Development of a papillation assay using constitutive promoters to find hyperactive transposases

Michael Tellier[2]* and Ronald Chalmers[1]*

*Corresponding authors: michael.tellier@path.ox.ac.uk, ronald.chalmers@nottingham.ac.uk,

[1] School of Life Sciences, University of Nottingham, Queen's Medical Centre, Nottingham, NG7 2UH, UK

[2] Current address: Sir William Dunn School of Pathology, University of Oxford, Oxford, OX1 3RF, UK

Abstract

Background

Transposable elements is an extremely diverse group of genetic elements encoding their own mobility. This ability has been exploited as a powerful tool for molecular biology and genomics techniques. However, transposition activity is regulated by *cis* and/or *trans* mechanisms because of the need to co-exist with their host. This represents a limitation to their usage as biotechnological tools. The development of screening assays and the improvement of current ones is therefore needed to find hyperactive transposases.

Results

We present in this study an improvement of the well-known papillation assay where in place of an inducible promoter, we designed a set of constitutive promoters cloned into a one or five copies vector in presence or absence of a ribosome binding site. This set of vectors provides a wide range of transposase expression and offers a more uniform expression of the transposase across cells compared to inducible promoters. These constructs can therefore be used to screen for hyperactive transposases or for transposases resistant to overproduction inhibition, a mechanism affecting DNA transposases such as Hsmar1, which decreases the transposition rate when the transposase concentration increases. We characterized and validated our set of vectors with the Hsmar1 transposase and took advantage of our approach to investigate the effects on the transposition rate of inserting mutations in the Hsmar1 dimer interface or of covalently binding two Hsmar1 monomer.

Conclusions

This improved papillation assay should be applicable to a wide variety of DNA transposases. It also provides a straightforward approach to screen transposase mutant libraries with a specific expression level to find hypoactive, hyperactive or overproduction inhibition resistant transposases. Our approach could also be useful for synthetic biology as a combination of the wild type or covalently bound Hsmar1 transposase with a library of weak promoters offers the possibility to find promoters expressing on average one or two proteins per cell.

Key words: transposon; transposase; overproduction inhibition; papillation assay; Hsmar1

Background

Transposable elements (TEs) are DNA sequences encoding their own ability to move in a genome from one place to another. They are found in virtually all organisms and are particularly present in eukaryotes where they can represent a high percentage of the genome (1-3). Originally described as selfish elements due to their ability to replicate independently of their host, TEs have now been shown to be important drivers of genome evolution (4, 5). Indeed, TEs can provide novel transcription factors binding sites, promoters, exons or poly(A) sites and can also be co-opted as micro RNAs or long intergenic RNAs (6-8).

TEs are therefore a diverse group of DNA sequences using a wide range of mechanisms to transpose inside and between hosts. For example, some DNA transposons like the *mariner* elements transpose through a “cut and paste” mechanism where a single copy of the transposon moves from one place to another without copying itself (9). Over the past several years, our group and others have described the mechanisms regulating the transposition rate of different DNA transposases, such as Hsmar1 or Mos1 (10-15). In Hsmar1, the central regulatory mechanism of transposition is overproduction inhibition (OPI) (16), a phenomenon emerging from the decreased binding affinity of a transposase dimer for the second end of the transposon after binding the first end (10). This decreased affinity increases the probability of another Hsmar1 dimer to bind the free end, blocking transposition at the same time. Thus, OPI curbs the transposition of Hsmar1 to avoid an exponential transposition rate which would prove catastrophic for the host (12).

100 However, OPI represents a limitation in the development of hyperactive
101 transposases, which are for example needed for transposon mutagenesis.
102 Several approaches can be used to overcome OPI such as modifying the
103 binding kinetics of the transposase to the ITR or the monomer-dimer
104 equilibrium. Indeed, we and others previously shown that most mutations in
105 the conserved WVPHEL motif in Himar1 and Hsmar1, which is involved in the
106 subunit interface, result in hyperactive transposases but at the cost of
107 producing nonproductive DNA double-strand breaks and therefore DNA
108 damage (17, 18).
109 The discovery of hyperactive transposases in bacteria has mostly been
110 accomplished by screening libraries of transposase mutants with the
111 papillation assay (Supplementary Figure 1a) (19, 20). This assay is based on
112 a promoter-less lacZ gene flanked by transposon ends. This reporter is
113 integrated in a silent region of the genome of *Escherichia coli*. The
114 transposase gene is provided *in trans* on a plasmid to simplify mutagenesis
115 and library handling. Transposition events into an expressed ORF give rise to
116 lacZ gene fusion proteins. When this happens within a colony growing on an
117 X-gal indicator plate, it converts the cell to a lac+ phenotype, which allows the
118 outgrowth of a blue microcolony (papillae) on a background of white cells. The
119 transposition rate is given by the number of papillae per colony and by the
120 rate of their appearance. The visual read-out of this assay simplifies the
121 screening of mutant libraries for discovering hypo- or hyper-active
122 transposases. The mating-out assay is a more quantitative assay using the
123 same reporter strain. Here, the rate of transposition is measured by

movement of the reporter from its chromosomal location into a conjugative plasmid.

In the current work, we present an improvement of the papillation assay by using a set of constitutive promoters cloned in a single or five copies vector in absence or presence of a ribosome binding site (RBS). This set of expression vectors allows us to express a transposase across a wide range of expression level facilitating the screening of hyperactive and/or OPI-resistant transposases. We used this set of vectors to compare an Hsmar1 monomer to a covalently bound Hsmar1 dimer and to test for hyperactivity and OPI-resistance several Hsmar1 mutants. We found that one Hsmar1 mutant in the dimer interface, R141L, is resistant to OPI in *E. coli*.

Results

Characterization of the constitutive promoters

The papillation assay provides a visual assessment of the transposition rate, which can be determined from the rate of papillae apparition and their number per colony (19). The transposition rate is dependent of the concentration and the activity of the transposase (12). Conventionally, transposition assays use inducible promoters which limits the accessible range of transposase expression and also result in cell-to-cell variability due to the unequal diffusion of the inducer in colonies. To overcome these limitations, we synthesized a set of five constitutive promoters (00, JJ, K, E, and W) based on (21). In addition, we created a null-expression vector by replacing the promoter and the RBS with a featureless, undistorted, DNA sequence consisting of a single repeat of the tetranucleotide $(GACT)_{n=44}$. To increase the range of expression

levels obtainable, we also created a variant of each promoter where the RBS has been abolished. The expression construct is shown in Figure 1A and is composed of the promoter and a RBS sequence, an NdeI and BamHI restriction sites to clone a gene of interest which can then be fused to a C-terminal 3x FLAG tag. To avoid any read-through transcription, the construct is flanked by terminator sequences. The expression constructs were cloned either into a one-copy vector or a five-copy vector, pBACe3.6 and pGHM491, respectively. The following nomenclature will be used: Bp1 to Bp6 represents the six promoters cloned into the single copy vector, Ip1 to Ip6 corresponds to the six promoters cloned into the five copy vector, the '-' and '+' represents the absence or the presence of a RBS, respectively, while the 'N' corresponds to the RBS composed of the GACT repeat in the p1 promoter.

We first determined the strength of each expression vector by inserting an *EGFP* gene in each vector to investigate by flow cytometry the amount of fluorescence produced (Figure S1B). To rank the expression vectors, we normalize their average fluorescence value against the strongest vector, Ip6+ (Figure 1B). Most of the one-copy expression vectors produce an amount of EGFP close to the detection threshold and therefore their ranking might not be representative. However, we can observe that the five-copy expression vectors produce more fluorescence than the one-copy vectors and that the vectors with a consensus RBS are also producing more fluorescence than the vectors without a RBS motif, except for the p1 promoter where the presence or absence of the RBS does not influence the amount of fluorescence produced. We confirmed the flow cytometry results for the strongest expression vectors expressing Hsmar1 by performing western blot with an

anti-Hsmar1 antibody (Figure 1C). We also compared by western blotting these constructs with the inducible promoter normally used for papillation assay, the pTac promoter (Figure 1D). Interestingly, two of our constructs (Ip5+ and Ip6+) produce a higher amount of Hsmar1 than the pTac promoter fully induced with 1 mM of IPTG.

Characterization of the papillation assay with the wild-type Hsmar1 transposase

To characterize the implementation of the constitutive promoters into the papillation assay, we used the wild-type Hsmar1 transposase as we have already well characterized its activity in papillation assay (18). We defined the transposition rate as the average number of papillae per colony after five days of incubation at 37°C. Bacterial cells stop dividing after a few days at 37°C because of carbon exhaustion, therefore affecting the number of visible papillae since late transposition events, i.e. a few divisions before the cells stop dividing because of carbon exhaustion, will not be visible because of the insufficient number of division. To overcome this limitation, we took advantage of the fact that bacterial cells having undergone a transposition event resulting in a fusion between a host gene and the lacZ gene, located in the transposon, will be able to use lactose as a carbon source. Thus, by adding lactose to the medium we should recover a higher number of papillae per colony since the lac^+ cells resulting from a late transposition event would be able to continue to grow by using lactose (Supplementary Figure 2). Indeed, by using the Ip3+ expression vector and a range of lactose concentration, we observed a correlation between the number of papillae per colony and the lactose

199 concentration (Supplementary Figure 2A/ to C/). We decided to use a
200 concentration of lactose of 0.1% for the papillation assay since it is
201 representing the best trade-off between the number of papillae per colony and
202 the size of the papillae to allow a proper quantitation at high transposition rate.
203 We next investigated the transposition rate of each expression vector with the
204 wild-type Hsmar1 transposase. Representative colony of the papillation assay
205 for each Hsmar1 expression vectors is shown in Figure 2A and their
206 respective whole plate pictures are shown in Supplementary Figure 3. As
207 expected from the wide-range of expression, we observed a 200-fold variation
208 in the average number of papillae per colony as the quantitation of the
209 different expression vectors shows in Figure 2B. To better visualize the
210 relationship between the expression vector strength and the transposition
211 rate, determined by the number of papillae per colony, we plotted the vector
212 strength determined by flow cytometry (Figure 1B) against the number of
213 papillae per colony (Figure 2C). As previously published *in vitro*, in *E. coli* and
214 in HeLa cells, the wild-type Hsmar1 transposase follows an inverse-
215 exponential relationship between transposase expression and transposition
216 rate (12, 22). Interestingly, the transposition rate peaks with the weakest
217 promoters in the single copy vector indicating that the transposition rate is
218 extremely sensitive to overproduction inhibition (OPI) resulting from a slight
219 increase in the transposase concentration. However, for the expression
220 vectors with a similar average promoter strength but a transposition rate
221 below the highest one we cannot determine whether their suboptimal
222 transposition rate is due to an insufficient amount of Hsmar1 transposase in
223 the cell or OPI.

Importantly, the transposition assay also provides a more precise approach to investigate weak promoter strength than flow cytometry with EGFP. Indeed, out of the 18 expression vectors having a relative promoter strength around 4% of Ip6+, the transposition assay shows a 10-fold change in transposition rate within this set of vectors (Figure 2C).

Bounding covalently two Hsmar1 monomers in a dimer affects the transposition rate

We recently published a new Hsmar1 construct where two monomers are covalently bound by a linker region (Figure 3A) (23). We use advantage of our approach to test whether the transposition rate of a covalently bound Hsmar1 dimer differs to the Hsmar1 monomer. At low expression level, we expect a covalently bound Hsmar1 dimer to transpose more efficiently than an Hsmar1 monomer because of the physical link in the covalent dimer, which keeps both monomers close to each other, and also because the covalent dimer requires a single translation event whereas the monomer requires two. We cloned the monomeric and dimeric construct in a set of expression vectors spanning from very low to high expression and performed a papillation assay.

Representative colony of the papillation assay of each expression vectors is shown in Figure 3B and their respective whole plate pictures are shown in Supplementary Figure 4. Interestingly, we observe a change in the number of papillae per colony with the lowest expression vectors, as shown by the quantitation in Figure 3C. When compared to the results obtained with Hsmar1 monomer, the covalent dimer transposition rate peaks at a different set of expression vectors, Bp2- and Bp3- for the covalent dimer and Ip2- and

Ip1+ for the monomer. These four expression vectors have a similar relative promoter strength, around 4% of Ip6+, indicating that the number of transposases expressed per cell is particularly low (see Discussion). Inversely, we do not observe any difference in the number of papillae per colony with stronger expression vectors such as Ip3+ and Ip6+ (Figure 3B to D). This indicates that a covalently bound Hsmar1 dimer is as sensitive to OPI as the monomer, which is also supported by the inverse relationship between the average promoter strength and the transposition rate for both the monomer and the covalent dimer (Figure 3D).

SETMAR transposition activity was lost during the same period than Hsmar1 transposase domestication

The Hsmar1 transposase was originally found in the human genome where an Hsmar1 transposase with several mutations is fused to a SET domain to form SETMAR (24-26). The domesticated Hsmar1 transposase is inefficient at performing transposition because of the mutation of the last DDD triad catalytic motif to a DDN (25, 26). In addition to the D282N mutation, we performed a papillation assay with a non-induced pTac promoter the 22 other mutations present in the human SETMAR to determine their effects on Hsmar1 transposition (Figure 4A). Most of the mutations present in the human SETMAR occurred outside the DNA binding domain and happened at the same time as the domestication of the Hsmar1 transposase. Representative colony of the papillation assay of each expression vectors is shown in Figure 4B and their respective whole plate pictures are shown in Supplementary Figure 5. In addition to D282N, two other mutations, C219A and S279L,

disrupt completely Hsmar1 transposition activity. Two other mutations located in the DNA binding domain, E2K and R53C, affect severely the transposition rate. In addition, seven other mutations located mostly in the catalytic domain mildly affect Hsmar1 transposition activity. Only one mutation, V201L, increases Hsmar1 transposition rate whereas the remaining mutations were neutral. Interestingly, nearly all deleterious mutations for transposition arose at the same time as the domestication of the Hsmar1 transposase.

Mutations in Hsmar1 dimer interface produce hyperactive mutants in bacteria

The mutagenic nature of transposable elements make them useful in screening for essential genes. However, OPI limits the transposition rate when the transposase concentration is too high (12). One way to overcome OPI is to decrease the stability of the Hsmar1 dimer to shift the monomer-dimer equilibrium to the inactive monomeric form. We decided to take advantage of our approach to investigate two Hsmar1 transposases mutated in the dimer interface, one known mutant, F132A (F460 in SETMAR (27)) and a novel one R141L (9). We used three vectors expressing Hsmar1 at a low (Bp1+), optimal (Ip1+) and high (Ip6+) level. Representative colony of the papillation assay of each expression vectors is shown in Figure 4C and their respective whole plate pictures are shown in Supplementary Figure 6. The average number of papillae per colony is indicated below each representative colony. Interestingly, both F132A and R141L transposases are hyperactive at low and optimal levels of expression when compared to WT. A higher transposition rate is also observed at high expression level for both mutants,

299 with R141L showing a stronger resistance to OPI than F132A. To confirm the
300 papillation assay, the mutants' transposition rate was also determined using
301 the mating-out assay, a more quantitative assay measuring the transposition
302 rate through the movement of the transposon reporter from its chromosomal
303 location into a conjugative plasmid (Table 1). The results of the mating-out
304 and transposition assays were similar. Interestingly, Hsmar1 R141L
305 transposition rate is not affected by the high transposase expression level
306 produced by Ip6+, as the rate remains similar between Ip1+ and Ip6+
307 whereas we observe a 147 fold decrease for the wild type transposase and a
308 17 fold decrease for the F132A mutant.
309

Construct	Transposition frequency	Mutant/W.T.
Ip1+ W.T.	4.73 (±1.02) x 10^{-5}	
Ip1+ F132A	9.73 (±4.53) x 10^{-4}	21
Ip1+ R141L	2.42 (±1.68) x 10^{-4}	5
Ip6+ W.T.	3.22 (±1.02) x 10^{-7}	
Ip6+ F132A	5.79 (±2.63) x 10^{-5}	180
Ip6+ R141L	3.24 (±1.43) x 10^{-4}	1006

310 **Table 1: Transposition frequencies of two Hsmar1 transposase mutants**
311 **expressed at optimal and high level.** The bacterial mating-out assays have been
312 done with the RC5097 strain and the Ip1+ or Ip6+ vectors. Transposition frequencies
313 are the average of three independent experiments ± standard error of the mean.
314

315 **Discussion**

We present here an improvement of the papillation assay using a set of constitutive promoters cloned into a single- or five-copies vector in absence or presence of a RBS. This range of expression vectors give us a better control of the transposase expression compared to inducible vectors such as the pTac. This is illustrated in Figure 2C where we observe a large variation in the number of papillae per colony across expression vectors producing EGFP fluorescence close to the background level. This indicates that the transposition rate is extremely sensitive to small variation in transposase concentration, in agreement with previous works from our group (10, 12).

We recently published a covalently bound Hsmar1 construct, where a single transcription and translation event is sufficient to synthesize an active Hsmar1 dimer (23). At low expression level, a change in the transposition rate is observed for the covalent dimer construct when compared to the monomeric construct whereas at higher expression level, the transposition rates are quite similar indicating that OPI is occurring for both the monomer and the covalent dimer Hsmar1 (Figure 3C and D). The change in transposition rate at low expression level is expected because a single translation event is needed for producing a covalently bound Hsmar1 dimer whereas the wild type Hsmar1 requires two translation events for producing an active transposase. Based on this idea, we can hypothesize that Bp2- and Bp3-, which provides the highest transposition rates for the covalent dimer, corresponds to weaker promoters than Ip2- and Ip1+, which provides the highest transposition rates for the monomeric Hsmar1 but lower transposition rate for the covalent dimer. Thus, Bp2- and Bp3- are likely to express on average less than two proteins per cell,

341 which is not sufficient to optimally promote transposition for the Hsmar1
342 monomer construct, whereas Ip2- and Ip1+ are likely to express on average
343 at least two proteins per cell, which starts to promote OPI for the covalent
344 dimer construct and therefore results in a lower transposition rate than Bp2-
345 and Bp3-.
346
347 A peculiarity of the Hsmar1 transposase is the presence in anthropoid
348 primates of SETMAR, a gene with new functions in gene regulation resulting
349 from the fusion between a SET gene and an Hsmar1 transposase (24)
350 (Tellier, M. and Chalmers, R., manuscript under review). It was shown that the
351 domesticated Hsmar1 transposase DNA binding domain was under purifying
352 selection whereas the catalytic domain was evolving under neutral selection,
353 with a mutation of the last D of the catalytic triad DDD to an N abolishing
354 SETMAR transposase activity (24-26). In addition to the mutation in the
355 catalytic triad, the domesticated Hsmar1 contains 22 other mutations with
356 three of them located in the DNA binding domain (E2K, R53C and D98N)
357 (Figure 4A and B). Out of the 23 mutations, only one was found to increase
358 the transposition rate in the papillation assay, V201L, whereas 12 mutations
359 were deleterious for the transposition rate with three of them abolishing it
360 completely (C219A, S279L and D282N). Interestingly, 11 of the 12 deleterious
361 mutations occurred at the same time as Hsmar1 was domesticated, resulting
362 in the abolition of SETMAR ability to promote transposition, and are therefore
363 common to all anthropoid primates. Two of the DNA binding mutants, E2K
364 and R53C, are deleterious to Hsmar1 transposition activity in a papillation
365 assay. It will be interesting to determine whether this effect is mediated

366 through a change in ITR binding efficiency, which could have modified
367 SETMAR's ability to bind ITRs in the genome and therefore its emerging
368 functions in regulating gene expression.
369
370 Transposases have become an important and versatile biotechnological tool
371 (28-30). The creation of hyperactive transposases can be advantageous over
372 its wild-type counterpart as for transposon mutagenesis for example. One of
373 the major mechanism limiting the transposition rate of *mariner* transposases is
374 OPI, which is occurring when there is an excess of transposase dimer per
375 transposon (11, 12). We previously show that mutating the conserved
376 WVPHEL motif, which is part of the Hsmar1 transposase dimer interface,
377 resulted in mostly hyperactive transposases (18, 31). Here we show that the
378 mutation of the residues F132 and R141, which are located in the subunit
379 interface (9, 27), also produces hyperactive transposases in *E. coli* with the
380 R141L mutant being OPI-resistant (Figure 4C and Table 1). The hyperactivity
381 of F132A and R141L mutants could be explained by the promotion of one of
382 the conformational change essential for transposition (11). The decreased
383 OPI-sensitivity could result from a decrease in the dimer stability, which shifts
384 the monomer-dimer equilibrium towards the monomeric form, and therefore
385 reduces the concentration of active transposases in the cell. Also, an unstable
386 dimer bound to a transposon end could be more likely to fall apart allowing the
387 recruitment of the previously bound end by another bound dimer, activating
388 transposition. This type of mutant is more likely to be found hyperactive only
389 in bacteria. Indeed, in mammalian cells the size of the nucleus and the bigger
390 ratio of non-specific DNA to specific DNA, i.e. the transposon ends account

for a smaller fraction of the genome, will increase the time necessary for a transposase to find a transposon end. Therefore, transposases with a weakened dimer interface are more likely to revert to an inactive monomeric state resulting in hypoactive mutants.

An unexpected outcome of our range of expression vectors is the realization that the transposition rate could be used as a better approach than EGFP fluorescence to compare the strength of a series of weak promoters. This is illustrated in Figure 1B where 18 of our expression vectors have a relative promoter strength comprised between 3 and 4% of Ip6+, our strongest vector. However, in Figure 2C, we observe a 10-fold difference in the transposition rate between these 18 expression vectors. It is interesting to note that the seven expression vectors with the highest transposition rate (Figure 2C) are either based on the p1 promoter, which is a GACT track without promoter activity, or proper promoters but without a RBS site indicating that a stochastic Hsmar1 mRNA and/or protein production provides the highest transposition rate in *E. coli*. We previously shown that OPI starts to occur when two transposase dimers are present in a single cell (12). This therefore shows that the expression vectors with the highest transposition rate (Bp1-, Ip1+ and Bp5-) are the closest to the production of a single Hsmar1 dimer per cell across the bacterial colony. Thus, the papillation assay with the Hsmar1 covalent dimer could be used to screen promoters to find a promoter expressing on average one protein per cell across a bacterial colony. The advantage of the papillation assay is that the transposition rate is quantified on multiple bacterial colonies which allows a comparison of the transposition

rate across a population of cells. We can therefore confidently differentiate between noise and real change in transposition efficiency due to OPI, as shown in Figure 2C.

Conclusions

We present in this study an improvement of the papillation assay using a set of constitutive promoters cloned in a one- or five-copies vector in presence or absence of a ribosome binding site. This set of expression vectors gives a better control of the transposase expression and the possibility to screen for hyperactive or OPI-resistant transposases by using one of the optimal or high expression vector, respectively. A potentially interesting approach to the synthetic biology community is an experimental method based on the covalently bound Hsmar1 transposase or the wild-type Hsmar1 that could be used for screening weak promoters to find a promoter expressing on average one or two proteins per cell, respectively.

Methods

Media and bacterial strains

Bacteria were grown in Luria-Bertani (LB) media at 37°C. The following antibiotics were used at the indicated concentrations: ampicillin (Amp), 100 µg/ml), chloramphenicol (Cm), 25 µg/ml, and spectinomycin (Spec), 100 µg/ml. The following Escherichia coli strains were used: RC5024 (identical to DH5α) [endA1 hsdR17 glnV44 thi-1 recA1 gyrA relA1 Δ(lacIZYA-argF)U169 deoR (φ80dlac Δ(lacZ)M15)], RC5094 [F- araD139 Δ(argF-lac)U169 rspL150 relA1 flbB5301 fruA25 deoC1 ptsF25 rpoS359::Tn*10*], RC5096 [F$^-$ fhuA2

Δ(lacZ)r1 glnV44 e14-(McrA-) trp-31 his-1 rpsL104 xyl-7 mtl-2 metB1 Δ(mcrC-mrr)114::*IS10* argE::Hsmar1-lacZ'-kanR] and RC5097 (= RC5096 pOX38::miniTn10-CAT).

Constitutive promoters

Alper et al previously generated and characterized a set of constitutive promoters based on pltetO ranging from strong down to very weak (21). We select the promoters 00, jj, K, E, and W (equivalent to p2, p3, p4, p5, and p6 in this study) and generate p1, a featureless tract of 44 GACT repeats which we represent an ideal promoter-less region (Table 1). Each promoter sequence is preceded by three terminator sequences and followed by a consensus or a null ribosome binding site (RBS) (and also a GACT RBS in the case of p1), a transposase gene, three Flag tag and a terminator sequence (Figure 1A).

Promoter name	Sequence	mRNA production value
p1	CTGACTGACTGACTGACTGACTGACTGACTGACTGACT GACTGACTGACTGACTGACTGACTGACTGACTGACTG ACTGACTGACTGACTGACTGACTGACTGACTGAC TGACTGACTGACTGACTGACTGACTGACTGACTGACTG ACTGACTGACTGACTGACTGACTGACCATATG	n.d.
p2 (00)	CAATTCCGACGTCTAAGGAAACCATTATCATGACATCA ACCTATAAAAATAGGCGTATCACGAGGCCCTCTCGTCT CCACCTCAAGCTCCCTATCTAGTGATAGCGATTGACAT	0.003

	CCCTATCAGTGACGGAGATATTGAGCACATCAGCAGG ACGCACTGACCACTTTAAGAAGGAGATATACATATG	
p3 (JJ)	CAATTCCGACGTCTAAGAAACCATTATTATCATGACATT AACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTC TTCACCTCGAGTCCCTATCAGTGATAGAGATTGACCTC CCTATCAGTGATAGAGATACTGAGCACATCAGCAGGA CGCACTGACCACTTTAAGAAGGAGATATACATATG	0.159
p4 (K)	CAATTCCGACGTCTAAGAAACCATTATTATCATGACATT AACCTATAAAAATAGGCGTATCACGAGGCCCTCTCGTC TTCACCTCGAGTCCCTATCAGTGATAGGGATTGACATC CCTATCAGTGATAGAGACACTGGGCACATCAGCAGGA CGCACTGACCACTTTAAGAAGGAGATATACATATG	0.299
p5 (E)	CAATTCCGACGCCTAAGAAACCATTATTATCATGACATT AGCCTATAAAAATAGGCGTACCACGAGGCCCTTTCGTC TTCACCTCGAGTCCCTATCAGTGATAGAGATTGACACC CCTATCAGTGATAGAGATACTGAGCACATCAGCAGGA CGCACTGACCACTTTAAGAAGGAGATATACATATG	0.743
p6 (W / pItetO)	CAATTCCGACGTCTAAGAAACCATTATTATCATGACATT AACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTC TTCACCTCGAGTCCCTATCAGTGATAGAGATTGACATC CCTATCAGTGATAGAGATACTGAGCACATCAGCAGGA CGCACTGACCACTTTAAGAAGGAGATATACATATG	1

Table 2: List of constitutive promoters.

Nomenclature (the letters indicated between brackets are from (21)), sequence, and strength of the constitutive promoters used in this study. n.d.: not determined.

Flow cytometry

RC5096 cells expressing EGFP were grown overnight at 37°C in LB medium supplemented with chloramphenicol or spectinomycin. The cultures were diluted in a 1:1000 ratio in fresh LB medium complemented with antibiotics and grown to mid-log phase (OD_{600} ~ 0.5). The cells were pelleted at 6,000g for 5 min, washed in 1X PBS twice, and resuspended in 500 µl of 1X PBS. Flow cytometry analysis was performed on 100,000 cells with a Beckman Coulter Astrios EQ.

Western blotting

Cells containing a derivative of pMAL-c2x were grown in LB supplemented with 100 µg/ml of ampicillin at 37°C until an OD_{600} of ~ 0.5 and were then induced with the required concentration of IPTG for 2 hours at 37°C. Cells containing pGHM491 or pBACe3.6 derivatives were grown in LB supplemented with respectively 100 µg/ml of spectinomycin or 50 µg/ml of chloramphenicol at 37°C for the same amount of time as the induced cells. Promoters' expression were analyzed by pelleting ~$1.5x10^9$ cells. The samples were resuspended in SDS sample buffer, boiled for 5 min, and loaded on 10% SDS-PAGE gels. Proteins were transferred to PVDF membrane, probed with an anti-Hsmar1 antibody (goat polyclonal, 1:500 dilution, ab3823, Abcam) followed by a horseradish peroxidase-conjugated anti-goat secondary antibody (rabbit polyclonal, 1:5000 dilution, ab6741, Abcam). Proteins were visualized by using the ECL system (Promega) and Fuji medical X-ray film (Fujufilm).

Papillation assay

The papillation assay and the reporter strain RC5096 have been described previously (Figure S1A) (18). Briefly, transposase expression vectors were transformed into the RC5096 strain. It is a lac$^-$ *E. coli* strain encoding a transposon containing a promoter-less lacZ and a kanamycin resistance gene flanked with Hsmar1 ends, which has been integrated in a silent genomic locus. In absence of LacZ, the strain produces white colonies on X-gal indicator plates. When the transposase is supplied in trans, the integration of a transposon into the correct reading frame of an active gene will produce a lacZ fusion protein. The descendants of this cell will become visible as blue papillae on X-gal indicator plates. RC5096 transformants were plated on LB-agar medium supplemented with 0.01% lactose, 40 µg/ml of X-gal and either 50 µg/ml of chloramphenicol or 100 µg/ml of spectinomycin. Plates were incubated 5 days at 37°C and photographed. The transposition rate is determined by the number of papillae per colony or by the rate of appearance of papillae on the colonies.

Mating-out assay

A chloramphenicol resistant derivative of the conjugative plasmid pOX38 has been introduced in the RC5096 papillation strains to create the donor strains RC5097. Briefly, RC5097 transformants and the recipient strain, RC5094, were grown overnight in LB supplemented with antibiotics at 37°C. The next day, respectively one and three volumes of RC5097 and RC5094 were centrifuged for 5 min at 6,000x g. Each pellet was resuspended in 3 ml of fresh LB, pool together, and incubated in a shaking water bath for 3 hours at 37°C. After the mating, the transposition events were detected by plating 200

511 µl of each culture on LB-agar medium supplemented with tetracycline and
512 kanamycin. The number of transconjugants was obtained by plating a 10^{-5}
513 fold dilution of each culture on LB-agar medium supplemented with
514 tetracycline and chloramphenicol. The plates were incubated overnight at
515 37°C and the transposition rate determined the next day by dividing the
516 number of kanamycin-resistant cells by the number of chloramphenicol
517 resistant cells.
518
519 **Funding**
520 This work was supported by a Wellcome Trust Grant [WT093160] to RC and a
521 Biotechnology and Biological Sciences Research Council Doctoral Training
522 Program Grant [BB/J014508/1] to MT
523
524 **Availability of data and materials**
525 All data generated or analyzed during this study are included in this published
526 article and its Additional files.
527
528 **Competing interests**
529 The authors declare that they have no competing interests.
530
531 **Authors' contributions**
532 MT and RC designed the study. MT performed the experiments, analyzed the
533 data and wrote the manuscript. All authors read and approved the final
534 manuscript.
535

References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.
2. Feschotte C, Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet. 2007;41:331-68.
3. Aziz RK, Breitbart M, Edwards RA. Transposases are the most abundant, most ubiquitous genes in nature. Nucleic Acids Res. 2010;38(13):4207-17.
4. Orgel LE, Crick FH. Selfish DNA: the ultimate parasite. Nature. 1980;284(5757):604-7.
5. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet. 2017;18(2):71-86.
6. Piriyapongsa J, Jordan IK. A family of human microRNA genes from miniature inverted-repeat transposable elements. PLoS One. 2007;2(2):e203.
7. Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet. 2013;9(4):e1003470.
8. Jangam D, Feschotte C, Betran E. Transposable Element Domestication As an Adaptation to Evolutionary Conflicts. Trends Genet. 2017;33(11):817-31.
9. Tellier M, Bouuaert CC, Chalmers R. Mariner and the ITm Superfamily of Transposons. Microbiol Spectr. 2015;3(2):MDNA3-0033-2014.
10. Claeys Bouuaert C, Chalmers R. Transposition of the human Hsmar1 transposon: rate-limiting steps and the importance of the flanking TA dinucleotide in second strand cleavage. Nucleic Acids Res. 2010;38(1):190-202.
11. Claeys Bouuaert C, Walker N, Liu D, Chalmers R. Crosstalk between transposase subunits during cleavage of the mariner transposon. Nucleic Acids Res. 2014;42(9):5799-808.
12. Claeys Bouuaert C, Lipkow K, Andrews SS, Liu D, Chalmers R. The autoregulation of a eukaryotic DNA transposon. Elife. 2013;2:e00668.
13. Dawson A, Finnegan DJ. Excision of the Drosophila mariner transposon Mos1. Comparison with bacterial transposition and V(D)J recombination. Mol Cell. 2003;11(1):225-35.
14. Auge-Gouillou C, Brillet B, Hamelin MH, Bigot Y. Assembly of the mariner Mos1 synaptic complex. Mol Cell Biol. 2005;25(7):2861-70.
15. Richardson JM, Dawson A, O'Hagan N, Taylor P, Finnegan DJ, Walkinshaw MD. Mechanism of Mos1 transposition: insights from structural analysis. EMBO J. 2006;25(6):1324-34.
16. Lohe AR, Hartl DL. Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. Mol Biol Evol. 1996;13(4):549-55.
17. Lampe DJ. Bacterial genetic methods to explore the biology of mariner transposons. Genetica. 2010;138(5):499-508.

583 18. Liu D, Chalmers R. Hyperactive mariner transposons are created by
584 mutations that disrupt allosterism and increase the rate of transposon end
585 synapsis. Nucleic Acids Res. 2014;42(4):2637-45.
586 19. Huisman O, Kleckner N. A new generalizable test for detection of
587 mutations affecting Tn10 transposition. Genetics. 1987;116(2):185-9.
588 20. Pajunen MI, Rasila TS, Happonen LJ, Lamberg A, Haapa-Paananen S,
589 Kiljunen S, et al. Universal platform for quantitative analysis of DNA
590 transposition. Mob DNA. 2010;1(1):24.
591 21. Alper H, Fischer C, Nevoigt E, Stephanopoulos G. Tuning genetic
592 control through promoter engineering. Proc Natl Acad Sci U S A.
593 2005;102(36):12678-83.
594 22. Blundell-Hunter G, Tellier M, Chalmers R. Transposase subunit
595 architecture and its relationship to genome size and the rate of transposition
596 in prokaryotes and eukaryotes. Nucleic Acids Res. 2018.
597 23. Claeys Bouuaert C, Chalmers R. A single active site in the mariner
598 transposase cleaves DNA strands of opposite polarity. Nucleic Acids Res.
599 2017;45(20):11467-78.
600 24. Cordaux R, Udit S, Batzer MA, Feschotte C. Birth of a chimeric primate
601 gene by capture of the transposase gene from a mobile element. Proc Natl
602 Acad Sci U S A. 2006;103(21):8101-6.
603 25. Liu D, Bischerour J, Siddique A, Buisine N, Bigot Y, Chalmers R. The
604 human SETMAR protein preserves most of the activities of the ancestral
605 Hsmar1 transposase. Mol Cell Biol. 2007;27(3):1125-32.
606 26. Miskey C, Papp B, Mates L, Sinzelle L, Keller H, Izsvak Z, et al. The
607 ancient mariner sails again: transposition of the human Hsmar1 element by a
608 reconstructed transposase and activities of the SETMAR protein on
609 transposon ends. Mol Cell Biol. 2007;27(12):4589-600.
610 27. Goodwin KD, He H, Imasaki T, Lee SH, Georgiadis MM. Crystal
611 structure of the human Hsmar1-derived transposase domain in the DNA
612 repair enzyme Metnase. Biochemistry. 2010;49(27):5705-13.
613 28. Atkinson H, Chalmers R. Delivering the goods: viral and non-viral gene
614 therapy systems and the inherent limits on cargo DNA and internal
615 sequences. Genetica. 2010;138(5):485-98.
616 29. Frokjaer-Jensen C, Davis MW, Sarov M, Taylor J, Flibotte S, LaBella
617 M, et al. Random and targeted transgene insertion in Caenorhabditis elegans
618 using a modified Mos1 transposon. Nat Methods. 2014;11(5):529-34.
619 30. Gayle S, Pan Y, Landrette S, Xu T. piggyBac insertional mutagenesis
620 screen identifies a role for nuclear RHOA in human ES cell differentiation.
621 Stem Cell Reports. 2015;4(5):926-38.
622 31. Bouuaert CC, Tellier M, Chalmers R. One to rule them all: A highly
623 conserved motif in mariner transposase controls multiple steps of
624 transposition. Mob Genet Elements. 2014;4(1):e28807.
625

626 **Legends**

627 **Figure 1. Characterization of the constitutive promoters.**

A/ The gene of interest, Hsmar1 in this case, is fused to 3x FLAG tag on its C-terminus and cloned downstream of one of six different promoters (see text for more details) with a ribosome binding site (RBS) present or not. The construct is located between terminator sequences (T) upstream and downstream to avoid read-through transcription. To further control the number of copies, the plasmid backbone is a one-copy, pBACe3.6, or five copies, pGMH491, vector.

B/ The promoter strength of each construct were determined by FACS after cloning an eGFP gene in each vector. The number 1 to 6 corresponds to one of the six different promoters. The single and five-copies vectors are annotated B or I, respectively. The vector without or with a consensus RBS contained a – or +, respectively. For the promoter 1, which is a repeat of the GACT motif, the RBS motif can also corresponds to a GACT motif and is named N (Bp1N and Ip1N). n=3.

C/ Comparison of Hsmar1 expression cloned in the strongest promoters by western blotting using an antibody against the C-terminus of Hsmar1.

D/ Comparison by western blotting of the strongest constitutive promoters against pTac promoter induced with different concentration of IPTG.

Figure 2. Characterization of the modified papillation assay with the wild-type Hsmar1 transposase.

A/ Representative colony of each vector expressing a wild-type Hsmar1 transposase. Whole plate pictures are presented in Supplementary Figure 3.

B/ Quantification of the number of papillae per colony. Average ± standard deviation of the mean of six representative colonies from three independent experiments.

C/ Plot of the average promoter strength (as defined in Figure 1B) versus the average number of papillae per colony (as defined in Figure 2B). As expected from the overproduction inhibition (OPI) mechanism, an inverse power law is observed between the promoter strength and the transposition rate.

Figure 3. Expression of a covalent Hsmar1 transposase affects the expression threshold needed to reach OPI.

A/ Schematic of the Hsmar1 monomer or covalent dimer. The difference between both constructs is the insertion of a stop codon upstream of the linker region in the monomer construct.

B/ Representative colony of each vector expressing either an Hsmar1 monomer or covalent dimer transposase. Whole plate pictures are presented in Supplementary Figure 4.

C/ Quantification of the number of papillae per colony. The expression vectors have been ordered by decreasing number of papillae per colony for the Hsmar1 monomer. Average ± standard deviation of the mean of six representative colonies from three independent experiments.

D/ Plot of the average promoter strength (as defined in Figure 1B) versus the average number of papillae per colony (as defined in Figure 3C).

Figure 4. Characterization by the papillation assay of different Hsmar1 mutants.

A/ Phylogenetic tree of anthropoid primates representing the apparition of mutations in the Hsmar1 domain of SETMAR. All the mutations present in the human SETMAR were tested by papillation assay to determine when deleterious mutations for Hsmar1 transposition appeared.

B/ Representative colony of pMAL-C2X expressing wild-type or mutant Hsmar1 transposase. Whole plate pictures are presented in Supplementary Figure 5.

C/ Different Hsmar1 mutants have been tested in low, optimal and high transposase expression level (Bp1+, Ip1+ and Ip6+, respectively). Representative colony of each papillation plate is shown. The average number of papillae per colony is indicated below the pictures. Whole plate pictures are presented in Supplementary Figure 6.

SF1. Representation of the modified papillation assay.

A/ The Hsmar1 (RC5096), which encodes a promoter-less *lacZ* gene and a kanamycin resistance marker, has been integrated at a transcriptionally silent locus in a lac- E. coli strain. In absence of transposase, the lacZ gene is not expressed therefore the strain produces white colonies on X-gal reporter plates. When the transposase is supplied in trans from a vector, if a transposon integrates into the ORF of a transcribed gene, a lacZ fusion protein will be produced. The cell's descendants will expressed LacZ and therefore will appear as blue papillae on X-gal reporter plates. Black arrow, promoter; open brackets, transposon ends; empty rectangle, transposase gene.

For the mating-out assay, a chloramphenicol resistant derivative of the conjugative plasmid pOX38 is introduced into the reporter strain. Transposition of the reporter into the plasmid is detected by selecting transconjugants after mating with a recipient strain on chloramphenicol and kanamycin.

B/ Example of FACS profile for each constitutive promoter expressing the eGFP gene. Ip0 is the negative control with nothing cloned in the vector.

SF2. Effect of lactose on the modified papillation assay.

A/ Representative colony of the Ip3+ Hsmar1 vector on different concentration of lactose. Whole plate pictures are presented in C/.

B/ Quantification of the number of papillae per colony. Average ± standard deviation of the mean of six representative colonies from three independent experiments.

C/ Whole plate pictures of the representative colony presented in Supplementary Figure 2A.

SF3. Characterization of the modified papillation assay with the wild-type Hsmar1 transposase.

Whole plate pictures of the representative colony presented in Figure 2A.

SF4. Expression of a covalent Hsmar1 transposase affects the expression threshold needed to reach OPI.

Whole plate pictures of the representative colony presented in Figure 3A.

SF5. Characterization by the papillation assay of different Hsmar1 mutants.

Whole plate pictures of the representative colony presented in Figure 4B.

SF6. Characterization by the papillation assay of different Hsmar1 mutants.

Whole plate pictures of the representative colony presented in Figure 4C.

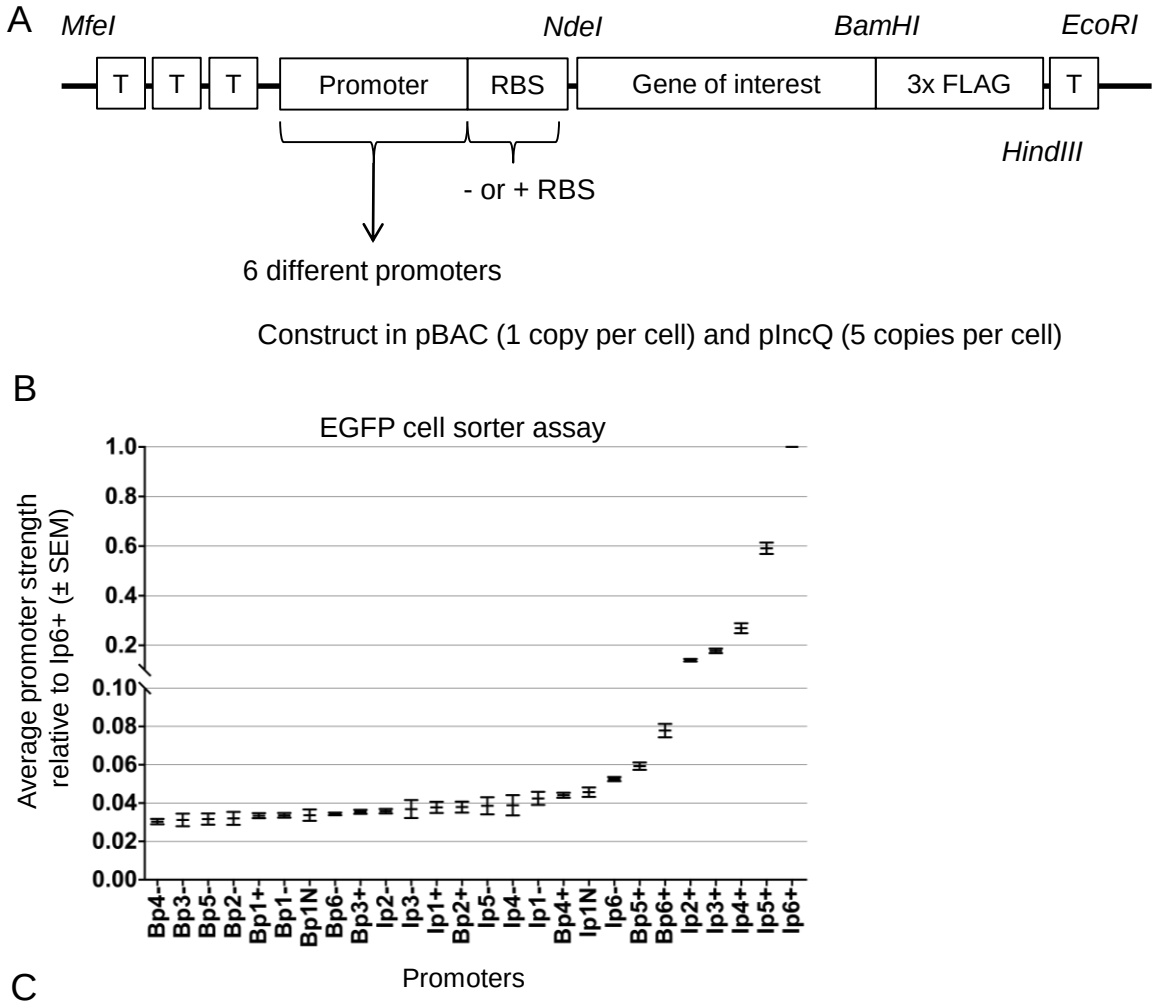

A
MfeI
NdeI
BamHI
EcoRI
T
T
T
Promoter
RBS
Gene of interest
3x FLAG
T
HindIII
- or + RBS
6 different promoters
Construct in pBAC (1 copy per cell) and pIncQ (5 copies per cell)
B
EGFP cell sorter assay
Average promoter strength relative to Ip6+ (± SEM)
1.0
0.8
0.6
0.4
0.2
0.10
0.08
0.06
0.04
0.02
0.00
Bp4-
Bp3-
Bp5-
Bp2-
Bp1+
Bp1-
Bp1N
Bp6-
Bp3+
Ip2-
Ip3-
Ip1+
Bp2+
Ip5-
Ip4-
Ip1-
Bp4+
Ip1N
Ip6-
Bp5+
Bp6+
Ip2+
Ip3+
Ip4+
Ip5+
Ip6+
Promoters
C
Bp5+ Bp6+ Ip2+ Ip3+ Ip4+ Ip5+ Ip6+
- Hsmar1 transposase
D
pTac
Ip1+ Ip2+ Ip3+ Ip4+ Ip5+ Ip6+
0 0.1 1
[IPTG] (mM)
- Hsmar1 transposase

Figure 1

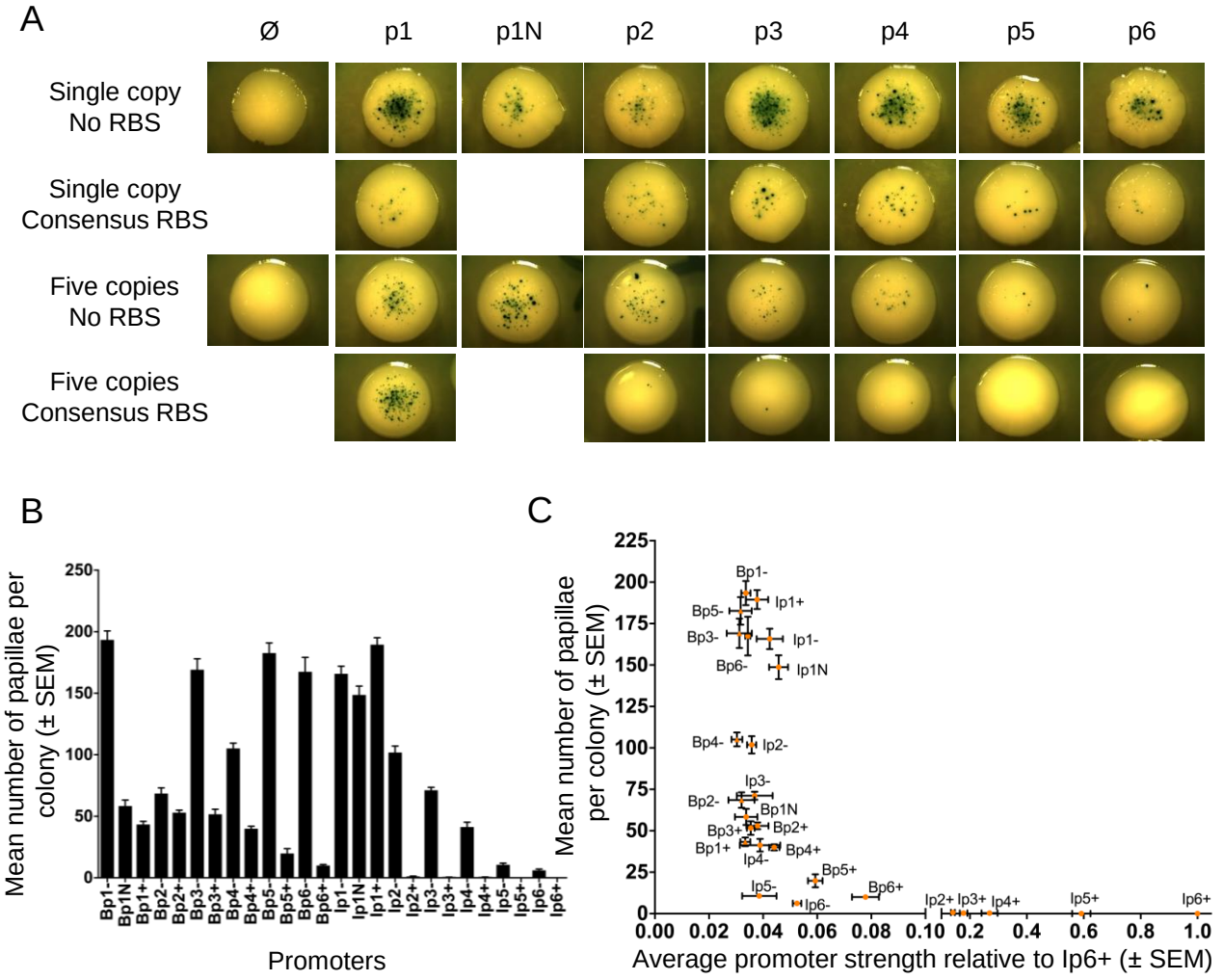
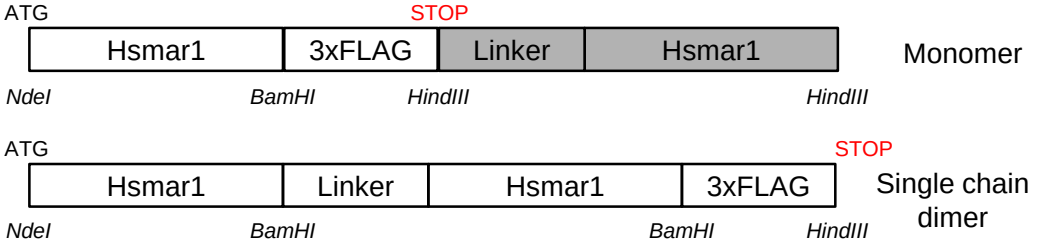

A
Ø
p1
p1N
p2
p3
p4
p5
p6
Single copy
No RBS
Single copy
Consensus RBS
Five copies
No RBS
Five copies
Consensus RBS
B
Mean number of papillae per colony (± SEM)
250
200
150
100
50
0
Bp1-
Bp1N
Bp1+
Bp2-
Bp2+
Bp3-
Bp3+
Bp4-
Bp4+
Bp5-
Bp5+
Bp6-
Bp6+
Ip1-
Ip1N
Ip1+
Ip2-
Ip2+
Ip3-
Ip3+
Ip4-
Ip4+
Ip5-
Ip5+
Ip6-
Ip6+
Promoters
C
Mean number of papillae per colony (± SEM)
225
200
175
150
125
100
75
50
25
0
Bp1-
Ip1+
Bp5-
Bp3-
Ip1-
Bp6-
Ip1N
Bp4-
Ip2-
Ip3-
Bp2-
Bp1N
Bp3+
Bp2+
Bp1+
Bp4+
Ip4-
Bp5+
Ip5-
Bp6+
Ip6-
Ip2+
Ip3+
Ip4+
Ip5+
Ip6+
0.00
0.02
0.04
0.06
0.08
0.10
0.2
0.4
0.6
0.8
1.0
Average promoter strength relative to Ip6+ (± SEM)

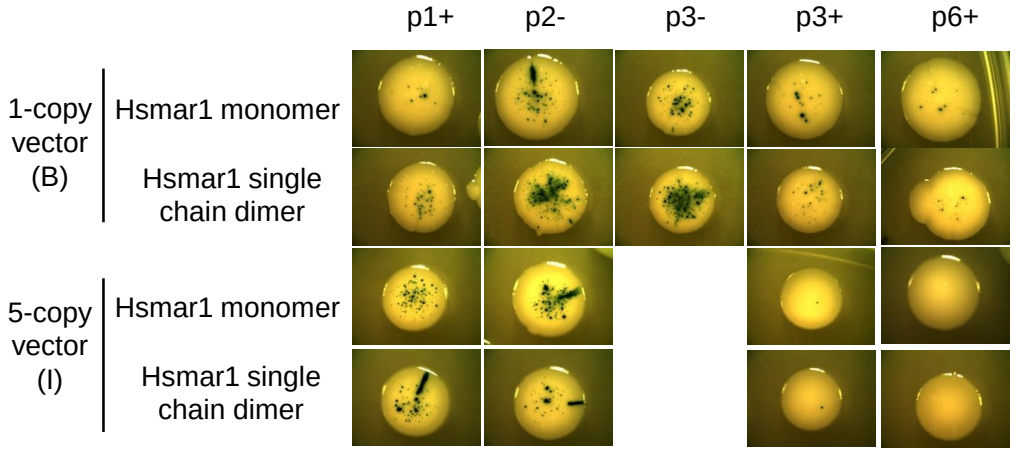
Figure 2

A

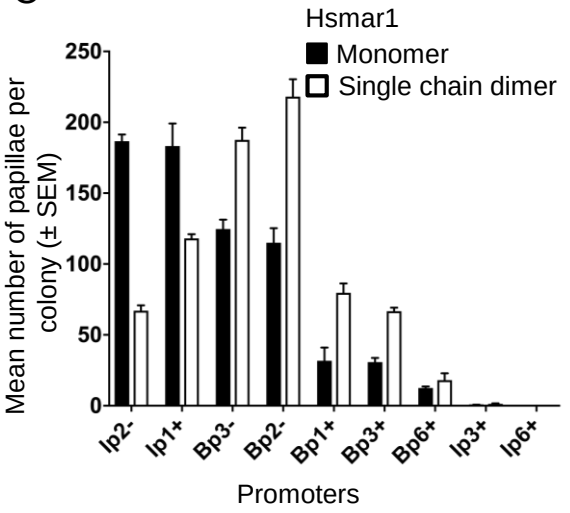
Covalent dimer constructs:



B



C



D

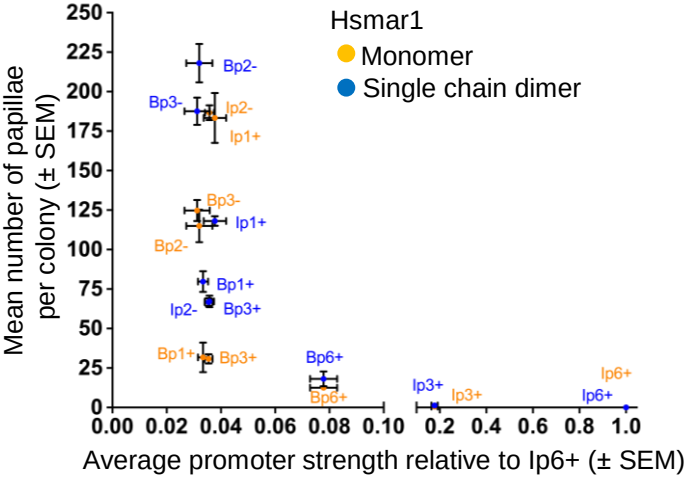
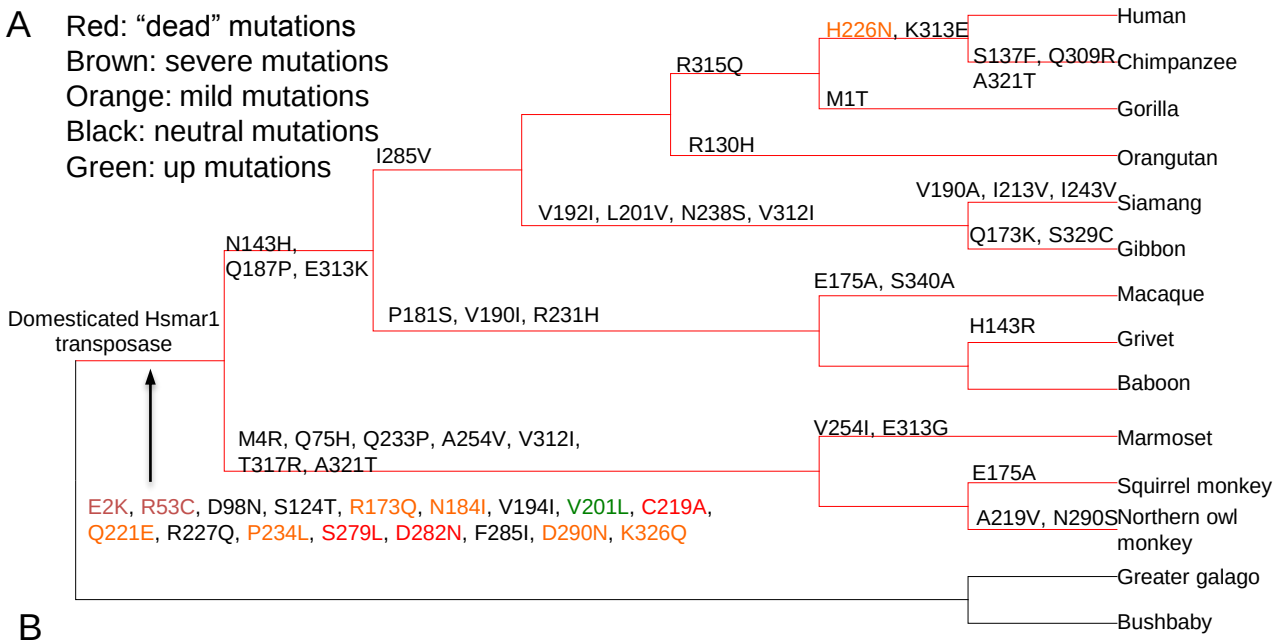
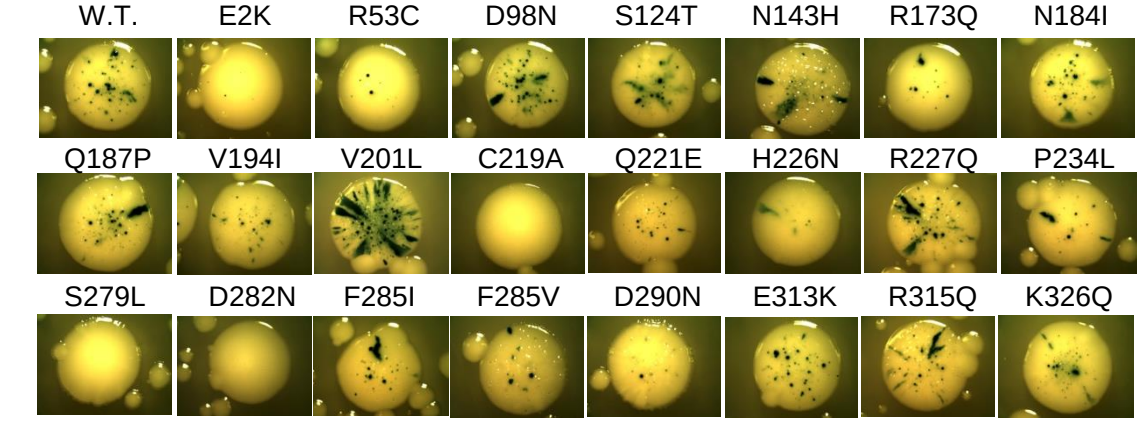


Figure 3


A
Red: “dead” mutations
Brown: severe mutations
Orange: mild mutations
Black: neutral mutations
Green: up mutations
Domesticated Hsmar1 transposase
E2K, R53C, D98N, S124T, R173Q, N184I, V194I, V201L, C219A, Q221E, R227Q, P234L, S279L, D282N, F285I, D290N, K326Q
N143H, Q187P, E313K
M4R, Q75H, Q233P, A254V, V312I, T317R, A321T
I285V
P181S, V190I, R231H
R315Q
H226N, K313E
S137F, Q309R
A321T
M1T
R130H
V192I, L201V, N238S, V312I
V190A, I213V, I243V
Q173K, S329C
E175A, S340A
H143R
V254I, E313G
E175A
A219V, N290S
Human
Chimpanzee
Gorilla
Orangutan
Siamang
Gibbon
Macaque
Grivet
Baboon
Marmoset
Squirrel monkey
Northern owl monkey
Greater galago
Bushbaby


B
W.T.
E2K
R53C
D98N
S124T
N143H
R173Q
N184I
Q187P
V194I
V201L
C219A
Q221E
H226N
R227Q
P234L
S279L
D282N
F285I
F285V
D290N
E313K
R315Q
K326Q

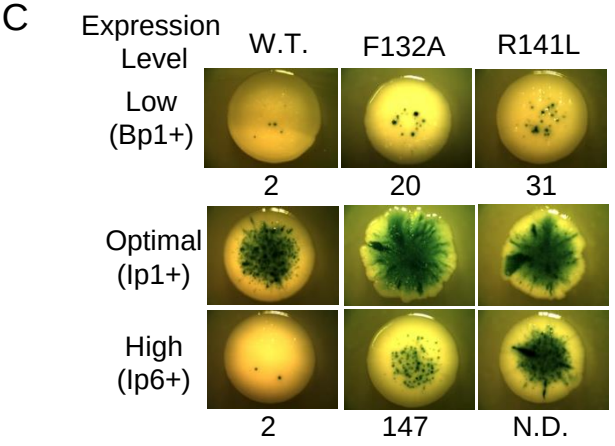

C
Expression Level
W.T.
F132A
R141L
Low (Bp1+)
2
20
31
Optimal (Ip1+)
High (Ip6+)
2
147
N.D.

Figure 4