positioned protomers are active at any given time. For ordered DNA strand cleavages and exchanges to form and then resolve a Holliday junction (HJ), the protomers need to switch roles during the recombination reaction, which accompanies isomerization of the HJ intermediate. It has been proposed that binding of the AB domain to the arm-type DNA, facilitates this isomerization process. The core binding domain (CB) recognizes the C and C' core binding sites of *attP* and those of *attB* (B and B'). The catalytic domain (CAT) is responsible for DNA cleavage and rejoining in the site-specific recombination reaction (51,53). To further optimize Int activity in human cells, 10 different single mutated Ints were constructed from oInt (Figure 2A): I43F (in the AB), E174K (CB) and E264G, R319G, D336V (CAT). These variants were selected based on previously described mutations used for lambda and HK022 Ints (46,54–57). We noted that two of the known lambda mutations that improved Int activity (E174K and E218K) carry Glu to Lys changes (56,58). Therefore, we tested additional replacements of acidic residues (Glu/Asp) potentially close enough to make DNA backbone contacts that could be changed to Lys. Thus, the mutants E134K, D149K (CB), and D215K, D278K, E309K (CAT) were constructed.

To examine the activity of these Int variants, we performed an assay in human HEK293T cells of a transient *trans* integrative recombination reaction. The assay utilized two plasmids as recombination substrates (42) (Figure 2B), the first substrate (pGH790) carried the *attB* site downstream to the *CMV* promoter. The second plasmid (p1331) carried *attP* located upstream to the open reading frame (ORF) of the green fluorescent protein (*GFP*) and downstream to transcription terminator (Stop). A successful *attB* × *attP* reaction forms a dimer plasmid encoding *CMV*-promoted *GFP* expression (Figure 2B). HEK293T cells were co-transfected with these two substrate plasmids, with or without an Int-expressing plasmid (oInt or one of its Int mutant derivatives, Supplementary Table S1d). 48 hours post-transfection GFP expression was analyzed by fluorescence-activated cell sorting (FACS). The quantified FACS data (Figure 2C and Supplementary Figure S1) showed that only two single Int mutants (E174K and D278K) demonstrated an increased activity, of 7% and 6.5% GFP expression, respectively, compared to 5.5% by oInt. All other 8 single mutants demonstrated similar or lower activity (between 5.5% and 0.2%) compared to 5.5% by oInt (Supplementary Figure S2). Next, we combined each of the two more efficient mutations (E174K and D278K) with some of the moderately active ones and assayed their activity. The double mutants E174K+I43F, E174K+R319G, and E174K+D278K showed elevated activity of 7.5%, 8% and 9.5% of GFP expression, respectively, compared to oInt (Figure 2C and Supplementary Figure S1). However, E174K+E264G and E174K+D336V showed a significantly lower activity (Supplementary Figure S2). Lastly, a constructed E174K+I43F+R319G triple mutant

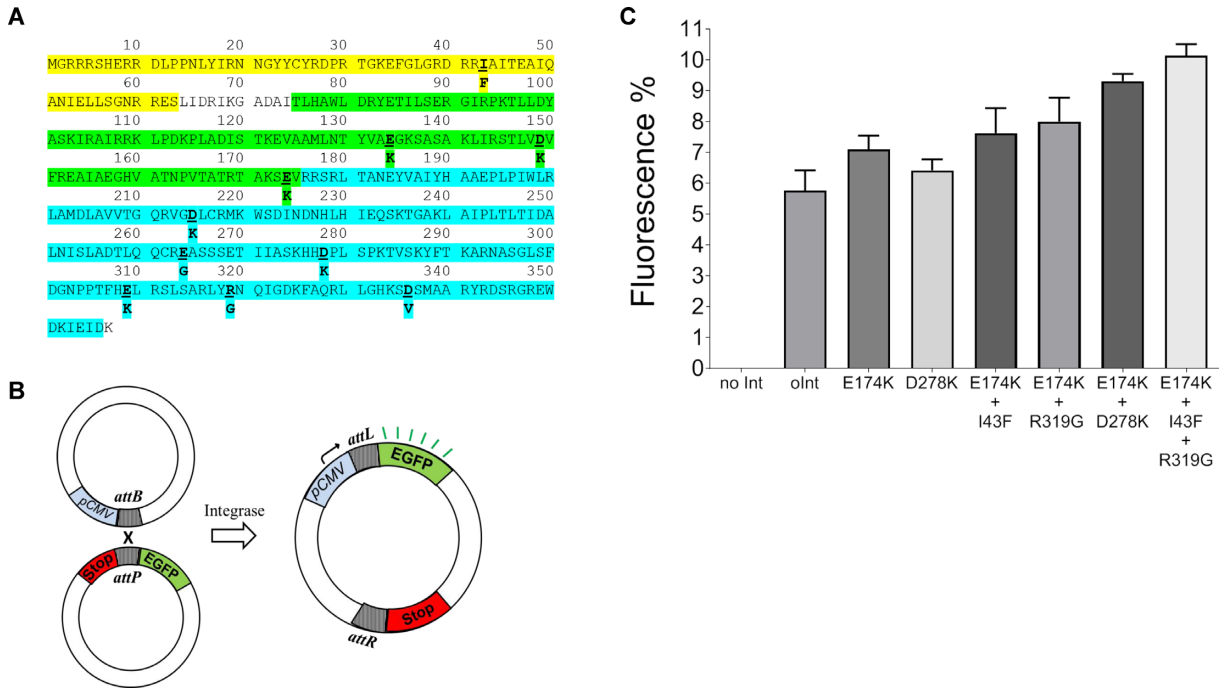

Figure 2. (**A**) HK022 Int protein sequence. Bold amino acids present mutated positions under the native amino acids. The N-terminal DNA binding domain (AB) (residues 1–63) are highlighted in yellow, core-binding (CB) domain (residues 75–175) and C-terminal catalytic domain (CAT) (residues 176–356) are highlighted in green and blue, respectively. (**B**) Scheme of transient *attB* x *attP* (*w.t.)* integration reaction using a promoter-*EGFP* trap assay. Stop - transcription terminator. (**C**) FACS data comparing GFP expression in transient *attB x attP* (w.t.) integration reaction using oInt and its variants. The bars show the mean values of three independent experiments each with three repeats; the error bars indicate standard deviation. Fluorescence % – GFP positive cells (%).

showed increased activity of 10% (Figure 2C and Supplementary Figure S1).

The various Int variants were examined on 8 different active *'attB'* sites with their compatible *'attP*'s (Table 1, Figure 3A, B and Supplementary Figure S3), using the same recombination assay described above (Figure 2B). Two of the *'attB'* sites (*HEXA3* and *ATM4*, Figure 3A and Supplementary Figure S3A) were previously reported (45), while another three *'attB'* pairs flank known mutations in the genes *CTNS* (chromosome 17), *DMD* (chromosome X) and *CFTR* (chromosome 7) that respectively cause the following diseases: Cystinosis, Duchene muscular dystrophy and Cystic fibrosis (59–61). Notably, different Int-mutants showed various efficiencies with the different *attB* sites. For instance, the triple mutant Int and the double mutant (E174K+D278K) were the most efficient Ints with the *w.t. att* sites (Figure 2C) but a lower efficiency with the *CTNS4* and *DMD3 sites* (Figure 3B). Thus, examining the optimal Int variant toward the targeted *'attB'* pair may give the prospect to achieve more efficient RMCE reaction.

Int-catalyzed RMCE reaction using human native '*attB*' sites

We designed an assay to examine if genomic '*attB'* sites that flank human deleterious mutations can enable successful Int-catalyzed genomic RMCE reactions. First, we constructed a target DNA docking platform that carried the *mCherry* reporter gene in the Flp-In™-293 cells. This 'docking' RMCE substrate plasmid (Figure 4A) carries two different human '*attB's* (*HEXA3* and *ATM4*, Table 1) that are 2.7 kb apart. *attB1* represents the *HEXA3* '*attB'* located downstream to the *EF1α* promoter, and *attB2* represents the *ATM4* '*attB'* located upstream to promoter-less ORF of the *mCherry* (Figure 4A). This docking plasmid also carried an *frt-Hyg*^{R} ORF sequence to be inserted into the chromosome of the Flp-In™-293 cells by an Flp-catalyzed SSR reaction followed by hygromycin selection. In addition, an 'incoming' plasmid was constructed (Figure 4B), that carries the relevant compatible *'attP'* sites (*attP1* and *attP*2 for *HEXA3* and *ATM4*, respectively) 4.3 kb apart. *attP1* is located upstream to the promoter-less ORF of *EGFP*, and *attP2* is located downstream to the *CMV* promoter (Figure 4B). A successful Int-catalyzed RMCE reaction between these two plasmids is expected to form a recombinant product that co-expresses both green GFP and red mCherry fluorescent products (Figure 4C). The validity of these two plasmids was initially tested transiently by co-transfecting HEK293T cells with both docking and incoming plasmids with and without Int (oInt and its variants), followed by 48 hours post-transfection FACS analysis. The quantified FACS data showed a mean increase of 5.5% of GFP and mCherry co-expression in the presence of oInt and 6.4% in the presence of E174K compared to the cells without Int (Figure 4D and Supplementary Figure S4A). E174K+I43F, E174K+R319G, E174K+D278K, and E174K+I43F+R319G Ints showed lower or non-significant RMCE activity compared to oInt (data not shown). To further verify that the *GFP* and *mCherry* co-expression properly indicated the occurrence of the expected RMCE reaction, extrachromosomal DNA was extracted from the

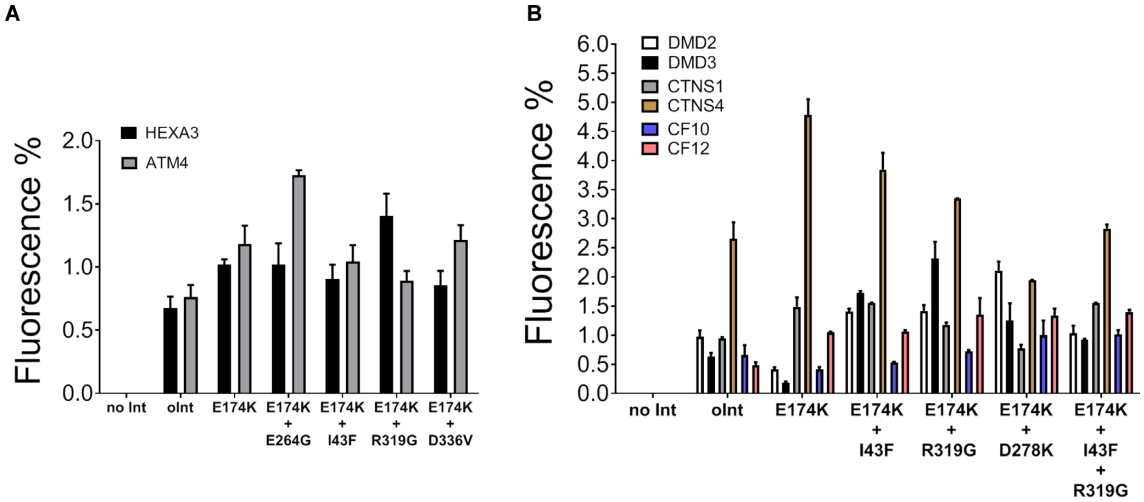

Figure 3. Transient in *trans HEXA3, ATM4, DMD2, DMD3, CTNS1, CTNS4, CF10* and *CF12 'attP'* x *'attB'* integration reaction using a promoter-*EGFP* trap assay in human HEK293T cells. Stop- transcription terminator. (**A**) FACS data comparing GFP expression in transient *attB x attP* integration reaction using oInt and its variants with *HEXA3* and *ATM4* sites. (**B**) FACS data of comparison GFP expression in transient *attB x attP* integration reaction using oInt and its variants with *DMD, CTNS* and *CF* pairs of *'attB'* sites. The bars show the mean values of three independent experiments each with three repeats; the error bars indicate standard deviation. Fluorescence % – GFP positive cells (%).

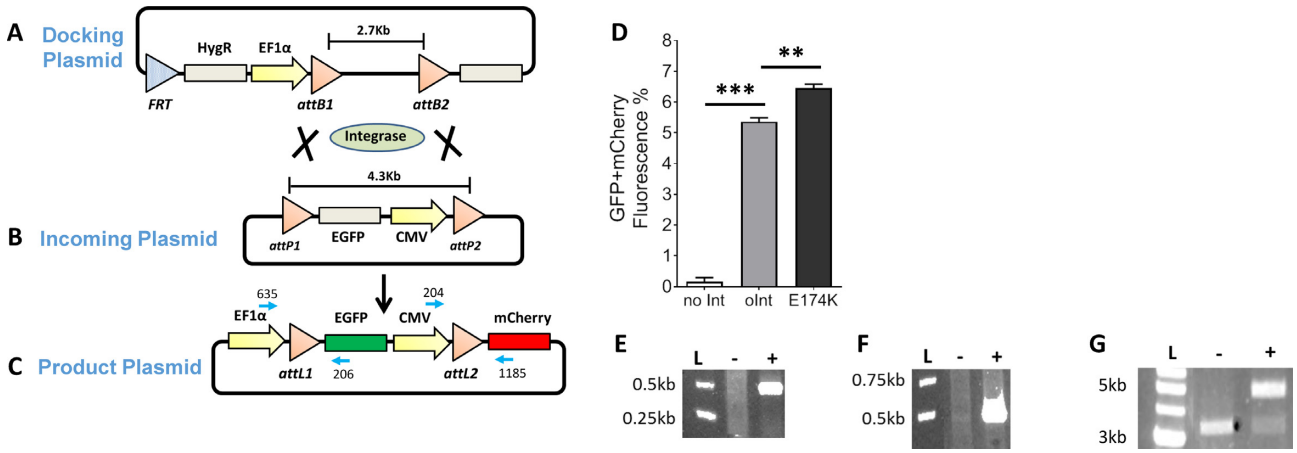

Figure 4. Int-catalyzed transient RMCE reaction in HEK293 cells. (**A**) Docking plasmid coding *EF1α* promoter–*'attB'1-'attB'2-mCherry* (ORF) sequence. (**B**) Incoming plasmid carrying *EGFP (ORF)-CMV* promoter cassette flanked by *'attP'1* and *'attP'2*. (**C**) Int-catalyzed RMCE product that co-expresses GFP and mCherry as a result of promoter trapping. (**D**) Bar graph show the FACS quantification mean values of GFP and mCherry co-expression of three independent experiments each with three repeats; the error bars indicate standard deviation. ***P-value ≤ 0.001 versus no Int; **P-value ≤ 0.01 versus oInt. (E–G) PCR analysis: (**E**) *EF1α–'attL'1-EGFP* junction (primers 635+206, 500 bp), (**F**) *CMV-'attL'2-mCherry* junction (primers 204+1185, 486 bp), (**G**) RMCE full exchanged cassette (primers 635+119, 4.6 Kb). *attB/P/L1- HEXA3 'att'* sites. *attB/P/L2-ATM4 'att'* sites. Blue arrows—primers used for PCR analysis. L-Appropriate fragments of 1 kb ladder. GFP+mCherry Fluorescence % – GFP and mCherry positive cells (%).

transfected cells, and tested by PCR (Figure 4C). The treated cells demonstrated the formation of the expected recombination junctions: *EF1α-attL1-EGFP* (500 bp) (Figure 4C and E), *CMV-attL2-mCherry* (486 bp) (Figure 4C and F), and of the complete RMCE product (4.6 kb) (Figure 4G), which were confirmed by sequencing.

These results demonstrated the validity of the two plasmids as substrates in proceeding towards a chromosomal RMCE reaction (Figure 5). Hence, the Flp-In™-293 cell line (Figure 5B) that carries a chromosomal *frt* recombination site downstream to the *SV40* promoter was co-transformed with the docking plasmid (Figure 5A) along with a Flp-expression plasmid (pOG-Flp) followed by hygromycin selection. The correct insertion of the docking plasmid in selected Hyg^R cells (Figure 5C) was confirmed by sequencing of the PCR product (415 bp, Figure 5C and G). Next, these dual *'attB'*-docked cells were co-transfected with the dual *'attP'* incoming plasmid (Figure 5D), with and without different Int-expressing plasmids. An Int-catalyzed chromosomal RMCE product (Figure 5E) is expected to demonstrate both promoter-trapped GFP and mCherry fluorescence products resulting from both recombination reactions. The FACS analysis of the Int-treated cells, 48 hours post-transfection, showed a mean of 0.45% increase in dual fluorescence by oInt activity and 0.7% by E174K compared to the untreated Int cells (Figure 5F and Supplementary Figure S4B). E174K+I43F, E174K+R319G, E174K+D278K and E174K+I43F+R319G Ints showed

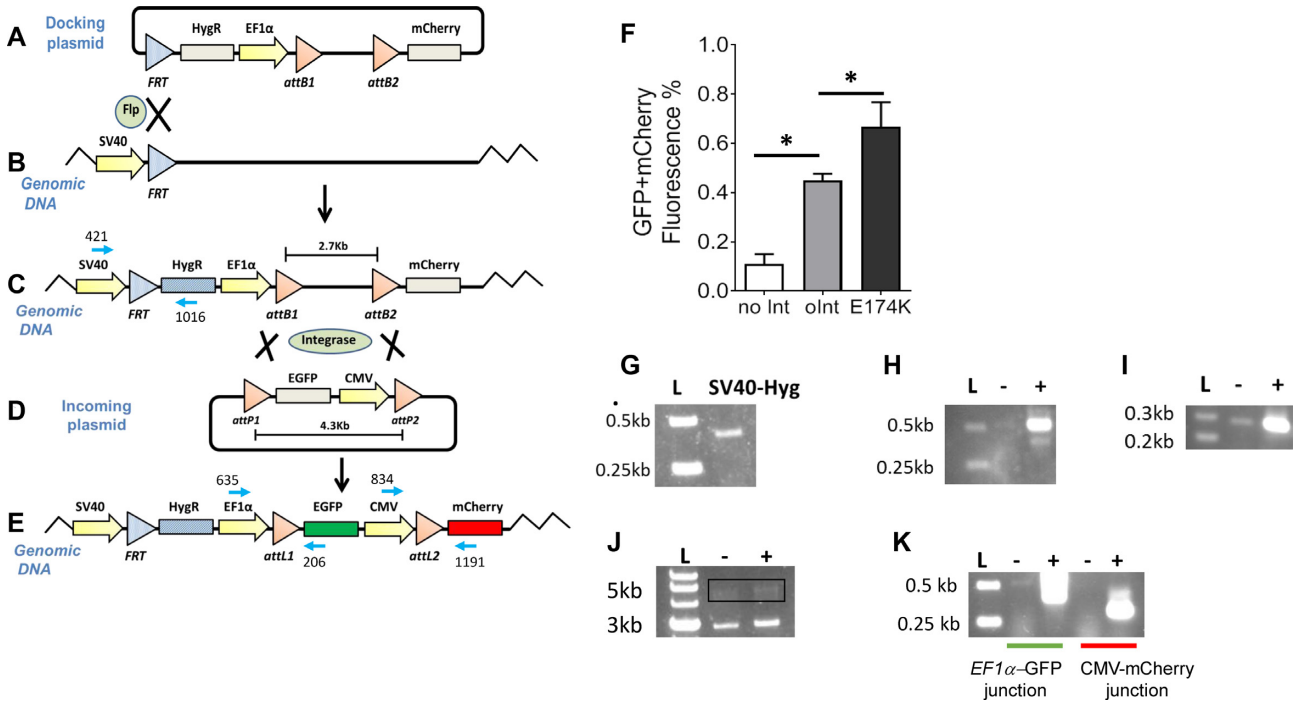

Figure 5. Int-catalyzed chromosomal RMCE reaction in the Flp-In™-293 cells. (**A**) Docking plasmid to be inserted into the chromosome, coding *frt-EF1α* promoter- *'attB'1-'attB'2 - mCherry* (ORF). (**B**) Flp-In™-293 chromosome. (**C**) Flp mediated integration product of the docking plasmid resulting Hyg^R cells. (**D**) Incoming plasmid coding *EGFP* (ORF) *-CMV* promoter flanked by *'attP' 1* and *'attP' 2* sites. (**E**) Int-catalyzed RMCE product that co-express GFP and mCherry. (**F**) Bar graph show the FACS quantification mean values of GFP and mCherry co-expression of three independent experiments each with three repeats; the error bars indicate standard deviation. *p-val ≤ 0.05 vs. no Int; *p-val ≤ 0.05 vs. oInt. (G–K) PCR analysis: (**G**) *SV40-FRT-*Hyg^R junction (primers 421+1016, 415 bp), (**H**) *EF1α–'attL'1-EGFP* junction (primers 635+206, 500 bp), (**I**) *CMV-'attL'2-mCherry* junction (primers 834+1191, 273bp). (**J**) RMCE full exchanged cassette (primers 635+1191, 4.6 kb). (**K**) Nested PCRs of *EF1α–'attL'1-EGFP* junction (primers 635+206, 500 bp) and *CMV-'attL'2-mCherry* junction (primers 834+1191, 273bp) on the purified recombinant 4.6 kb PCR as template. *attB/P/L1- HEXA3 'att'* sites. *attB/P/L2-ATM4 'att'* sites. Blue arrows – primers used for PCR analysis. L-Appropriate fragments of 1 kb or 100 bp ladders. GFP+mCherry Fluorescence % – GFP and mCherry positive cells (%).

lower or non-significant RMCE activity compared to oInt (data not shown). PCR and sequencing analyses by the appropriate primers and the chromosomal DNA of the transfected cells as a template confirmed the presence of the expected recombination junction products *EF1α-attL1-EGFP* (500 bp) (Figure 5H), and *CMV-attL2-mCherry* (273bp) (Figure 5I) only in the Int-treated cells. The PCR analysis for the expected full 4.6 kb RMCE product (Figure 5E, primers 635+1191) revealed a weak band for the expected product (Figure 5J), dominated by the shorter 3.2 kb PCR product of the non-recombined docking chromosomal cassette (Figure 5C). Therefore, this weak 4.6 kb product was gel-purified and used as a template for a nested PCR reaction that confirmed the presence of both expected recombination junctions only in the Int-treated cells (Figure 5K). The expected sequence of these PCR products was confirmed by sequencing. These results are a proof of concept that Int variants can successfully catalyze chromosomal RMCE with two different active *'attB'* sites in the absence of any selective pressure or fluorescence-based sorting.

Int-catalyzed RMCE reactions in native *'attB'* sites at the human genes *CTNS* and *DMD*

Our next aim was to demonstrate a successful Int-based RMCE reaction in native locations of the human genes. We targeted genes whose disease-causing mutations are flanked by two active native *'attB'* sites. Therefore, we examined Int-catalyzed RMCE reactions in the *CTNS* and *DMD* genes using their appropriate flanking *'attB'* sites in HEK293 cells by GFP trap assays. In the *CTNS* case, we chose the *CTNS1* and *CTNS4 'attB'* sites (Table 1) which are 7.6 kb apart and flank a region containing the *CTNS* promoter and *exons 1* to *3* (Figure 6B). Deleterious mutations in this region are associated with Cystinosis disease (59). The compatible constructed incoming plasmid (Figure 6A) carried a *CMV*-promoted *EGFP* ORF (lacking polyA) followed by a P2A ribosomal skipping sequence and the sequence of the *CTNS exon 3* splice donor (for RNA splicing), all flanked by the relevant *'attP'*s (*CTNS1* and *CTNS4*), 1.7 kb apart. Successful Int-catalyzed RMCE is expected to replace the genomic sequence between the two *CTNS 'attB'*s with the incoming cassette between its two *'attP'*s (Figure 6C). Thus, the RMCE genomic recombinant is expected to transcribe an mRNA of the *EGFP-P2A-Exons 4–12* sequence (Figure 6D). The P2A ribosomal skipping site will lead to the translation of two peptides, GFP and a distal peptide of CTNS. HEK293 cells were co-transfected with this incoming plasmid along with a plasmid expressing one of the Int variants (oInt, E174K, E174K+I43F, E174K+R319G, E174K+D278K and E174K+I43F+R319G). FACS analysis of transformed cells have shown that the E174K+I43F Int variant gave the highest RMCE efficiency with a mean 0.6% GFP fluorescence compared to 0.45% by oInt and the untransformed cells (Figure 6E and Supplementary Fig-

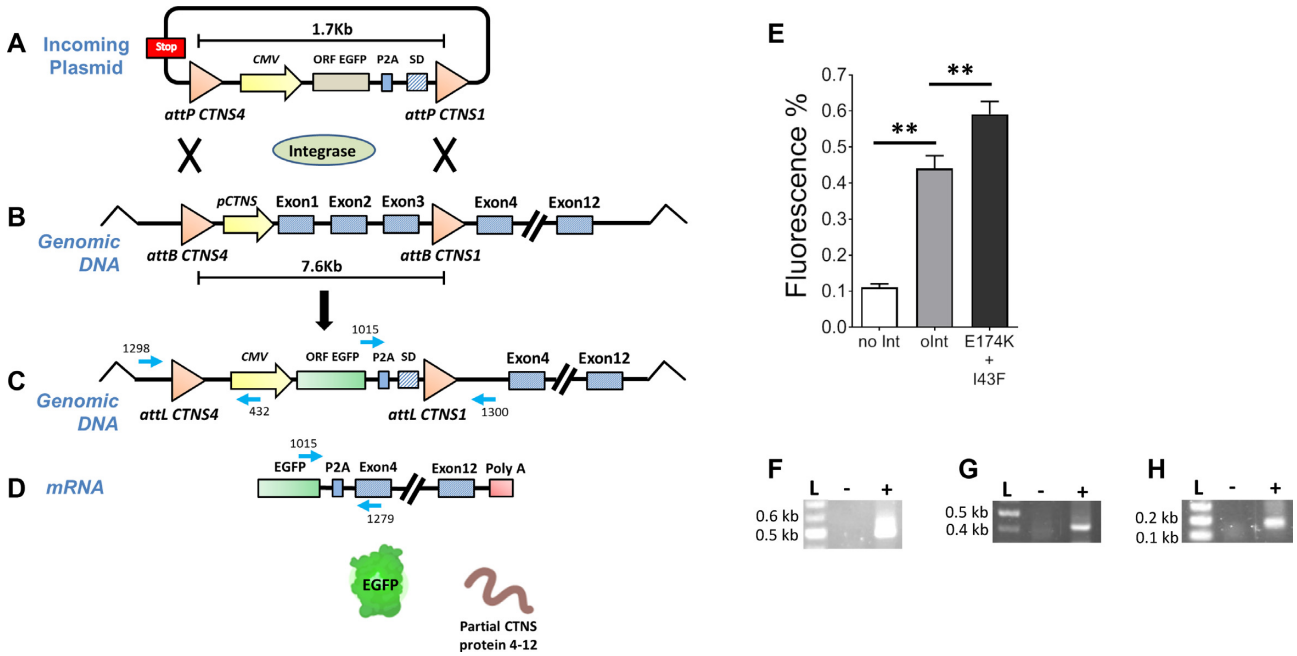

Figure 6. Scheme of Int-catalyzed RMCE on *'attB'*s in the *CTNS* gene. (**A**) Incoming plasmid coding *CMV-EGFP* (ORF- lacking poly A)- P2A-SD all flanked by *'attP' CTNS4* and *'attP' CTNS1* sites. (**B**) Genomic *CTNS* locus with *'attB' CTNS4* and *'attB' CTNS1* sites that flanks the *CTNS* promoter-*Exons1–3* cassette. (**C**) The RMCE reaction product. (**D**) mRNA of the RMCE expressing *EGFP*-P2A upstream to *exons 4–12*. (**E**) Bar graph show the FACS quantification mean values of GFP expression of three independent experiments each with three repeats; the error bars indicate standard deviation. **p-val ≤ 0.01 vs. no Int; **p-val ≤ 0.01 vs. oInt. (F–H) PCR analyses: (**F**) Proximal *CTNS* locus-*'attL'CTNS4-CMV* junction (primers 1298+432, 500 bp). (**G**) *EGFP*-P2A-SD-*'attL'CTNS1-Intron3* junction (primers 1015+1300, 400 bp). (**H**) *EGFP*-P2A-*Exon4* mRNA junction (primers 1015+1279, 177 bp). SD- Splicing donor. P2A- 2a peptide ribosome skipping. Stop- transcription terminator. L – Appropriate fragments of 100 bp ladder. Fluorescence %- GFP positive cells (%).

ure S5A). E174K, E174K+R319G, E174K+D278K and E174K+I43F+R319G Ints showed lower or non-significant RMCE activities compared to oInt (data not shown). Chromosomal DNA and mRNA extracted from the transfected cells were used as templates for PCR reactions with the proper primers (Figure 6C and D). PCR results demonstrated the formation of the expected recombinant junctions *CTNS-'attL'CTNS4-CMV* of 500 bp (Figure 6C and F), *EGFP-2A-SD-'attL'CTNS1-Intron3* of 400 bp (Figure 6C and G), as well as the mRNA expected *EGFP-P2A-Exon4* junction of 177 bp (Figure 6D and H). The correct sequence of all PCR products was confirmed by next-generation sequencing (NGS).

A second RMCE reaction targeted the *DMD* gene, using *DMD2* and *DMD3 'attB'* sites (Table 1) that are 23 kb apart located in *introns 43* and *44* respectively, and flank *exon 44* (Figure 7B). Deleterious mutations in this region are associated with Duchene muscular dystrophy (60). Here we used a GFP promoter trap whose incoming plasmid (Figure 7A) carried a splicing acceptor (SA), a ribosomal skipping site (P2A), and the ORF of *EGFP* with a polyA sequence, all flanked by the two relevant *'attP'*s (*DMD2* and *DMD3*), 1.4 kb apart. Like *CTNS*, a processed mRNA is expected to be translated to an active GFP and a proximal DMD peptide. FACS analysis of transformed cells have shown that the E174K+I43F+R319G Int variant gave the highest RMCE efficiency with a mean 0.55% GFP fluorescence compared to 0.3% by oInt and the untransformed cells (Figure 7E and Supplementary Figure S5B). E174K, E174K+I43F, E174K+R319G and E174K+D278K Ints showed lower or non-significant RMCE activities compared to oInt (data not shown). Chromosomal DNA and mRNA extracted from the transfected cells served as templates for PCR reactions with the proper primers. PCR results confirmed and demonstrated the expected recombinant junctions: *Intron43-'attL' DMD2-SA-P2A-EGFP* (700 bp) (Figure 7C and F), *EGFP-'attR' DMD3-Intron44* (800 bp) (Figure 7C and G) and the mRNA *Exon43-P2A-EGFP* junction (229 bp) (Figure 7D and H). The correct sequence of all PCR products was confirmed by NGS. RMCE reactions were also demonstrated in *exons 45* and *52* in the *DMD* gene (Supplementary Figures S6 and S7). The used *DMD4* and *DMD5 'attB'* sites (Table 1) are 41 kb apart located in *introns 44* and *45* respectively, flank *exon 45* (Supplementary Figure S6B). *DMD6* and *DMD7* sites (Table 1) that are 58 kb apart located in *introns 51* and *52* respectively, flank *exon 52* (Supplementary Figure S7B). Preliminary FACS results demonstrated the occurrence of oInt-catalyzed RMCE in both locations with a 0.35% and 0.2% of GFP mean fluorescence in *exons 45* and *52* respectively (Supplementary Figures S6E–F and S7E–F). Other Int variants showed lower or non-significant RMCE activities compared to oInt (data not shown).

DISCUSSION

Despite the progress achieved in the last two decades in the field of gene therapy, human genome-editing tools still

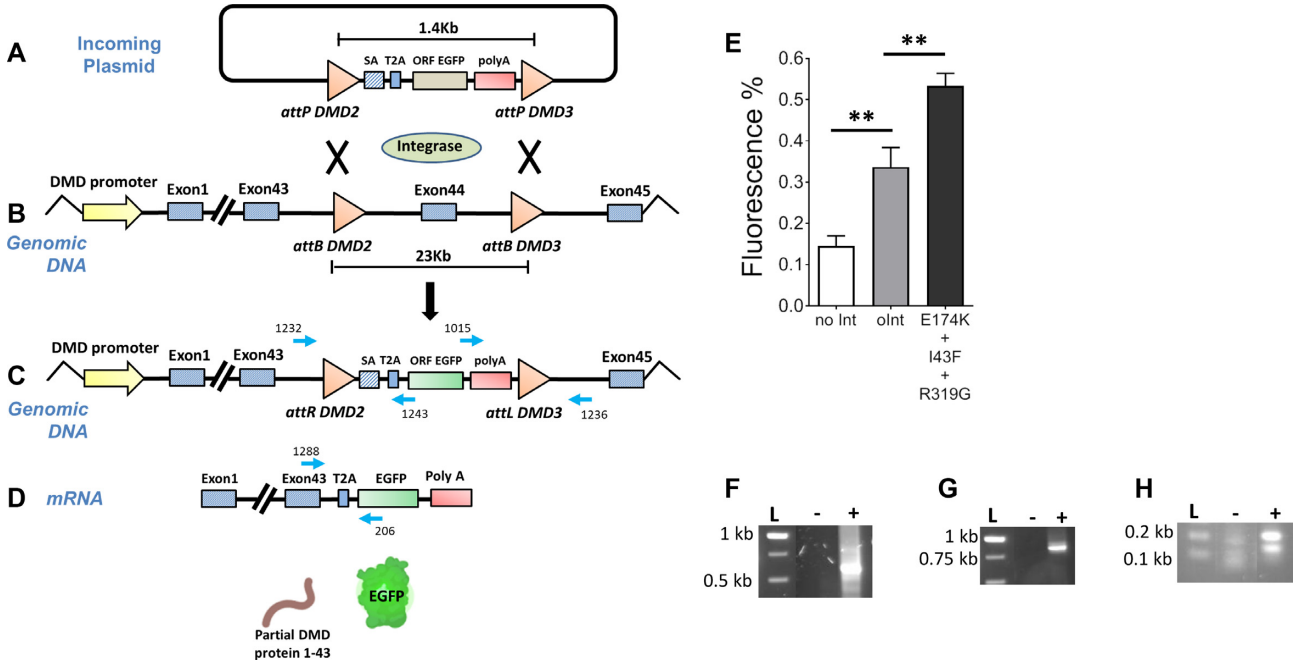

Figure 7. Scheme of Int-catalyzed RMCE in the *DMD* gene on exon44 flanking *'attB'*s. (**A**) Incoming plasmid coding promoter-less *EGFP*-ORF-SA-P2A-Poly A all flanked by *'attP' DMD2* and *'attP' DMD3* sites. (**B**) Genomic *DMD* locus with *'attB' DMD2* and *'attB' DMD3* sites in *introns 43* and *44* respectively and flanking *exon 44*. (**C**) The RMCE reaction product. (**D**) mRNA of the RMCE expressing proximal *exons 1–43* and *EGFP*. (**E**) Bar graph show the FACS quantification mean values of GFP expression of three independent experiments each with three repeats; the error bars indicate standard deviation. **p-val ≤ 0.01 vs. no Int; **p-val ≤ 0.01 vs. oInt. (F–H) PCR analyses: (**F**) *Intron43-'attL'DMD2*-SA-P2A-*EGFP* junction (primers 1232+1243, 700 bp). (**G**) *EGFP-'attR'DMD3-Intron44* junction (primers 1015+1236, 800 bp). (**H**) *Exon43*-P2A-*EGFP* mRNA junction (primers 1288+206, 229 bp). SA, splicing acceptor. P2A, 2a peptide ribosome skipping. Stop, transcription terminator. L-Appropriate fragments of 1 kb or 100 bp ladders. Fluorescence % – GFP positive cells (%).

need to improve in efficiency, safety, sequence-specificity, and large DNA fragments targeting. In this study, we aimed to demonstrate a novel potential gene therapy tool based on the Int-catalyzed RMCE. Previously we identified pairs of human native active *'attB'* sites that can be used to correct *'attB'*-flanked mutations by the Int-catalyzed RMCE reactions. The *'attB'* sites have a short 17 bp recombining site and a promiscuous 7 bp overlap sequence. In contrast to the *'attB'* sites, the RSs of commonly used tyrosine recombinases like Cre and Flp were not identified in the human genome. Consequently, additional strategies were utilized to carry out a genome editing, either by constructing a variety of recombinases with altered target specificities (62,63), or by the insertion of their specific recombination sites into the targeted DNA sequence (27). HK022 Int is active in human cells without the need to supply the prokaryotic IHF accessory protein unlike lambda's *w.t.* Int (42,43). These features of HK022 Int offer a simpler means for correcting human mutations by RMCE.

We hypothesize that the mechanism of the Int-catalyzed RMCE reaction is sequential. The reaction begins with a single *trans* integration reaction that takes place between one *'attP'* of the incoming plasmid and its compatible chromosomal *'attB'*. This step creates an integrated intermediate of the incoming plasmid. Thereafter, a *cis* reaction occurs between the second *'attP'* × *'attB'* pair resulting in the fully-integrated intermediate to the final RMCE product (44,64). According to our hypothesis, we predict that the efficiency of the first *trans* integration reaction is a major limiting step towards the formation of the RMCE product. Therefore, an optimal Int variant for RMCE must either improve the first step of the integration *in trans* reaction activity, and/or reduce the *cis* reverse excision activity of the first recombination intermediate. Hence, to identify a more efficient Int variant we examined the constructed Int variants in a single *'attP'* × *'attB'* *trans* reaction (Figure 2). The integration analysis revealed that only two Int variants (E174K and D278K) increased the integration activities compared to oInt (Figure 2C).

The construction of HK022 Int variants was based on lambda's Int structural information. The E174K mutant of lambda Int (also known as Int-h) proved to be IHF-independent in *E. coli*, and more active *in trans* site-specific recombination reactions (56). Furthermore, it showed an enhanced DNA binding activity (65) that supports its enhanced activity *in trans*. This mutation has also been used to allow lambda Int to be active in human cells (66–69). The E174K mutation in HK022 Int is in the inter-domain linker (I160–R176), which corresponds with the location of the mutation in lambda Int. It is likely that this CB-CAT linker is highly flexible when Int is not bound to DNA, which would affect the entropic cost of DNA binding and thereby DNA binding affinity. Thus, we predict that lysine residue substitution might enhance the DNA binding affinity by stabilizing interaction with the DNA and/or by constraining the movement of the inter-domain linker (70). We

suppose that E174K and D278K, which replaced the negatively charged Glu/Asp by the positively charged lysine near the DNA, enhances Int activity, most likely, by introducing new ionic interactions with the DNA backbone (Figure 8). We could have expected the same for E309K as it is also near the DNA backbone. Yet, E309 is close to the active site and is hydrogen-bonded to R179, an important residue for positioning Tyr342 in the active site. This might explain why the E309K mutation significantly compromises Int activity.

The constructed Integrase variants based on E174K Int showed higher recombination activities with the different *'attB'* sites compared to oInt (E174K+I43F, E174K+E264G, E174K+R319G, E174K+D278K, E174K+I43F+D336V) (Figure 3A, B and Supplementary Figure S3). Based on the crystal structure (Figure 8), I43 is far away from the arm-site DNA and faces the adjacent N-terminal domain within the Int tetramer (Figure 8B). It was suggested earlier that the I43F mutation in lambda Int affects the allosteric rearrangements that drive resolution of the intermediate Holliday junction towards recombined products (55). The R319G mutation is located on the CAT surface (Figure 8E) which plays a key role in the catalytic activity and the regulation of site-specific recombination. Notably, it has been previously described that mutations in lambda Int E319 reduce substrate specificity and increase its activity toward HK022 sites (46). In the present work, the single I43F and R319G mutations did not increase transient trans integrative recombination activity (Supplementary Figure S2) but they did enhance E174K activity in a double mutant context (Figure 2C and Supplementary Figure S1). Moreover, the I43F and R319G mutations contributed mostly toward improving *in* trans integrative recombination in the triple mutant context (Figure 2C and Supplementary Figure S1). However, in the context of transient RMCE (Figure 4D), and in stable genomic RMCE in the human cell-line Flp-In™-293 (Figure 5F), the most active Int variant was E174K showing 6.4% and 0.7% RMCE activity, respectively (Supplementary Figure S4). Although, in the *CTNS* and *DMD* genes the most efficient mutants were E174K+I43F and E174K+I43F+R319G reaching 0.6% and 0.55% RMCE activity, respectively (Figures 6E, 7E, and Supplementary Figure S5A-B). The differences in stable RMCE efficiency between the three cases could be due to the distance between their targeted *'attB'* pairs (2.7 Kb, 7.6 Kb and 23 Kb, respectively), the *'attB'* pairs activity, the different Int variants efficiency, and/or the genomic location accessibility.

The RMCE reactions within the *CTNS* and *DMD* genes were initially performed in their native sequence contexts (in HEK293 strain). Hence, probable reduction in cell viability compared to the unrecombined cells may have originated from replacing the native sequence with the *EGFP* reporter in either gene. Although, the FACS data of the Int-treated cells showed elevations of GFP expression by 0.2–0.7% compared to the untreated control, and the occurrences of the expected RMCE reactions were confirmed by PCRs and sequencing. These achieved preliminary efficacies are encouraging in the light of the already successfully used gene therapy approach based on the Adeno-associated virus system that have similar efficacies (between 0.2% and 0.5%) (71,72). To further study and increase the efficiency of the HK022 Int recombination reaction, we plan to perform sequence consensus analysis (73–75) with structure-based examination of potential scaffold-optimizing mutations (76–78). This will help identify and manipulate critical residues that are important for the *att* sites' recognition and binding. Additionally, the correction of targeted mutated genes by Int-based RMCE approach in tissue cultures (*ex vivo*) and mice models (*in vivo*) will be performed using systemic treatment (72,79). Such an approach of mutated genes correction with the appropriate sequence may favor cured cells and even show increased efficiencies.

Off-target events interfere with successful gene targeting attempts and present a safety concern. Therefore, we performed a blast search for possible identical copies of the identified minimal 17 bp active *'attB'* sequences in the human genome. The analysis of human *'attB'* sites showed a varying number of repetitions within the human genome, ranging from a single copy to several copies. *DMD2, DMD4, CTNS1* and *CTNS4* have a single copy, on the other hand, *CF12*, *DMD7, HEXA3, DMD5*, *DMD3*, *DMD6, CF10* and *ATM4* all have additional copies (Supplementary Table S3). However, none of these additional copies of the paired sites are on the same chromosome, except for the single site of *CF12* and one of the *CF10* additional sites that are located on chromosome No.18. The distance between the two sites is 18,000 kb, deeming it irrelevant for a productive RMCE reaction. Despite the presence of additional sites, we successfully detected productive RMCE reactions even for the case that uses the *ATM4* '*attB*' site out of 41 additional sites. Even though single-site integration is less likely to occur due to its kinetically favored reverse excision reaction (27), for future gene therapy applications, the selection of a pair of *'attB'* sites is essential to avoid potential off-target effects. Therefore, we suggest choosing a pair of '*attB*' sites, each having a single copy in the human genome, as in the case of *CTNS*, or ones that do not have additional sites on the same non-targeted chromosome.

In our previous publication (45), we tested Int-catalyzed off-target integration activity of various *'attP'* sites in the *E. coli* genome by using a selection force. While all tested w.t. *attP* integrations occurred exclusively with their matching w.t. *attB* site, in 8.5% selected colonies of four different *'attP*'s (*HEXA*3 and 7, *ATM*4 and 8) integration were reported as off-targeted. However, we later realized that this assay did not take into consideration false-positive colonies. Thus, to re-examine this alleged off-target activity, we applied a more restrictive two-step assay (described in Materials and Methods and in Supplementary Figure S8). The assay confirmed a complete absence of off-target integration in the *E. coli* genome catalyzed either by *w.t.* or E174K Int using the same four *'attP'* sites. This further demonstrates the low risk of off-target recombination based on *'attB' sites.* In addition, the use of Integrase-based binary system as an anti-cancer treatment did not show any broad toxicity effects as known in alternative technologies in the lungs and spleen of the healthy mice. TUNEL assay that detects double strand breaks, did not show any DNA damage in the healthy mice tissue biopsies taken 45 days after systemic injections (twice a week) (79). These findings strengthen our claim that HK022 Int-based gene therapy is

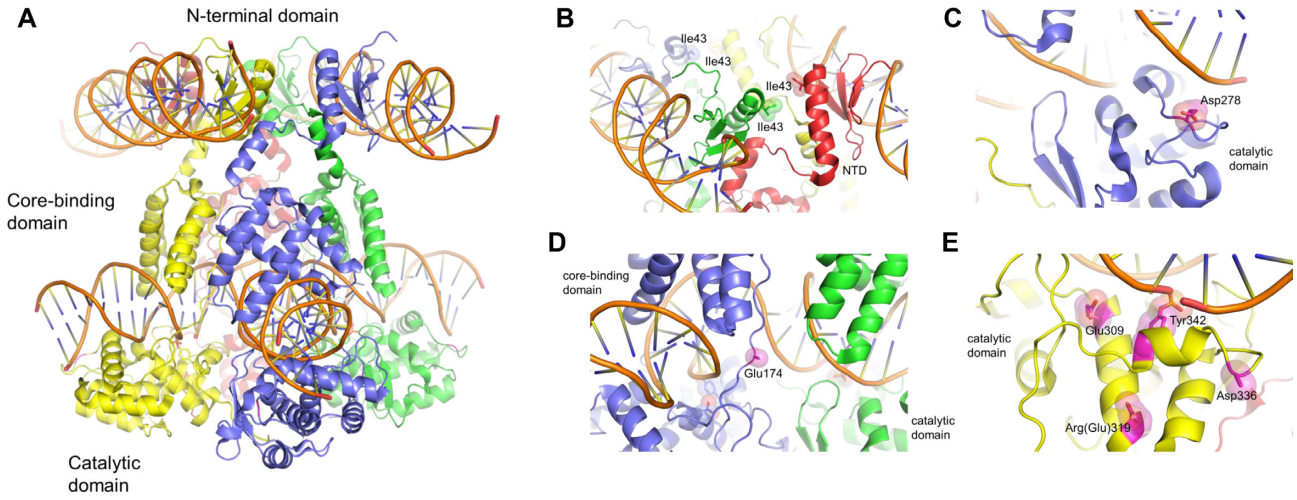

Figure 8. (**A**) Crystal structure of the lambda integrase tetramer bound to the arm and core-type DNAs (PDB ID: 1z1b). The four protein molecules are colored differently. (**B**) Interaction between the four N-terminal domains, sandwiched between two arm-type DNAs. I43 is located at the protein-protein interface. (**C**) D278 positioned close to the distal region of core-type DNA, that is, away from the center of the Holliday junction. (**D**) The position of lambda Int E174 in the linker between core-binding and catalytic domains, close to the DNA backbone. Note that E174K mutant was used in the crystallographic studies, and due to the flexibility of the Lys side chain that residues were modeled as an alanine. (**E**) Y342 covalently bonded to the cleaved DNA 3' end, and surrounding residues E309, E319 (R in HK022 Int), D336 (modeled as A due to side chain flexibility) from the Int's catalytic domain.

a safe strategy. In the future, we aim to examine off-target effects by using unbiased, high-throughput assay as the linear amplification-mediated high-throughput genome-wide translocation sequencing (80), to detect and quantify DNA double-stranded breaks (DSBs) in the context of much larger human genome in mutated tissue culture and mice models. These experiments will help provide a better understanding regarding off-target activity.

The HK022 Int-RMCE approach based on native '*attB*' sites that flank deleterious mutations along with improved variants of Int, is a promising human genome-editing and gene-therapy tool. Future studies of this approach will aim to further optimize HK022 Int-based RMCE efficiency in the human genome and demonstrate gene correction in appropriate mutated tissue cultures and in mice models using cDNA of the targeted gene. Since many human disorders arise from loss-of-function mutations, this strategy could be applied to a wide variety of diseases. Hence, presenting a universal therapeutic approach that could restore normal expression levels of any targeted gene, regardless of the specific deleterious mutation. This method holds the potential to treat diseases such as Cystinosis, Duchene muscular dystrophy and Cystic fibrosis that are caused by several mutations in the same gene (81–83). Furthermore, the Int-based RMCE addresses the insert size limitation; in the present work, we demonstrated its capability to exchange a human chromosomal fragment over 20 kb long. Moreover, this strategy allows a promoter-less stable transgene expression by relying on the endogenous promoter expression of the targeted gene as in the case of the *DMD* gene. This is preferred over the use of promiscuously integrating vectors, which are sometimes associated with clonal expansion (84) and oncogenesis (15). Here, we described a novel genome editing method that relies on the ability of HK022 Int to recombine with native RSs within the human genome. Thus, providing the basis for a promising gene therapy tool that is universal, easy to deliver, safe, stable and reliable.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank all members of the Adi Barzel laboratory for fruitful discussions.

FUNDING

Colton Foundation, Israel (to M.K.); National Institutes of Health, US [R35-GM118047 to H.A.]. Funding for open access charge: Deutsche Forschungsgemeinschaft.
Conflict of interest statement. None declared.

REFERENCES

1. Cheng,J.K. and Alper,H.S. (2014) The genome editing toolbox: a spectrum of approaches for targeted modification. *Curr. Opin. Biotechnol.*, **30C**, 87–94.
2. Anguela,X.M. and High,K.A. (2018) Entering the modern era of gene therapy. *Annu. Rev. Med.*, **70**, 273–288.
3. Kitada,T., DiAndreth,B., Teague,B. and Weiss,R. (2018) Programming gene and engineered-cell therapies with synthetic biology. *Science*, **359**, aad1067.
4. Maeder,M.L. and Gersbach,C.A. (2016) Genome-editing technologies for gene and cell therapy. *Mol. Ther.*, **24**, 430–446.
5. Cornu,T.I., Mussolino,C. and Cathomen,T. (2017) Refining strategies to translate genome editing to the clinic. *Nat. Med.*, **23**, 415–423.
6. Roy,B., Zhao,J., Yang,C., Luo,W., Xiong,T., Li,Y., Fang,X., Gao,G., Singh,C.O., Madsen,L. *et al.* (2018) CRISPR/Cascade 9-mediated genome editing-challenges and opportunities. *Front Genet.*, **9**, 240.
7. Xu,X., Wan,T., Xin,H., Li,D., Pan,H., Wu,J. and Ping,Y. (2019) Delivery of CRISPR/Cas9 for therapeutic genome editing. *J. Gene Med.*, **21**, 3107.

8. Hendrickx,R., Stichling,N., Koelen,J., Kuryk,L., Lipiec,A. and Greber,U.F. (2014) Innate immunity to adenovirus. *Hum. Gene Ther.*, **25**, 265–284.
9. Milone,M.C. and O'Doherty,U. (2018) Clinical use of lentiviral vectors. *Leukemia*, **32**, 1529–1541.
10. Wang,D., Tai,P.W.L. and Gao,G. (2019) Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discov.*, **18**, 358–378.
11. Kawakami,K., Largaespada,D.A. and Ivics,Z. (2017) Transposons as tools for functional genomics in vertebrate models. *Trends Genet*., **33**, 784–801.
12. Kebriaei,P., Izsvak,Z., Narayanavari,S.A., Singh,H. and Ivics,Z. (2017) Gene therapy with the sleeping beauty transposon system. *Trends Genet*., **33**, 852–870.
13. Navarro,C. (2017) The mobile world of transposable elements. *Trends Genet*., **33**, 771–772.
14. Nayak,S. and Herzog,R.W. (2010) Progress and prospects: immune responses to viral vectors. *Gene Ther*., **17**, 295–304.
15. Fischer,A., Hacein-Bey-Abina,S. and Cavazzana-Calvo,M. (2010) 20 years of gene therapy for SCID. *Nat. Immunol.*, **11**, 457–460.
16. Ma,Y., Zhang,L. and Huang,X. (2014) Genome modification by CRISPR/Cas9. *FEBS J.*, **281**, 5186–5193.
17. Fu,Y., Foden,J.A., Khayter,C., Maeder,M.L., Reyon,D., Joung,J.K. and Sander,J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822–826.
18. Ihry,R.J., Worringer,K.A., Salick,M.R., Frias,E., Ho,D., Theriault,K., Kommineni,S., Chen,J., Sondey,M., Ye,C. *et al.* (2018) p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.*, **24**, 939–946.
19. Haapaniemi,E., Botla,S., Persson,J., Schmierer,B. and Taipale,J. (2018) CRISPR-Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med.*, **24**, 927–930.
20. Kosicki,M., Tomberg,K. and Bradley,A. (2018) Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.*, **36**, 765–771.
21. Byrne,S.M. and Church,G.M. (2015) CRISPR-mediated gene targeting of human induced pluripotent stem cells. *Curr. Protoc. Stem Cell Biol.*, **35**, 5A–22.
22. Holkers,M., Maggio,I., Henriques,S.F., Janssen,J.M., Cathomen,T. and Goncalves,M.A. (2014) Adenoviral vector DNA for accurate genome editing with engineered nucleases. *Nat. Methods*, **11**, 1051–1057.
23. Maggio,I., Holkers,M., Liu,J., Janssen,J.M., Chen,X. and Goncalves,M.A. (2014) Adenoviral vector delivery of RNA-guided CRISPR/Cas9 nuclease complexes induces targeted mutagenesis in a diverse array of human cells. *Sci. Rep.*, **4**, 5105.
24. Li,K., Wang,G., Andersen,T., Zhou,P. and Pu,W.T. (2014) Optimization of genome engineering approaches with the CRISPR/Cas9 system. *PLoS. ONE.*, **9**, e105779.
25. Olorunniji,F.J., Merrick,C., Rosser,S.J., Smith,M.C.M., Stark,W.M. and Colloms,S.D. (2017) Multipart DNA assembly using site-specific recombinases from the large serine integrase family. *Methods Mol. Biol.*, **1642**, 303–323.
26. Bland,M.J., Ducos-Galand,M., Val,M.E. and Mazel,D. (2017) An att site-based recombination reporter system for genome engineering and synthetic DNA assembly. *BMC. Biotechnol.*, **17**, 62.
27. Meinke,G., Bohm,A., Hauber,J., Pisabarro,M.T. and Buchholz,F. (2016) Cre recombinase and other tyrosine recombinases. *Chem. Rev.*, **116**, 12785–12820.
28. Stark,W.M. (2017) Making serine integrases work for us. *Curr. Opin. Microbiol.*, **38**, 130–136.
29. Schlake,T. and Bode,J. (1994) Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry*, **33**, 12746–12751.
30. Seibler,J. and Bode,J. (1997) Double-reciprocal crossover mediated by FLP-recombinase: a concept and an assay. *Biochemistry*, **36**, 1740–1747.
31. Wallace,H.A.C., Marques-Kranc,F., Richardson,M., Luna-Crespo,F., Sharpe,J.A., Hughes,J., Wood,W.G., Higgs,D.R. and Smith,A.J.H. (2007) Manipulating the mouse genome to engineer precise functional synthenic replacements with human sequence. *Cell*, **128**, 197–209.
32. Turan,S., Zehe,C., Kuehle,J., Qiao,J. and Bode,J. (2013) Recombinase-mediated cassette exchange (RMCE) - a rapidly-expanding toolbox for targeted genomic modifications. *Gene*, **515**, 1–27.
33. Schlake,T. and Bode,J. (1994) Use of mutated flp recognition target (frt) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry*, **33**, 12746–12751.
34. Wirth,D., Gama-Norton,L., Riemer,P., Sandhu,U., Schucht,R. and Hauser,H. (2007) Road to precision: recombinase-based targeting technologies for genome engineering. *Curr. Opin. Biothecnol.*, **18**, 411–419.
35. Haenebalcke,L., Goossens,S., Naessens,M., Kruse,N., Farhang,G.M., Bartunkova,S., Haigh,K., Pieters,T., Dierickx,P., Drogat,B. *et al.* (2013) Efficient ROSA26-based conditional and/or inducible transgenesis using RMCE-compatible F1 hybrid mouse embryonic stem cells. *Stem Cell Rev. Rep.*, **9**, 774–785.
36. Zhang,L., Inniss,M.C., Han,S., Moffat,M., Jones,H., Zhang,B., Cox,W.L., Rance,J.R. and Young,R.J. (2015) Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line. *Biotechnol. Prog.*, **31**, 1645–1656.
37. Pei,Y., Sierra,G., Sivapatham,R., Swistowski,A., Rao,M.S. and Zeng,X. (2015) A platform for rapid generation of single and multiplexed reporters in human iPSC lines. *Sci. Rep.*, **5**, 9205.
38. Liang,X., Sun,L., Yu,T., Pan,Y., Wang,D., Hu,X., Fu,Z., He,Q. and Cao,G. (2016) A CRISPR/Cas9 and Cre/Lox system-based express vaccine development strategy against re-emerging Pseudorabies virus. *Sci. Rep.*, **6**, 19176.
39. Weisberg,R.A., Gottesmann,M.E., Hendrix,R.W. and Little,J.W. (1999) Family values in the age of genomics: comparative analyses of temperate bacteriophage HK022. *Annu. Rev. Genet.*, **33**, 565–602.
40. Yagil,E., Dolev,S., Oberto,J., Kislev,N., Ramaiah,N. and Weisberg,R.A. (1989) Determinants of site-specific recombination in the lambdoid coliphage HK022. An evolutionary change in specificity. *J. Mol. Biol.*, **207**, 695–717.
41. Malchin,N., Goltsman,J., Dabool,L., Gorovits,R., Bao,Q., Droge,P., Yagil,E. and Kolot,M. (2009) Optimization of coliphage HK022 Integrase activity in human cells. *Gene*, **437**, 9–13.
42. Kolot,M., Meroz,A. and Yagil,E. (2003) Site-specific recombination in human cells catalyzed by the wild-type integrase protein of coliphage HK022. *Biotechnol. Bioeng.*, **84**, 56–60.
43. Harel-Levy,G., Goltsman,J., Tuby,C.N.J.H., Yagil,E. and Kolot,M. (2008) Human genomic site-specific recombination catalyzed by coliphge HK022 integrase. *J. Biotechnol.*, **134**, 45–54.
44. Voziyanova,E., Malchin,N., Anderson,R.P., Yagil,E., Kolot,M. and Voziyanov,Y. (2013) Efficient Flp-Int HK022 dual RMCE in mammalian cells. *Nucleic Acids Res*., **41**, e125
45. Kolot,M., Malchin,N., Elias,A., Gritsenko,N. and Yagil,E. (2015) Site promiscuity of coliphage HK022 integrase as tool for gene therapy. *Gene Ther*., **22**, 602.
46. Dorgai,L., Yagil,E. and Weisberg,R. (1995) Identifying determinants of recombination specificity: construction and characterization of mutant bacteriophage integrases. *J. Mol. Biol.*, **252**, 178–188.
47. Steyert,S.R. and Pineiro,S.A. (2007) Development of a novel genetic system to create markerless deletion mutants of Bdellovibrio bacteriovorus. *Appl. Environ. Microbiol.*, **73**, 4717–4724.
48. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) In: *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, NY.
49. Unger,T., Jacobovitch,Y., Dantes,A., Bernheim,R. and Peleg,Y. (2010) Applications of the Restriction Free (RF) cloning procedure for molecular manipulations and protein expression. *J. Struct. Biol.*, **172**, 34–44.
50. Kolot,M., Silberstein,N. and Yagil,E. (1999) Site-specific recombination in mammalian cells expressing the Int recombinase of bacteriophage HK022. *Molec. Biol. Reports*, **26**, 207–213.
51. Biswas,T., Aihara,H., Radman-Livaja,M., Filman,D., Landy,A. and Ellenberger,T. (2005) A structural basis for allosteric control of DNA recombination by lambda integrase. *Nature*, **435**, 1059–1066.
52. Weisberg,R.A. and Landy,A. (1983) Site-specific recombination in phage lambda. In: Hendrix,R.W., Roberts,J.W., Stahl,F.W. and Weisberg,R.A. (eds). *Lambda II*. Cold Spring Harbor Laboratory, NY, pp. 211–250.

53. Franz,B. and Landy,A. (1995) The Holliday junction intermediates of lambda integrative and excisive recombination respond differently to the bending proteins integration host factor and excisionase. *EMBO J.*, **14**, 397–406.
54. Miller,H.I., Mozola,M.A. and Friedman,D.I. (1980) int-h: an int mutation of phage lambda that enhances site-specific recombination. *Cell*, **20**, 721–729.
55. Tay,Y., Ho,C., Droge,P. and Ghadessy,F.J. (2010) Selection of bacteriophage lambda integrases with altered recombination specificity by in vitro compartmentalization. *Nucleic Acids Res.*, **38**, e25.
56. Lange-Gustafson,B.J. and Nash,H.A. (1984) Purification and properties of Int-h, a variant protein involved in site-specific recombination of bacteriophage lambda. *J. Biol. Chem.*, **259**, 12724–12732.
57. Siau,J.W., Chee,S., Makhija,H., Wai,C.M., Chandra,S.H., Peter,S., Droge,P. and Ghadessy,F.J. (2015) Directed evolution of lambda integrase activity and specificity by genetic derepression. *Protein Eng. Des. Sel.*, **28**, 211–220.
58. Wu,Z., Gumport,R.I. and Gardner,J.F. (1997) Genetic analysis of second-site revertants of bacteriophage lambda integrase mutants. *J Bacteriol*, **179**, 4030–4038.
59. Shotelersuk,V., Larson,D., Anikster,Y., McDowell,G., Lemons,R., Bernardini,I., Guo,J., Thoene,J. and Gahl,W.A. (1998) CTNS mutations in an American-based population of cystinosis patients. *Am. J. Hum. Genet.*, **63**, 1352–1362.
60. Koenig,M., Hoffman,E.P., Bertelson,C.J., Monaco,A.P., Feener,C. and Kunkel,L.M. (1987) Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell*, **50**, 509–517.
61. Kerem,B., Rommens,J.M., Buchanan,J.A., Markiewicz,D., Cox,T.K., Chakravarti,A., Buchwald,M. and Tsui,L.C. (1989) Identification of the cystic fibrosis gene: genetic analysis. *Science*, **245**, 1073–1080.
62. Bogdanove,A.J., Bohm,A., Miller,J.C., Morgan,R.D. and Stoddard,B.L. (2018) Engineering altered protein-DNA recognition specificity. *Nucleic Acids Res.*, **46**, 4845–4871.
63. Olorunniji,F.J., Rosser,S.J. and Stark,W.M. (2016) Site-specific recombinases: molecular machines for the genetic revolution. *Biochem. J.*, **473**, 673–684.
64. Malchin,N., Molotsky,T., Borovok,I., Voziyanov,Y., Kotlyar,A.B., Yagil,E. and Kolot,M. (2010) High efficiency of a sequential recombinase-mediated cassette exchange reaction in Escherichia coli. *J. Mol. Microbiol. Biotechnol.*, **19**, 117–122.
65. Patsey,R.L. and Bruist,M.F. (1995) Characterization of the interaction between the lambda intasome and attB. *J. Mol. Biol.*, **252**, 47–58.
66. Lorbach,E., Christ,N., Schwikardi,M. and Droge,P. (2000) Site-specific recombination in human cells catalyzed by phage lambda integrase mutants. *J. Mol. Biol.*, **296**, 1175–1181.
67. Christ,N., Corona,T. and Droge,P. (2002) Site-specific recombination in eukaryotic cells mediated by mutant lambda integrases: Implications for synaptic complex formation and the reactivity of episomal DNA segments. *J. Mol. Biol.*, **319**, 305–314.
68. Christ,N. and Droge,P. (2002) Genetic manipulation of mouse embryonic stem cells by mutant lambda integrase. *Genesis.*, **32**, 203–208.
69. Corona,T., Bao,Q.Y., Christ,N., Schwartz,T., Li,J.M. and Droge,P. (2003) Activation of site-specific DNA integration in human cells by a single chain integration host factor. *Nucleic Acids Res.*, **31**, 5140–5148.
70. Aihara,H., Kwon,H.J., Nunes-Duby,S.E., Landy,A. and Ellenberger,T. (2003) A conformational switch controls the DNA cleavage activity of lambda integrase. *Mol. Cell*, **12**, 187–198.
71. Barzel,A., Paulk,N.K., Shi,Y., Huang,Y., Chu,K., Zhang,F., Valdmanis,P.N., Spector,L.P., Porteus,M.H., Gaensler,K.M. *et al.* (2015) Promoterless gene targeting without nucleases ameliorates haemophilia B in mice. *Nature*, **517**, 360–364.
72. Borel,F., Tang,Q., Gernoux,G., Greer,C., Wang,Z., Barzel,A., Kay,M.A., Shultz,L.D., Greiner,D.L., Flotte,T.R. *et al.* (2017) Survival Advantage of Both Human Hepatocyte Xenografts and Genome-Edited Hepatocytes for Treatment of alpha-1 Antitrypsin Deficiency. *Mol. Ther.*, **25**, 2477–2489.
73. Lehmann,M., Pasamontes,L., Lassen,S.F. and Wyss,M. (2000) The consensus concept for thermostability engineering of proteins. *Biochim. Biophys. Acta*, **1543**, 408–415.
74. Durani,V. and Magliery,T.J. (2013) Protein engineering and stabilization from sequence statistics: variation and covariation analysis. *Methods Enzymol.*, **523**, 237–256.
75. Cirri,E., Brier,S., Assal,R., Canul-Tec,J.C., Chamot-Rooke,J. and Reyes,N. (2018) Consensus designs and thermal stability determinants of a human glutamate transporter. *Elife*, **7**, 40110.
76. Korkegian,A., Black,M.E., Baker,D. and Stoddard,B.L. (2005) Computational thermostabilization of an enzyme. *Science*, **308**, 857–860.
77. Borgo,B. and Havranek,J.J. (2012) Automated selection of stabilizing mutations in designed and natural proteins. *Proc. Natl. Acad. Sci. U. S. A*, **109**, 1494–1499.
78. Buss,O., Rudat,J. and Ochsenreither,K. (2018) FoldX as protein engineering tool: better than random based approaches? *Comput. Struct. Biotechnol. J.*, **16**, 25–33.
79. Elias,A., Gritsenko,N., Gorovits,R., Spector,I., Prag,G., Yagil,E. and Kolot,M. (2018) Anti-cancer binary system activated by bacteriophage HK022 integrase. *Oncotarget.*, **9**, 27487–27501.
80. Hu,J., Meyers,R.M., Dong,J., Panchakshari,R.A., Alt,F.W. and Frock,R.L. (2016) Detecting DNA double-stranded breaks in mammalian genomes by linear amplification-mediated high-throughput genome-wide translocation sequencing. *Nat. Protoc.*, **11**, 853–871.
81. David,D., Princiero,B.S., Elmonem,M.A., Oliveira,A.F., Soliman,N., van den Heuvel,B., Gijsbers,R. and Levtchenko,E. (2019) Molecular basis of cystinosis: geographic distribution, functional consequences of mutations in the CTNS gene, and potential for repair. *Nephron*, **141**, 133–146.
82. Fokkema,I.F., Taschner,P.E., Schaafsma,G.C., Celli,J., Laros,J.F. and den Dunnen,J.T. (2011) LOVD v.2.0: the next generation in gene variant databases. *Hum. Mutat.*, **32**, 557–563.
83. Veit,G., Avramescu,R.G., Chiang,A.N., Houck,S.A., Cai,Z., Peters,K.W., Hong,J.S., Pollard,H.B., Guggino,W.B., Balch,W.E. *et al.* (2016) From CFTR biology toward combinatorial pharmacotherapy: expanded classification of cystic fibrosis mutations. *Mol. Biol. Cell*, **27**, 424–433.
84. Aiuti,A., Biasco,L., Scaramuzza,S., Ferrua,F., Cicalese,M.P., Baricordi,C., Dionisio,F., Calabria,A., Giannelli,S., Castiello,M.C. *et al.* (2013) Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott-Aldrich syndrome. *Science*, **341**, 1233151.