



## Aberystwyth University

### *Terminator Operon Reporter*

Zampini, Massimiliano; Mur, Luis; Rees Stevens, Pauline; Pachebat, Justin Alexander; Newbold, Charles; Hayes, Finbarr; Kingston-Smith, Alison

*Published in:*  
Scientific Reports

*DOI:*  
[10.1038/srep26572](https://doi.org/10.1038/srep26572)

*Publication date:*  
2016

*Citation for published version (APA):*

Zampini, M., Mur, L., Rees Stevens, P., Pachebat, J. A., Newbold, C., Hayes, F., & Kingston-Smith, A. (2016). Terminator Operon Reporter: Combining a transcription termination switch with reporter technology for improved gene synthesis and synthetic biology applications. *Scientific Reports*, 6, [26572].  
<https://doi.org/10.1038/srep26572>

#### **Document License** CC BY

#### **General rights**

Copyright and moral rights for the publications made accessible in the Aberystwyth Research Portal (the Institutional Repository) are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Aberystwyth Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Aberystwyth Research Portal

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

tel: +44 1970 62 2400  
email: [is@aber.ac.uk](mailto:is@aber.ac.uk)

# SCIENTIFIC REPORTS



OPEN

## Terminator Operon Reporter: combining a transcription termination switch with reporter technology for improved gene synthesis and synthetic biology applications

Received: 02 March 2016

Accepted: 04 May 2016

Published: 25 May 2016

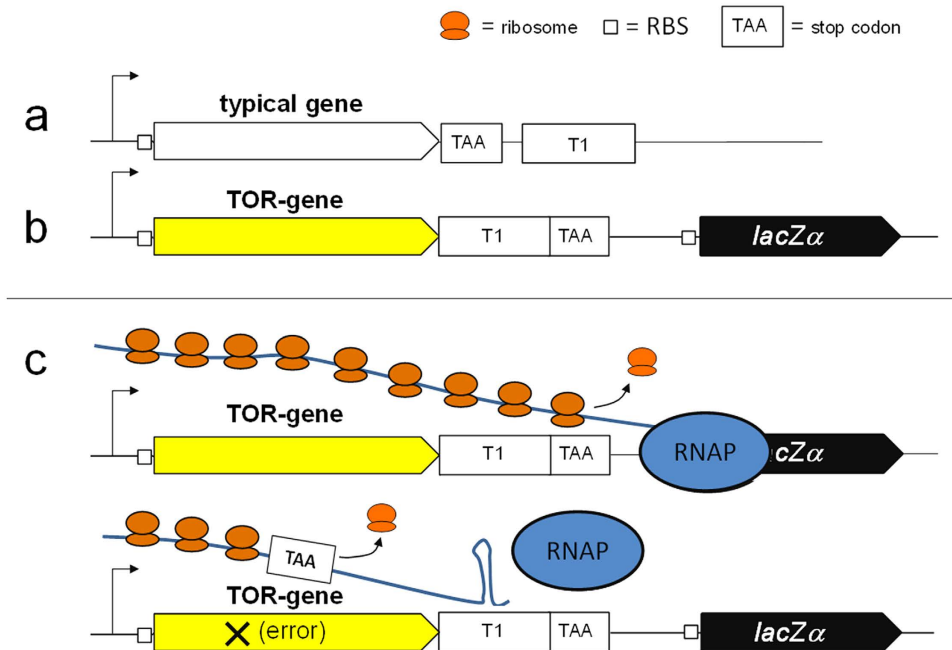
Massimiliano Zampini<sup>1</sup>, Luis A. J. Mur<sup>1</sup>, Pauline Rees Stevens<sup>1</sup>, Justin A. Pachebat<sup>1</sup>, C. James Newbold<sup>1</sup>, Finbarr Hayes<sup>2</sup> & Alison Kingston-Smith<sup>1</sup>

Synthetic biology is characterized by the development of novel and powerful DNA fabrication methods and by the application of engineering principles to biology. The current study describes Terminator Operon Reporter (TOR), a new gene assembly technology based on the conditional activation of a reporter gene in response to sequence errors occurring at the assembly stage of the synthetic element. These errors are monitored by a transcription terminator that is placed between the synthetic gene and reporter gene. Switching of this terminator between active and inactive states dictates the transcription status of the downstream reporter gene to provide a rapid and facile readout of the accuracy of synthetic assembly. Designed specifically and uniquely for the synthesis of protein coding genes in bacteria, TOR allows the rapid and cost-effective fabrication of synthetic constructs by employing oligonucleotides at the most basic purification level (desalted) and without the need for costly and time-consuming post-synthesis correction methods. Thus, TOR streamlines gene assembly approaches, which are central to the future development of synthetic biology.

Synthetic biology is an emerging discipline that aims to apply engineering principles to biology and has the potential to drive a revolution in biotechnology<sup>1</sup>. The discipline has progressed particularly as a result of the implementation of novel and powerful DNA construction techniques that allow for efficient assembly and manipulation of sequences both *in vitro* and *in vivo*<sup>2,3</sup>. At the most basic level among DNA fabrication strategies, gene synthesis<sup>4</sup> is a key pillar supporting synthetic biology.

Despite extraordinary advances in recent years, the chemistry behind gene synthesis is not yet an error-free process. A fundamental source of sequence errors in synthetic genes derives from the presence of upstream mutations, most commonly deletions or insertions, in the oligonucleotide pool used to prepare the assembly<sup>5</sup>. Several methods have been developed to reduce the error rate in the final construct such as the use of highly purified oligonucleotides, the development of functional assays based on intrinsic detectable properties of the assembled gene, or the addition of a reporter tag to the synthetic element<sup>6,7</sup>. However, assays based on the detection of catalytic activities are limited to enzymes, and reporter tags expressed in fusion proteins may fold inappropriately thereby impairing activity<sup>7</sup>. Alternative error removal strategies involve highly controlled assembly conditions, the use of mismatch binding or mismatch cleavage enzymes, or site-directed mutagenesis approaches in the post assembly stages. All of these approaches are costly and time consuming, and sequencing of multiple samples is normally necessary to identify clones harbouring the correct sequence<sup>8</sup>.

<sup>1</sup>Institute of Biological, Environmental and Rural Sciences, Edward Llwyd Building, Aberystwyth University, Aberystwyth SY23 3FG, UK. <sup>2</sup>Faculty of Life Sciences, University of Manchester, Manchester M13 9PL, UK. Correspondence and requests for materials should be addressed to M.Z. (email: M.Zampini@outlook.com) or F.H. (email: Finbarr.Hayes@manchester.ac.uk) or A.K.-S. (email: ahk@aber.ac.uk)



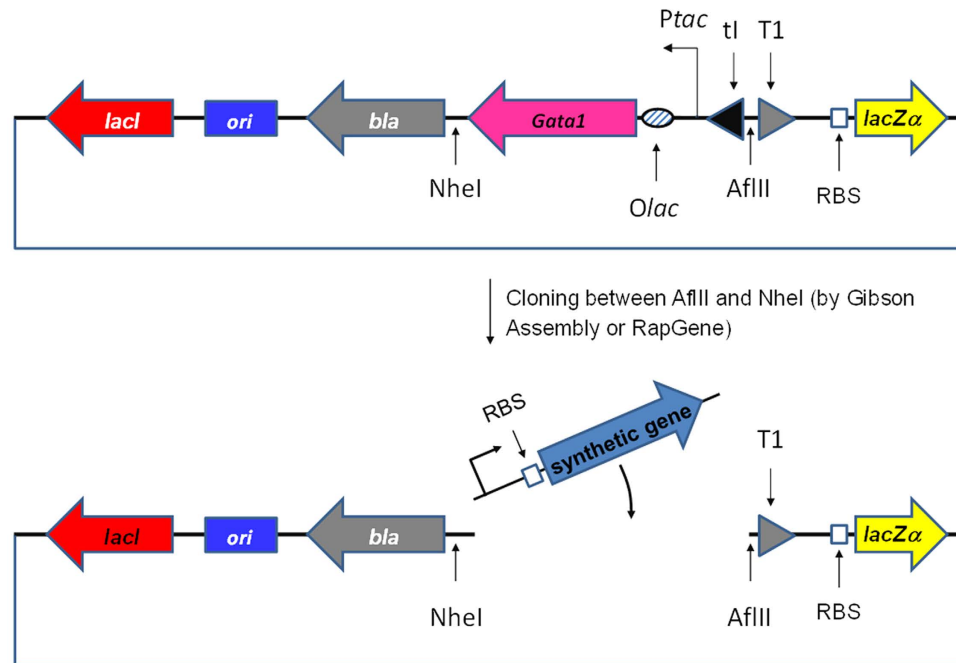
**Figure 1. Principle of TOR.** (a) The coding sequence in a typical protein coding gene is followed by a stop codon (TAA) and then by a transcription terminator (T1). (b) In the TOR platform the transcription terminator is placed upstream of the stop codon and the reporter follows as part of the same operon. (c, top) In TOR T1 is translated as a fusion tag and becomes inactive or is significantly impaired, thus the reporter can be expressed as the second gene of the operon. (c, bottom) In the presence of a premature stop codon arising after a nonsense or a frame-shift mutation, T1 will be active and the reporter is not expressed.

Developed with the aim of overcoming current challenges in gene synthesis, Terminator Operon Reporter (TOR) described here is a novel gene synthesis platform whose principle relies on the established synchrony between transcription and translation in bacteria. Whereas in eukaryotes these two fundamental steps of protein synthesis are spatially and temporally separated, transcription and translation occur simultaneously in prokaryotes. Thus, as RNA polymerase transcribes the 3'-end of bacterial mRNA, ribosomes bind to the 5'-end permitting concurrent translation of the newly synthesized transcripts. This synchrony allows translation to interfere with transcription in a number of well-documented cases. For example, ribosomes bound to bacterial mRNAs during transcription support efficient protein synthesis by promoting mRNA stability in several ways: masking endoribonuclease sites associated with mRNA decay, inhibiting the activity of RNA hairpin dependent transcriptional pausing signals, or hindering transcript access by transcription termination factors<sup>9–12</sup>. Transcriptional attenuation is another example of a ribosome-mediated event commonly found in amino acid biosynthetic operons in bacteria<sup>13</sup>. This type of transcriptional control depends on the ribosome-dependent folding of regulatory RNA stem-loops structures in the 5' untranslated region of the operon<sup>14,15</sup>. Whilst such mechanisms are part of the natural regulatory machinery in prokaryotes, Wright and Hayward demonstrated almost three decades ago how ribosomes can efficiently inhibit the activity of intrinsic terminators located within the coding sequence (CDS) in artificial genetic contexts<sup>16</sup>. The TOR method developed here is based upon this principle elucidated by Wright and Hayward. The strategy combines an in-frame, switchable transcription terminator and a downstream reporter gene to monitor the precision of synthetic gene assembly.

## Results

TOR is a gene assembly strategy in which the sequence of interest is assembled from oligonucleotides and cloned so that the strong T1 transcription terminator<sup>17</sup> is positioned in-frame with the synthetic element (Fig. 1). The termination activity of T1 is abolished or significantly impaired when it is translated in *Escherichia coli*<sup>16</sup> thereby allowing expression of a *lacZα* reporter gene located downstream, as the second gene of a bicistronic operon. Thus, when the sequence of the synthetic gene is correct, read-through occurs of the T1 transcription terminator and the *lacZα* gene is expressed to generate blue colonies on medium containing X-Gal. In contrast, frame-shift mutations lead to a stop codon shortly after the erroneous sequence. As the most common errors in synthetic oligonucleotides are deletions and insertions<sup>5,18</sup>, frame-shift mutations along the CDS are the most frequent. In constructs harbouring frame-shift mutations the T1 terminator is active and *lacZα* is not expressed, so recombinant plasmids produce white (clear) colonies.

TOR was tested with the prokaryotic *aadA1* gene (792 bp; GenBank HQ880267.1) that encodes aminoglycoside adenylyltransferase which confers spectinomycin and streptomycin resistance in *E. coli*, and with the eukaryotic *gfp* gene (717 bp; GenBank L29345.1) that codes for *Aequorea victoria* green fluorescent protein<sup>19</sup>. These genes were selected as we have recently described their assembly by RapGene, a ligation independent approach for gene



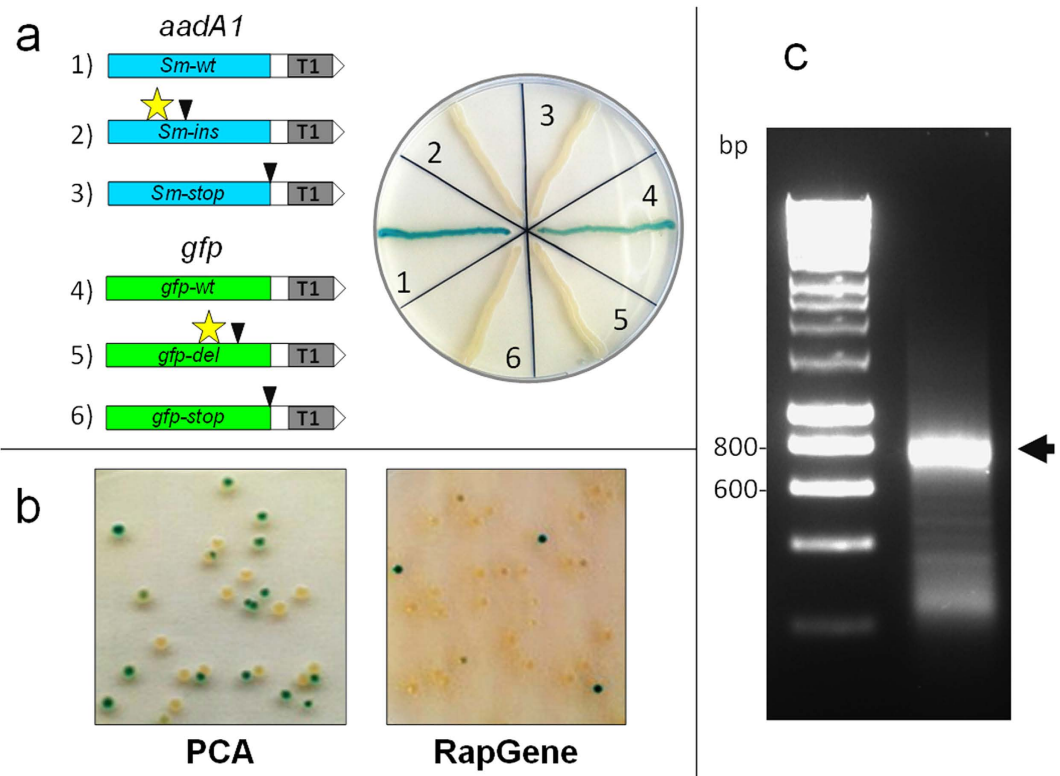
**Figure 2. pRG2 structure (not to scale).** pRG2 is the plasmid employed throughout this study. It features a wild-type *lacI* gene, the toxic and inducible *Gata-1*, the *bla* gene conferring resistance to ampicillin or carbenicillin, a pMB1 type origin of replication, the *lacZ $\alpha$*  gene and two terminators with opposite orientations,  $\lambda$ T1 and T1. Simultaneous digestion of pRG2 with AflIII, NheI and T4 DNA polymerase in the presence of dTTPs results in a plasmid having LIC-compatible ends suitable for RapGene assembly. On the contrary, ends produced by simultaneous NheI and AflIII digestion of the vector in the absence of T4 DNA polymerase were used to create overlap regions in the PCA product suitable for the Gibson Assembly. The reporter gene *lacZ $\alpha$*  is a transcriptional fusion with the synthetic element. The second terminator  $\lambda$ T1 was introduced to reduce leaky expression of *Gata-1* potentially resulting from engineering the original pGATA region upstream of the toxic element. Ori, origin of replication. NheI, restriction site for NheI. AflIII, restriction site for AflIII. *Olac*, *lac* operator. *Ptac*, *tac* promoter.

synthesis in *E. coli*<sup>20</sup>. The cloning reporter vector used throughout this study was pRG2 (Fig. 2). This plasmid is a positive selection vector based on the toxic gene *Gata1*<sup>21</sup> and is derived from the previously described pRG1.0 vector<sup>20</sup>. As TOR was developed as an extension of RapGene<sup>20</sup>, the promoter and the ribosome binding site (RBS) were designed to be part of the pool of oligonucleotides in the assembly (Supplementary Fig. S1), as detailed in Methods. To ensure that frame-shift mutations located at the C-terminus of a synthetic gene also resulted in transcription termination at T1, two additional stop codons were embedded in the vector in reading frames two and three, respectively within an AflIII site and just before the beginning of T1 (Supplementary Fig. S2).

Different *E. coli* promoters and terminators were tested for use in the TOR system. Preliminary successful results were obtained with a variant of the *E. coli*  $P_{recA}$  promoter, referred to here as  $P_{recA}^*$ , and the T1 terminator<sup>17</sup>. The sequence of  $P_{recA}^*$  differs from its wild-type sequence at one nucleotide of the  $-35$  region (TTGATA was changed to TTGACA)<sup>22</sup>. In *recA*<sup>+</sup> strains,  $P_{recA}$  is induced by RecA itself via proteolytic cleavage of the LexA repressor<sup>22</sup>. However the *E. coli* NEB5 $\alpha$  strain used in this study is a *recA1*<sup>-</sup> strain. This mutation has been reported to abolish completely all known RecA functions, including the protease activity towards LexA<sup>23</sup>. Thus,  $P_{recA}$  and  $P_{recA}^*$  are expected to be constitutively repressed in NEB5 $\alpha$ . Two additional non-inducible promoters were also tested to assess the robustness of the TOR platform under different conditions: the *E. coli*  $P_{BAD}$  core promoter comprising only the sequence from  $-35$  to  $-10$  plus external 4 bp on both sides and no additional regulatory elements<sup>24</sup>, and the  $P_{tac}$  promoter<sup>25</sup> comprising the  $-35$  to  $-10$  region with an additional 5 bp upstream and 4 bp downstream from the original pGATA plasmid (Supplementary Fig. S1).

TOR was tested initially on six genes comprising wild-type and nonsense mutants of *aadA1* and *gfp*, all expressed from  $P_{recA}^*$  (Fig. 3a, Supplementary Fig. S3). These genes, containing either wild-type sequences or a stop codon upstream of T1, showed that the TOR platform could easily discriminate between correct and incorrect sequences. In fact, on X-Gal-containing plates, colonies with the correct sequences appeared dark green or blue depending on the incubation time. This colony colour reflects read-through of the T1 terminator and consequent expression of the downstream *lacZ $\alpha$*  gene (Fig. 1). On the contrary, genes possessing a stop codon upstream of T1 resulted in white colonies on the same medium due to transcription termination at T1 and impairment of *lacZ $\alpha$*  expression. Colour development was assessed after a minimum of 30 hours incubation at 37 °C. (Fig. 3a).

In a following approach *gfp* was synthesized by either RapGene or polymerase cycling assembly<sup>26</sup> (PCA) (Fig. 3b). These techniques (Supplementary Fig. S4) were used throughout the study as standard gene synthesis methods to test TOR efficiency. In a first round of TOR, with the synthetic *gfp* expressed from  $P_{recA}^*$ ,



**Figure 3. Application of TOR to *aadA1* and *gfp* genes.** TOR was applied to six test constructs generated from *aadA1* and *gfp* by standard molecular biology techniques, and to *gfp* assembled by either RapGene or PCA assembly. (a) Preliminary assays on test samples. After approx. 36 hours incubation wild-type samples (1 and 4) conferred a dark blue colony colour on CIX, whereas the four mutants carrying a stop codon (black triangles) within or at the end of the CDS showed a white colour on the same medium. Stars correspond to frame-shift deletions or insertions that generate a premature stop codon further downstream. Grey boxes within the constructs correspond to the terminator T1. (b) TOR was applied to a gene synthesis assay for the assembly of *gfp* by PCA or RapGene methods starting from a pool of 24 overlapping oligonucleotides. (c) Electrophoresis of a typical *gfp* PCA product. The band corresponding to the insert (arrow) was gel-extracted and cloned by the Gibson assembly in pRG2.

fifteen randomly chosen TOR-positive candidates obtained by PCA assembly of *gfp* and 13 samples from those derived by RapGene were sequenced (Supplementary Data 1). Analysis of chromatograms showed that 66% of the positive samples identified by PCA and 61% of those obtained by RapGene contained the fully correct *gfp* sequence, while the remaining candidates contained either missense mutations or in-frame deletions (Table 1, Supplementary Data 1). In a second round of TOR, *gfp* was built only by PCA and placed under the control of either  $P_{recA}^*$ ,  $P_{BAD}$  or  $P_{tac}$  promoters (Supplementary Figs. S1 and S5). Plasmids from fifteen transformants, randomly selected among the positive samples, were sequenced for each promoter. Here 93%, 86% and 73% of the positive samples contained a correct *gfp* sequence driven by the  $P_{recA}^*$ ,  $P_{BAD}$  or  $P_{tac}$  promoters, respectively (Table 1). In the case of candidates expressed using  $P_{recA}^*$ , we additionally sequenced 15 TOR-negative transformants (Fig. 4, Supplementary Data 1). As expected, the latter group contained incorrectly assembled synthetic elements, featuring one or more frame-shift mutations followed by a stop codon emerging soon after the mutation site.

**Accuracy and error rates in TOR.** As reported in other studies<sup>18,20</sup> constructs harbouring full-length inserts can be used to determine the error rate. The error rate is defined as the total number of bases mutated (deleted, inserted or substituted) divided by the number of bases sequenced multiplied by 100. This value represents the statistical occurrence of a single mutation every 100 bp, and allows for comparisons between methods that use different strategies and different insert lengths. The error rate in TOR with RapGene was 0.136 from 13 samples sequenced from blue colonies on X-Gal, while the values determined for PCA assembly from a total of 60 positive samples varied between 0.009 and 0.045 (Table 1, column 6), with an average of 0.024. This latter value corresponds to 0.24 errors per kb, or equivalently to 24% sequences containing at least an error, for a 1 kb gene. Thus, with 76% correct sequences, less than two transformants per TOR cloning need to be sequenced to identify a correct synthetic gene of 1 kb.

The accuracy of gene synthesis in TOR was also determined as the ratio between the transformants harbouring a fully correct *gfp* gene divided by the total number of TOR-positive samples sequenced (Table 1, column 3). For the 714-bp *gfp* gene synthesized by PCA the accuracy was 79.5% from four separate trials revealing that values

1		2	3	4	5	6
		Number of samples sequenced	Number and % of samples with fully correct <i>gfp</i> CDS	Number and % of samples with deletions/insertions in <i>gfp</i> CDS	Number and % of samples with substitutions in <i>gfp</i> CDS	Error rate
1 <sup>st</sup> round	1 <sup>st</sup> PCA assembly ( $P_{recA}^*$ )	15	10 (10/15 = 66%)	2 (13%)	2 (13%)	0.022
	1 <sup>st</sup> RapGene assembly ( $P_{recA}^*$ )	13	8 (8/13 = 61%)	3 (23%)	2 (15%)	0.136
2 <sup>nd</sup> round	2 <sup>nd</sup> PCA assembly ( $P_{recA}^*$ )	15	14 (93%)	1 (6%)	1 (6%)	0.009
	2 <sup>nd</sup> PCA assembly ( $P_{BAD}$ )	15	13 (86%)	2 (13%)	0 (0%)	0.045
	2 <sup>nd</sup> PCA assembly ( $P_{tac}$ )	15	11 (73%)	2 (13%)	3 (20%)	0.020

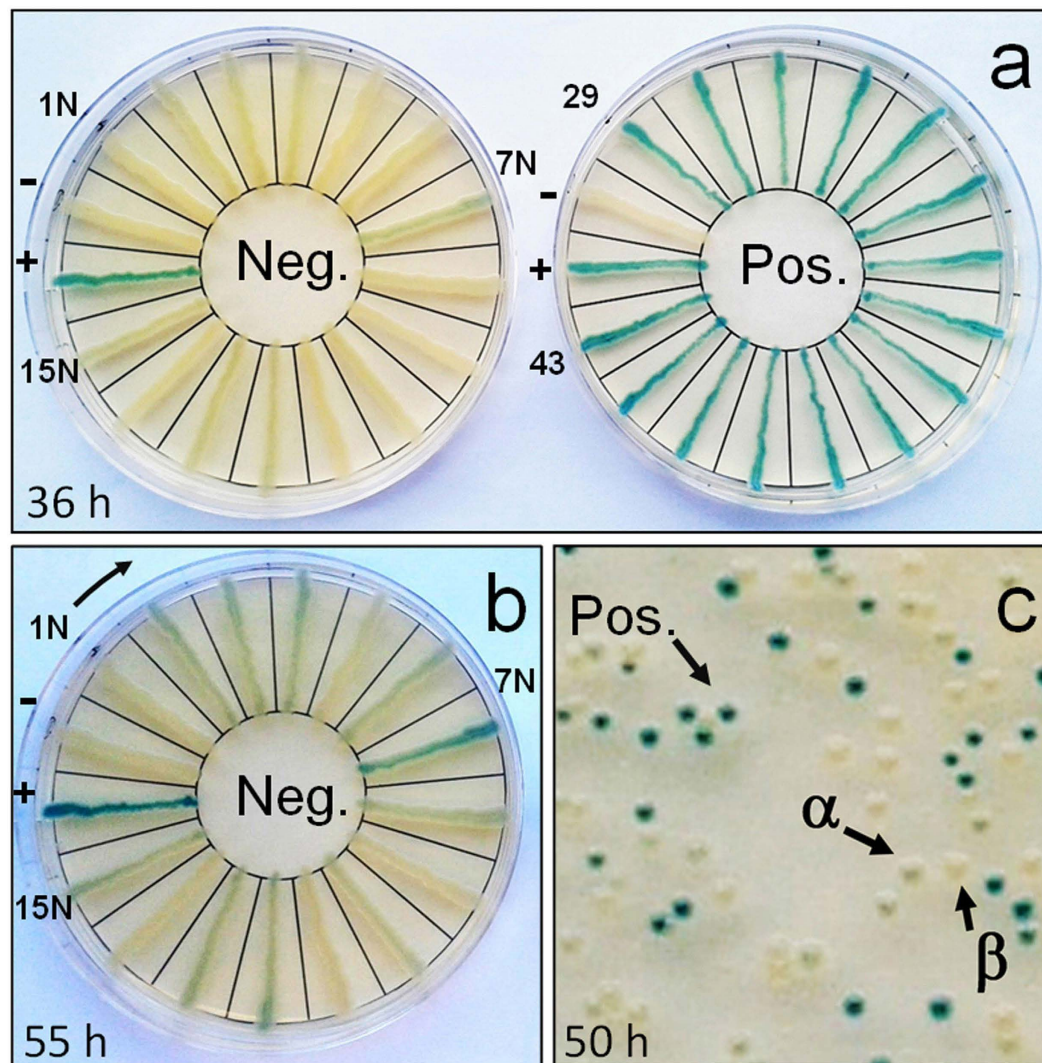
**Table 1. Efficiency of TOR gene synthesis for the assembly of *gfp* through PCA or RapGene methods.** A total of 73 TOR-positive samples derived from one round of RapGene and four rounds of PCA assembly (in the latter case using either  $P_{recA}^*$ ,  $P_{tac}$  or  $P_{BAD}$ ) were sequence verified.

determined by both methods were similar. Thus, the efficiency and reliability of TOR synthesis for genes of ~1 kb is very competitive with market standards, as less than two TOR-positive transformants need to be sequenced to identify the correct sample (<https://www.thermofisher.com/uk/en/home/life-science/cloning/gene-synthesis/gene-strings-dna-fragments.html>). Extrapolating the observed error rate (0.024) to longer synthetic genes, the predicted value of 0.48 errors per 2 kb genes would allow identification of correct constructs of this length by sequencing as few as two transformants per cloning assay. Crucially, only desalted oligonucleotides were used here and no post synthesis DNA correction methods were required for the assembly method. Without TOR selection, the probability of producing an intact *gfp* gene using desalted, 60 nt oligonucleotides is expected<sup>26,27</sup> to be 0.93<sup>22</sup> = 20%, rather than the 76–79.5% observed with TOR. Here, 0.93 is the sequence accuracy estimated for the oligonucleotides used in this work (see below and Supplementary Fig. S6), and 22 is the number of oligonucleotides necessary for *gfp* assembly. This value is comparable with that determined experimentally, both in a previous study<sup>20</sup> (~15%) and here: ~35% colonies per TOR cloning are blue of which 79.5% are correct resulting in an accuracy of 27%.

**False positives in TOR.** The analysis of sequence data from the PCA and RapGene samples revealed that false positives were present in each of the five independent assays for *gfp* assembly (Table 1). Approximately 2% of the constructs contained a single missense mutation, and an additional 2% showed in-frame deletions or insertions within the CDS. TOR cannot eliminate these types of false positives which are however very rare. We noticed that a potential second class of false positives could be observed after prolonged growth at 37 °C. In fact, after extended incubation times several colonies among the negative samples developed a light-green colony colour (Fig. 4c). To investigate this observation further, the fifteen negative samples previously sequenced were studied in more detail (Fig. 4a,b). These defective *gfp* genes were all expressed from the  $P_{recA}^*$  promoter. Eleven samples showed a light-green colony colour when incubated on CIX medium (LB agar plus carbenicillin [120 µg/ml], IPTG [1 mM] and X-Gal [0.0001%]) for up to 55 hours at 37 °C. In particular, sample *gfp-ins*, which sequence analysis revealed to possess a single insertion mutation after position 183 of the coding sequence, developed a strong green colour after 48 hours of incubation at 37 °C (Supplementary Fig. S7). However, a difference between *gfp-ins* and wild-type *gfp* was still evident at this timepoint, especially when cells expressing these mutants were observed as single colonies. All of the sequenced negative samples possessed at least one stop codon within *gfp*, but colony colour development differed significantly. Therefore, we excluded that colony darkening was due to inefficiency of T1 as a transcription terminator in this genetic context. To test whether the preceding false positives expressed from the  $P_{recA}^*$  promoter behaved similarly when expressed from alternative promoters, a plate assay was performed using the same *aadA1* and *gfp* mutants described in Fig. 3a, but expressed under the control of  $P_{BAD}$  or  $P_{tac}$  (Supplementary Fig. S8). Only *gfp-ins* expressed from  $P_{tac}$  produced an indistinguishable blue colour compared to wild-type *gfp*. Therefore, although high level of expression of the reporter may produce false-positives such as *gfp-ins*, sequencing data for  $P_{tac}$  transformants showed that the latter is still a reliable and “fast” TOR promoter (73% accuracy; Table 1, Supplementary Data 1), with false positives from the second class being negligible (accounting for only one out of the 73 TOR-positive samples sequenced, i.e. ~1%). While we cannot currently explain the light-green colour of some of the negative samples, we suggest that translation re-initiation at cryptic ribosome binding sites within the CDS may account for this effect. Ribosome recruitment at internal sites has been recently reported to be more common than expected, (not only for heterologous gene expression), as a mechanism that can affect the rate of transcription more than the codon usage<sup>28</sup>.

In TOR, a cryptic RBS-ATG/GTG pair located on frame two or three could also be involved in this effect. In fact, as a result of a frame-shift, a short sequence upstream of the mutation site on either of these frames will become in-frame with the CDS.

An appropriate codon optimization of the synthetic gene - common practice in gene synthesis - could avoid the emergence of a cryptic RBS, and likely exclude the possibility of weak expression of the reporter gene in TOR-negative samples.



**Figure 4. Analysis of TOR-negative samples obtained by TOR.** (a,b) Comparison between the phenotypes of 15 PCA-positive versus 15 PCA-negative samples (both under  $P_{recA}^*$ ). Eleven out of fifteen samples gradually darkened after prolonged incubation. A clear difference was still observable between positive and negative samples, but *gfp-ins* (sample N7) darkened more than the others. (c) TOR-negative samples showed two different colony colours upon cloning, green ( $\alpha$ ) and yellow ( $\beta$ ). Neg., TOR-negative samples; Pos., TOR-positive samples. 1N, 7N, 15N, samples 1, 7, 15 from the TOR-negative samples.

**Accuracy of the oligonucleotides used in this study.** Long deletions were occasionally observed in constructs following PCA cloning (Supplementary Data 1). As the inserts amplified by PCA were always gel extracted, these mutations were likely the result of *in vivo* DNA rearrangement events. To quantify this effect, the *NheI*-*AfeII* fragment bearing the wild-type *gfp* sequence construct in pRG2 under the control of  $P_{recA}^*$  was amplified and cloned back into the same vector by the Gibson procedure (Supplementary Fig. S6). Transformants were plated both on LB plus Carbenicillin and on CIX plates. Thirty colonies that appeared white on the former were chosen randomly and screened by PCR to identify possible DNA deletions. No rearrangements were observed (Supplementary Fig. S6a). Transformants plated on CIX mostly developed into colonies with a blue colour, but 10.6% (15/141) were white (Supplementary Fig. S6b). Ten samples randomly chosen among the latter were sequenced and mutations were identified only in the sequence of the reverse oligonucleotide used for PCR amplification (Supplementary Fig. S6c). This assay confirmed that the background level of *in vivo* rearrangements observed in this study was very low, and also provides an accurate quantification of the sequence accuracy associated with the oligonucleotides used in this study (approx. 93%). Typical accuracy levels for desalted 60 nt oligonucleotides may be as low as 60%. However, a much higher percentage of oligonucleotides (93% determined here) can be considered correct for selected applications such as TOR as several types of 5' deletions appear to be tolerated by this assembly method, as oligonucleotide gaps are filled *in vivo*.

**Colony colour development in TOR depends on codon usage.** Whereas colony colour development for positive samples was a function of the incubation time in TOR, the colour conferred by the *gfp* gene on CIX

plates was intrinsically less intense than that produced by *aadA1* (sample *Sm-wt*, Fig. 3a; Supplementary Fig. 8a). Wright and Hayward noted that when a terminator was in frame with the CDS, events expected to slow down the movement of ribosomes on the mRNA, e.g., binding of fusidic acid, allowed restoration of termination<sup>16</sup>. Based on this principle, a *gfp* mutant containing several rare codons was predicted to confer a brighter green colour on CIX plates compared to wild-type *gfp*. In contrast, the co-expression of rare tRNAs from plasmid pRARE2 (see Methods) in a strain expressing a *gfp* with rare codons, was expected to rescue the dark colony colour. These predictions were confirmed (Supplementary Figs. S9 and S10). As rare codons are expected to slow down the movement of the ribosomes on the transcript, fewer ribosomes will be able to populate the downstream mRNA region. Thus here, a weak but detectable transcription termination at T1 would be observed, and colony colour intensity would decrease even in the absence of a frame-shift mutation.

## Discussion

In this work we have described and characterized TOR, a screening strategy that employs a universal reporter assay for the quick and cost-effective identification of correctly assembled synthetic protein-coding genes. The approach relies on the conditional activation of a transcription terminator in response to sequence errors originating at the assembly step. The effect of TOR-based selection was significant as almost 80% of approx. 1 kb sequences selected by TOR were fully correct. Thus, no more than two transformants per cloning experiment need to be sequenced with this method to identify a correctly assembled 1 kb synthetic gene. Crucially, the technology allows the quick synthesis of standard length genes at affordable prices, as only desalted oligonucleotides are used to achieve the same level of accuracy of more complex industrial procedures. In addition, the predicted yields of approximately 50% for a 2 kb assembly would allow the commercially viable synthesis of even longer synthetic genes in a one-pot assay, especially when the TOR platform was coupled with current methods used in gene synthesis, e.g., highly purified oligonucleotides and enzymatic post-assembly error correction methods. Fragments of 2–3 kb are typically close to the upper limit for one-pot gene synthesis approaches, as yields for these long constructs may drop easily below values of commercial interest.

The results achieved by TOR are relevant for several reasons. First, as low-quality desalted oligonucleotides are not a limiting factor in achieving high accuracy by the TOR method, this technology is an efficient alternative option to third party gene synthesis services, especially if a simple in-house oligonucleotide synthesizer was available. Use of TOR in industrial settings will be an additional tool to improve yields and reduce gene synthesis costs by helping to overcome current limitations associated with the maximum gene length that is commercially viable. Moreover, the need for more complex techniques are avoided when accuracy levels similar to those obtained by the TOR method are sufficient.

Second, the potential use of reporter genes made up of entire metabolic pathways will further enhance the sensitivity and flexibility of the method allowing, for example, detection of very low expression levels for the reporter, and so also the synthesis of toxic genes. In fact, as TOR relies on the expression of the synthetic gene of interest, this method would not be suitable for the assembly of toxic proteins, although, as shown here for  $P_{BAD}$  driven expression, TOR is also effective under low expression levels that could be compatible with toxic effects. Moreover, the use of an antibiotic resistance marker as a reporter gene, will allow only positive clones to grow on appropriate antibiotic-containing plates. This is important when longer synthetic genes are assembled, as the lower yields expected with extended constructs could require too many colonies to be grown on the same plate.

Third, due to its intrinsic ability to select for full-length CDS mutants derived from random mutagenesis, the TOR platform could be explored for additional applications including directed evolution or phage-display-like approaches.

Thus, the TOR gene synthesis method described here is a fast, cost effective and accurate gene assembly platform for improved applications in synthetic biology.

## Methods

**Plasmids and strains.** Plasmid pRG2 (Fig. 2) is a derivative of pRG1.0<sup>19</sup>. It features a wild-type *lacI* promoter and a novel reporter system which consists of two elements introduced just upstream of the toxic *Gata1* gene: (i) the efficient *E. coli* T1 transcription terminator<sup>17</sup> designed to be in-frame with the synthetic gene upon cloning, followed by (ii) the *lacZ $\alpha$*  gene and native RBS from the cloning vector pUC19<sup>29</sup>. The *lacZ $\alpha$*  gene encodes the  $\alpha$ -fragment derived from  $\beta$ -galactosidase that is employed in blue-white screening for cloning in *E. coli*<sup>30</sup>. Since pRG2 was designed as an extension of the RapGene assembly method<sup>20</sup>, the  $P_{sm}$  promoter located upstream of the *NheI* restriction site in pRG1.0 is absent in plasmid pRG2. In fact the assembly of a >600 bp synthetic element by RapGene was associated with the generation of insert-free re-ligated vectors<sup>20</sup> and the presence of a plasmid-encoded promoter in pRG2 could easily produce false positive results due to the expression of the *lacZ $\alpha$*  from  $P_{sm}$ . Thus, instead of  $P_{sm}$ , variable promoters were introduced in pRG2 as part of the synthetic insert generated from several oligonucleotides. Plasmid pRG2 was produced by Gibson assembly from four DNA fragments derived from pGATA, pUC19 (*lacZ $\alpha$* ) and a six oligonucleotide generated PCA fragment ( $\lambda$ TI-T1). The T1 terminator sequence<sup>17</sup> was derived from the Registry of Standard Biological Parts ([http://parts.igem.org/Main\\_Page](http://parts.igem.org/Main_Page)) (part:BBa\_B0010) and was placed downstream of the  $\lambda$ I terminator<sup>31</sup>. This latter was oriented in the direction of *Gata1* to avoid the possibility that genetic manipulation upstream of *Gata1* (in the T1-*lacZ $\alpha$*  region) might introduce a promoter-like sequence inducing constitutive expression of the toxic GATA-1. *E. coli* NEB5 $\alpha$  (New England Biolabs) was used throughout the study. Plasmid pRARE2 was purified from *E. coli* strain Rosetta™ 2(DE3) (Merck Millipore) and used subsequently in the NEB5 $\alpha$  strain.

**TOR accuracy determination.** When determining the error rate in TOR, full-length inserts were arbitrarily defined as those samples showing deletions/insertions smaller than 15 bp, i.e., 25% of the average size oligonucleotide used in the assembly.



**DNA assembly and mutagenic strategies.** Recombinant DNA was generated mainly by the Gibson Assembly Cloning Kit (New England Biolabs) following the manufacturer's instructions. For PCA assembly of *gfp* constructs, plasmid pRG2 was simultaneously digested with NheI and AflII for 1–2 hours at 37 °C. After spin column purification the vector was used directly for cloning. For RapGene assembly of *gfp* constructs, T4 DNA polymerase and 1 mM dTTPs were also added to the digestion mixture and cloning was performed as previously described<sup>19</sup>. PCA assays were performed with Phusion DNA polymerase (New England Biolabs) using oligonucleotides that were 60 nt in length on average. One microliter of each oligonucleotide (from a 100 μM stock) was added to 1 ml 10 mM Tris, 50 mM NaCl (pH 8.0). One microliter of this mix was used as template in a 50 μl PCR reaction containing 0.25 μM external primers, Phusion DNA polymerase (1U/50 μl), 200 μM each dNTPs and HF Phusion buffer. Amplification primers were the same external oligonucleotides used in the assembly. PCA amplification of *gfp* required up to 50 cycles of PCR at constant annealing temperature (65 °C). However, in the first round of PCA amplification (Table 1) annealing temperatures varied from 70 °C to 50 °C in steps of 5 °C per 10 cycles. The amplified PCR product was purified by gel extraction before cloning with the Gibson Assembly Cloning Kit using *E. coli* NEB5α chemically competent cells.

**CIX plates.** Screening assays were performed on CIX plates (LB agar plus carbenicillin [120 μg/ml], IPTG [1 mM] and X-Gal [0.0001%]). Plates were stored at 5 °C and used within 10 days of preparation. Colour development on CIX plates (Supplementary Fig. S11) required unexpectedly the presence of IPTG, despite pRG2 lacking any *lac* operator-like sequences. The very low expression of *lacZα* observed here – typically about 10 Miller units for the strongest constructs (Supplementary Fig. S10) – coupled with LacI overexpression could possibly account for the observed silencing effect<sup>32</sup> (i.e. all site-specific DNA-binding proteins bind non-specifically although with low affinity to any DNA sequence<sup>33</sup>). The low levels of *lacZα* activity may also suggest that T1 is still partly active in termination, even when fully translated.

**Oligonucleotides and sequencing.** Sigma-Aldrich was the supplier for oligonucleotides (Supplementary Table S1) which were obtained at the lowest purification level (desalted), re-suspended in sterile 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA, and stored at –80 °C. Sequencing reactions were performed at GATC-Biotech (Germany).

**ONPG assay.** Bacteria were grown on LB agar plus Carbenicillin [120 μg/ml], IPTG [1 mM] (without X-Gal) for 48 hours. From these plates, suspensions were prepared (OD<sub>600</sub> = 0.2) in 3 ml NaCl 0.9%. Then 100 μl of these samples were mixed with 900 μl Z-buffer [60 mM Na<sub>2</sub>HPO<sub>4</sub>, 40 mM NaH<sub>2</sub>PO<sub>4</sub>, 10 mM KCl, 1 mM MgSO<sub>4</sub>, plus β-mercaptoethanol (50 mM, freshly added)] in 15 ml Falcon tubes. After adding 20 μl of CHCl<sub>3</sub>, 20 μl 0.1% SDS (sodium dodecyl sulphate) and 200 μl ONPG (ortho-Nitrophenyl-β-galactoside) [4 mg/ml in water], samples were vortexed and incubated at 28 °C in a waterbath for 55 hours, before stopping the reaction with 500 μl 1 M Na<sub>2</sub>CO<sub>3</sub>. OD<sub>600</sub> was measured for all the samples at 420 nm and 550 nm, against a blank made of the same buffer plus 200 μl of NaCl 0.9%. The negative controls consisted of NEB5α transformed with either pRG2 or pRG2 plus pRARE2. Miller Units (MU) were determined using the formula  $1000 \times (OD_{420} - [1.75 \times OD_{550}]) / (t \times v \times OD_{600})$  (t = minutes, v = millilitres). MU from the controls were subtracted from MU of the samples.

## References

1. Pleiss, J. The promise of synthetic biology. *Appl Microbiol Biotechnol* **73**, 735–739 (2006).
2. Czar, M. J., Anderson, J. C., Bader, J. S. & Peccoud, J. Gene synthesis demystified. *Trends Biotechnol* **27**, 63–72 (2009).
3. Chao, R., Yuan, Y. & Zhao, H. Recent advances in DNA assembly technologies. *FEMS Yeast Res* **15**, 1–9 (2015).
4. Hughes, R. A., Miklos, A. E. & Ellington, A. D. Gene synthesis: methods and applications. *Methods Enzymol* **498**, 277–309 (2011).
5. Ma, S., Saaem, I. & Tian, J. Error correction in gene synthesis technology. *Trends Biotechnol* **30**, 147–154 (2012).
6. Stemmer, W. P., Cramer, A., Ha, K. D., Brennan, T. M. & Heyneker, H. L. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164**, 49–53 (1995).
7. Denis-Quanquin, S. *et al.* Protein expression from synthetic genes: selection of clones using GFP. *J Biotechnol* **131**, 223–230 (2007).
8. Genscript. Gene Synthesis Handbook. Second Edition (2014) Available at: <http://www.genscript.com/gsf/files/techfiles/GSH-2edApr2014.pdf>. (Accessed: 25th February 2016).
9. Deana, A. & Belasco, J. G. Lost in translation: the influence of ribosomes on bacterial mRNA decay. *Genes Dev* **19**, 2526–2533 (2005).
10. Landick, R., Carey, J. & Yanofsky, C. Translation activates the paused transcription complex and restores transcription of the *trp* operon leader region. *Proc Natl Acad Sci USA* **82**, 4663–4667 (1985).
11. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci USA* **97**, 7090–7095 (2000).
12. Richardson, J. P. Loading Rho to terminate transcription. *Cell* **114**, 157–159 (2003).
13. Yanofsky, C. Transcription attenuation: once viewed as a novel regulatory strategy. *J Bacteriol* **182**, 1–8 (2000).
14. Santangelo, T. J. & Artsimovitch, I. Termination and antitermination: RNA polymerase runs a stop sign. *Nat Rev Microbiol* **9**, 319–329 (2011).
15. Gusarov, I. & Nudler, E. The mechanism of intrinsic transcription termination. *Mol Cell* **3**, 495–504 (1999).
16. Wright, J. J. & Hayward, R. S. Transcriptional termination at a fully rho-independent site in *Escherichia coli* is prevented by uninterrupted translation of the nascent RNA. *EMBO J* **6**, 1115–1119 (1987).
17. Orosz, A., Boros, I. & Venetianer, P. Analysis of the complex transcription termination region of the *Escherichia coli rrmB* gene. *Eur. J Biochem* **201**, 653–659 (1991).
18. Gibson, D. G. Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. *Nucleic Acids Res* **37**, 6984–6990 (2009).
19. Inouye, S. & Tsuji, F. I. *Aequorea* green fluorescent protein. Expression of the gene and fluorescence characteristics of the recombinant protein. *FEBS Lett* **341**, 277–280 (1994).
20. Zampini, M. *et al.* RapGene: a fast and accurate strategy for synthetic gene assembly in *Escherichia coli*. *Sci Rep* **5**, 11302 (2015).
21. Trudel, P., Provost, S., Massie, B., Chartrand, P. & Wall, L. pGATA: a positive selection vector based on the toxicity of the transcription factor GATA-1 to bacteria. *Biotechniques* **20**, 684–693 (1996).
22. Brent, R. & Ptashne, M. Mechanism of action of the *lexA* gene product. *Proc Natl Acad Sci USA* **78**, 4204–4208 (1981).

23. Kawashima, H., Horii, T., Ogawa, T. & Ogawa, H. Functional domains of *Escherichia coli* RecA protein deduced from the mutational sites in the gene. *Mol Gen Genet* **193**, 288–292 (1984).
24. Schleif, R. Regulation of the L-arabinose operon of *Escherichia coli*. *Trends Genet* **16**, 559–565 (2000).
25. de Boer, H. A., Comstock, L. J. & Vasser, M. The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. *Proc Natl Acad Sci USA* **80**, 21–25 (1983).
26. TerMaat, J. R., Pienaar, E., Whitney, S. E., Mamedov, T. G. & Subramanian, A. Gene synthesis by integrated polymerase chain assembly and PCR amplification using a high-speed thermocycler. *J Microbiol Methods* **79**, 295–300 (2009).
27. Smith, H. O., Hutchison, C. A., 3rd, Pfannkoch, C. & Venter, J. C. Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc Natl Acad Sci USA* **100**, 15440–15445 (2003).
28. Li, G. W., Oh, E. & Weissman, J. S. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484**, 538–41 (2012).
29. Yanisch-Perron, C., Vieira, J. & Messing, J. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103–119 (1985).
30. Langley, K. E., Villarejo, M. R., Fowler, A. V., Zamenhof, P. J. & Zabin, I. Molecular basis of beta-galactosidase alpha-complementation. *Proc Natl Acad Sci USA* **72**, 1254–1257 (1975).
31. Bermúdez-Cruz, R. M., Chamberlin, M. J. & Montañez, C. NusA is involved in transcriptional termination on lambda *tI*. *Biochimie* **81**, 757–764 (1999).
32. Rodionov, O., Lobočka, M. & Yarmolinsky, M. Silencing of genes flanking the P1 plasmid centromere. *Science* **283**, 546–549 (1999).
33. de la Rosa, M. A., Koslover, E. F., Mulligan, P. J. & Spakowitz, A. J. Dynamic strategies for target-site search by DNA-binding proteins. *Biophys J* **98**, 2943–2953 (2010).

## Acknowledgements

This work was supported by the Biotechnology and Biological Sciences Research Council [grant numbers BBS/E/W/10964A to A.K.-S., M.Z., C.J.N. and P.R.S].

## Author Contributions

M.Z. invented the TOR system, designed the study, and performed the research. M.Z., F.H., A.K.-S., L.A.J.M., J.A.P. and C.J.N. analysed the data. P.R.S. assisted in PCA, cloning assays, and sequencing. L.A.J.M. and J.A.P. suggested possible applications to random mutagenesis assays. F.H. and M.Z. devised the alternative initiation of translation model. M.Z., F.H. and A.K.-S. wrote the manuscript. L.A.J.M., J.A.P. and C.J.N. contributed to editing of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Zampini, M. *et al.* Terminator Operon Reporter: combining a transcription termination switch with reporter technology for improved gene synthesis and synthetic biology applications. *Sci. Rep.* **6**, 26572; doi: 10.1038/srep26572 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>