

Systematic Evaluation of Site-Specific Recombinant Gene Expression for Programmable Mammalian Cell Engineering

Nuša Pristovšek,^{†,ID} Saranya Nallapareddy,[†] Lise Marie Grav,[†] Hooman Hefzi,^{‡,§} Nathan E. Lewis,^{‡,§} Peter Røgbjerg,^{†,ID} Henning Gram Hansen,[†] Gyun Min Lee,^{*,†,II} Mikael Rørdam Andersen,[†] and Helene Fastrup Kildegaard^{*,†,ID}

[†]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, 2800 Kgs. Lyngby, Denmark

[‡]Departments of Pediatrics and Bioengineering, University of California, San Diego, La Jolla, California 92093, United States

[§]The Novo Nordisk Foundation Center for Biosustainability at the University of California, San Diego School of Medicine, La Jolla, California 92093, United States

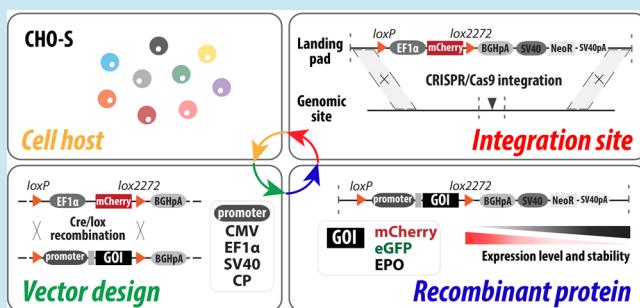
^{II}Department of Biological Sciences, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea

^{*}Department of Biotechnology and Biomedicine, Technical University of Denmark, Søltofts Plads, Building 223, 2800 Kgs. Lyngby, Denmark

Supporting Information

ABSTRACT: Many branches of biology depend on stable and predictable recombinant gene expression, which has been achieved in recent years through targeted integration of the recombinant gene into defined integration sites. However, transcriptional levels of recombinant genes in characterized integration sites are controlled by multiple components of the integrated expression cassette. Lack of readily available tools has inhibited meaningful experimental investigation of the interplay between the integration site and the expression cassette components. Here we show in a systematic manner how multiple components contribute to final net expression of recombinant genes in a characterized integration site. We develop a CRISPR/Cas9-based toolbox for construction of mammalian cell lines with targeted integration of a landing pad, containing a recombinant gene under defined 5' proximal regulatory elements. Generated site-specific recombinant cell lines can be used in a streamlined recombinase-mediated cassette exchange for fast screening of different expression cassettes. Using the developed toolbox, we show that different 5' proximal regulatory elements generate distinct and robust recombinant gene expression patterns in defined integration sites of CHO cells with a wide range of transcriptional outputs. This approach facilitates the generation of user-defined and product-specific gene expression patterns for programmable mammalian cell engineering.

KEYWORDS: gene expression, mammalian cell engineering, site-specific integration, CRISPR, synthetic biology, Chinese hamster ovary cells



The ability to precisely control and maintain the level of recombinant gene expression is crucial to mammalian cell engineering. To ensure that recombinant genes are stably expressed in the host cells, genomic integration is needed, which is typically still conducted in a random or semi-random manner (*i.e.*, insertion of cassettes into actively transcribed regions of the genome). When random or semi-random integration is applied, unpredictable and heterogeneous transgene expression is observed due to variability in gene copy numbers and the nature of the chromosomal integration site (so-called position effect).^{1,2} Accordingly, laborious screening for the desired expression levels is required.^{3,4} In contrast, methods applying targeted integration into predefined loci pave the way toward rational design of cells with predictable transgene expression.⁵

The use of the type II clustered regularly interspaced short palindromic repeats(CRISPR)/ CRISPR-associated protein 9(Cas9) system has enabled more efficient and precise genome editing.⁶ Harnessing the homology-directed repair (HDR) mechanism, error-free targeted integration of an exogenous DNA donor template into the genome can be achieved following a double-strand break induced by Cas9.⁷ HDR-mediated gene targeting has been used extensively in mammalian cell engineering with applications ranging from biopharmaceutical production, gene therapy to gene function studies.⁸ What these

Received: October 27, 2018

Published: February 26, 2019



fields have in common is that they are all struggling to overcome the obstacles of variable transgene expression, clonal variation, and oncogenic transformation (chromosomal rearrangements).

Therapeutic glycoproteins are traditionally produced using Chinese hamster ovary (CHO) cells, where the final yield and stability of the protein production are important aspects of cost-effective production pipelines.⁹ To achieve predictable and desirable protein production, protocols for CRISPR/Cas9-based site-specific integration^{5,10–12} and recombination-based site-specific integration^{13–18} have been developed, sometimes used in combination,^{19–21} that ensure controlled genome integration into predefined loci. Recombination-based systems require a prior establishment of master cell lines containing a marker gene with recombinase recognition sites, referred to as landing pads, allowing for recombinase-mediated cassette exchange (RMCE). CRISPR/Cas9-based targeted integration and RMCE result in predictable and desirable recombinant gene expression patterns when employed into previously characterized, safe genomic loci (*i.e.*, safe harbors). Safe harbors are genomic sites that are (a) positioned in open chromatin areas that are generally transcriptionally active, (b) immune to mono- or biallelic disruption of the target locus upon transgene integration, and (c) not receptive to transcriptional perturbations of endogenous genes upon transgene integration.²² There are a few widely used safe harbors provided for engineering in the human genome, such as AAVS1, CCR5, and ROSA26.²³ However, there is a continually growing repository of integration sites supplied for mammalian cell engineering,^{12,20,24} albeit their proper validation as safe harbors is needed.

Apart from the chromosomal context, the nature of the expression construct used to generate recombinant cell lines has an equally important impact on protein yield and stability.²⁵ A range of vector elements can modulate transgene expression, ranging from chromatin-modifying elements (such as matrix attachment regions (MARs), insulators, ubiquitously acting chromatin opening elements (UCOEs), and stabilizing antirepressors (STARs)) to natural and engineered promoters.²⁶ Site-specific expression levels with diverse vector elements have been studied in different fields of biology.^{14,27,28} When applying targeted integration into predefined loci, these studies have confirmed that the expression level of the transgene is highly dependent on the interplay between the promoter and the target genomic locus. Since methods that would predict the interaction of the integrated cassette with the genome are not available, both safe harbor sites and performance of different expression cassettes in such chromosomal contexts can currently only be verified experimentally.

To the best of our knowledge, no available toolbox has been published so far that would allow streamlined evaluation of site-specific stable gene expression levels under different expression cassettes. To this end, we present a modular toolbox for systematic screening of vector elements in predefined integration sites, using the CHO-S cell line as a model. By studying different expression cassette designs in newly discovered safe harbors using targeted integration, we demonstrate that desired (*e.g.* high and/or stable) expression levels in defined chromosomal loci are restricted to a specific cassette design. We explore four different layers contributing to the final net expression level, namely, clonal background, chromosomal context, vector design, and type of recombinant protein by studying five parental clones, three integration sites, six 5' proximal regulatory elements, and three recombinant proteins, respectively. This renders a total of 270 possible

combinations between the four features, and we here examine 66 (25%) of those combinations experimentally. A more detailed understanding of the performance of vector elements when embedded in a defined chromosomal context is not only of general interest to the mammalian cell engineering community, but is also critical knowledge when designing suitable expression cassettes for engineering more complex expression patterns. Furthermore, since there is a need to further diversify elements for gene expression, this tool allows for identification of a wide range of robust gene expression patterns providing means toward programming mammalian cells with dynamic functionalities.

RESULTS

Carefully studied integration sites constitute preferable genome safe harbors for recombinant gene expression. Therefore, this study set out to identify putative safe harbors in CHO-S cells to use them for systematic characterization of their capacity for stable recombinant gene expression, and thereby provide them to the mammalian cell engineering community. Next, to systematically assess recombinant gene expression patterns in these newly identified integration sites, we developed a modular CRISPR/Cas9-based toolbox. Specifically, this toolbox was designed to inspect the interactions between the four layers influencing the final level of recombinant gene expression: cell host, the integration site, the vector construct, and the type of recombinant protein. We inspected 66 randomly selected combinations between the four layers, and named the samples after the integrated components from each layer.

Identification of Putative Safe Harbor Integration Sites in the CHO-S Genome. To increase the number of well-characterized genomic loci for stable and predictable recombinant expression, we chose a combined strategy of identifying sites; the sites were identified based on either experimental expression stability or *in silico* approach. Site A derives from a low-copy, relatively high-producing cell line with randomly integrated Rituximab, generated using our in-house developed clone screening platform (for detailed description of Rituximab clone generation, refer to *Supporting Methods*).²⁹ This cell line exhibited stable and relatively high Rituximab productivity over two months of culturing during which no gain or loss in Rituximab gene copy number was observed (*Supporting Figure S1*). Proprietary TLA technology was used to identify the genomic locus of integration in this cell line,³⁰ unveiling site A. Full integration of intact Rituximab expression cassette in site A was independently confirmed with targeted MiSeq deep-sequencing (*Supporting Figure S2*). The results here validated that the integration event did not disrupt the host genome in the proximity of integration (*i.e.* chromosomal rearrangement) and therefore this locus can be used for targeted reverse engineering. On the other hand, sites T2 and T9 were selected based on transcriptomic data of neighboring genes and model predictions and were previously used as integration sites in our studies.^{21,31} By analyzing published RNA-seq data sets from CHO-S cell line³² and identifying regions with consistently high expression profiles across 40 samples, sites T2 and T9 were chosen as integration site candidates. We hypothesized that these two sites would be good integration sites owing to consistent chromatin accessibility for transcription. Site T2 is positioned in-between two highly expressed genes and site T9 is in proximity of a highly expressed gene *Rrm1*, that is metabolically essential for DNA synthesis,³³ as additionally confirmed by the genome-scale CHO metabolic model.³² The percentile ranks from expression

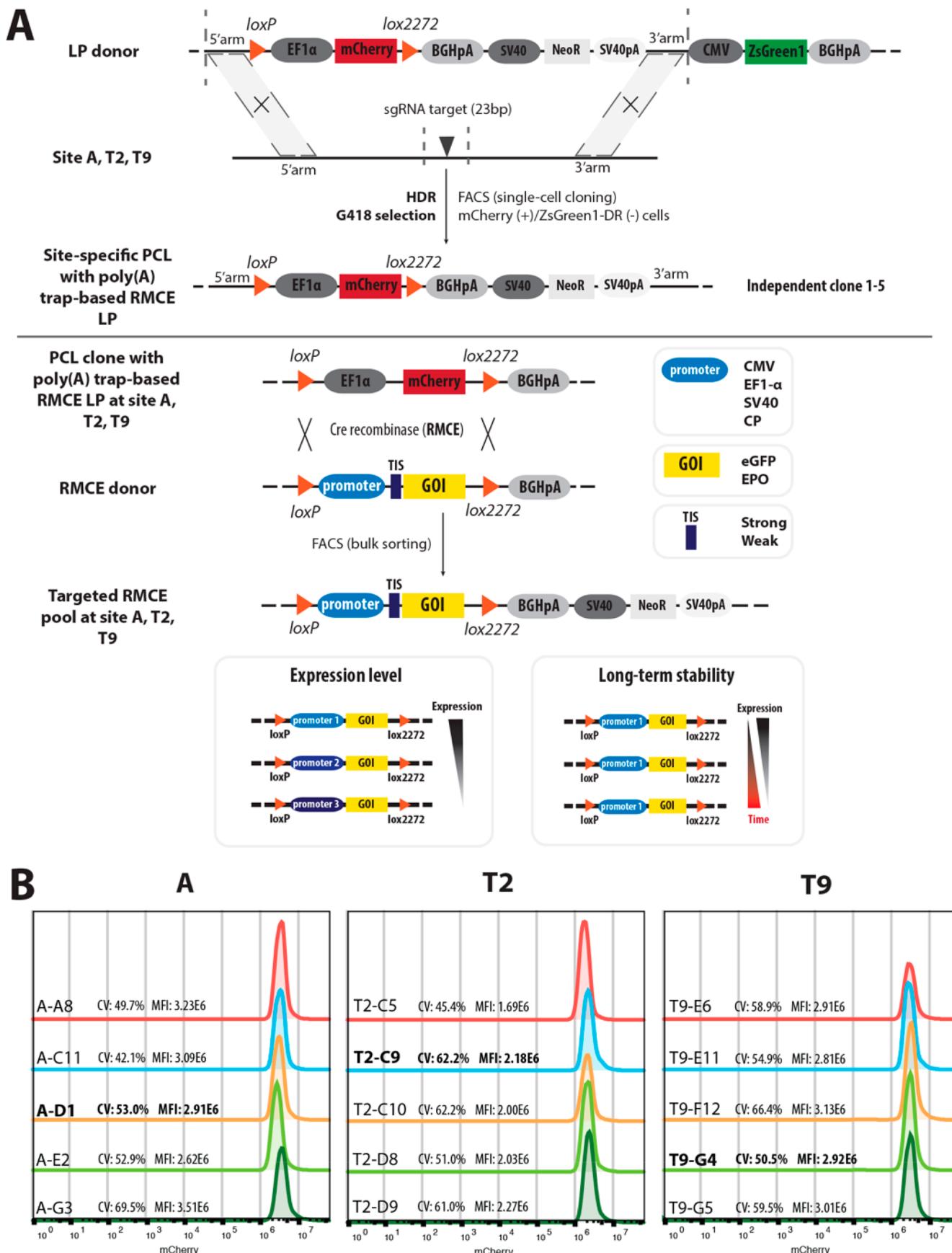


Figure 1. continued

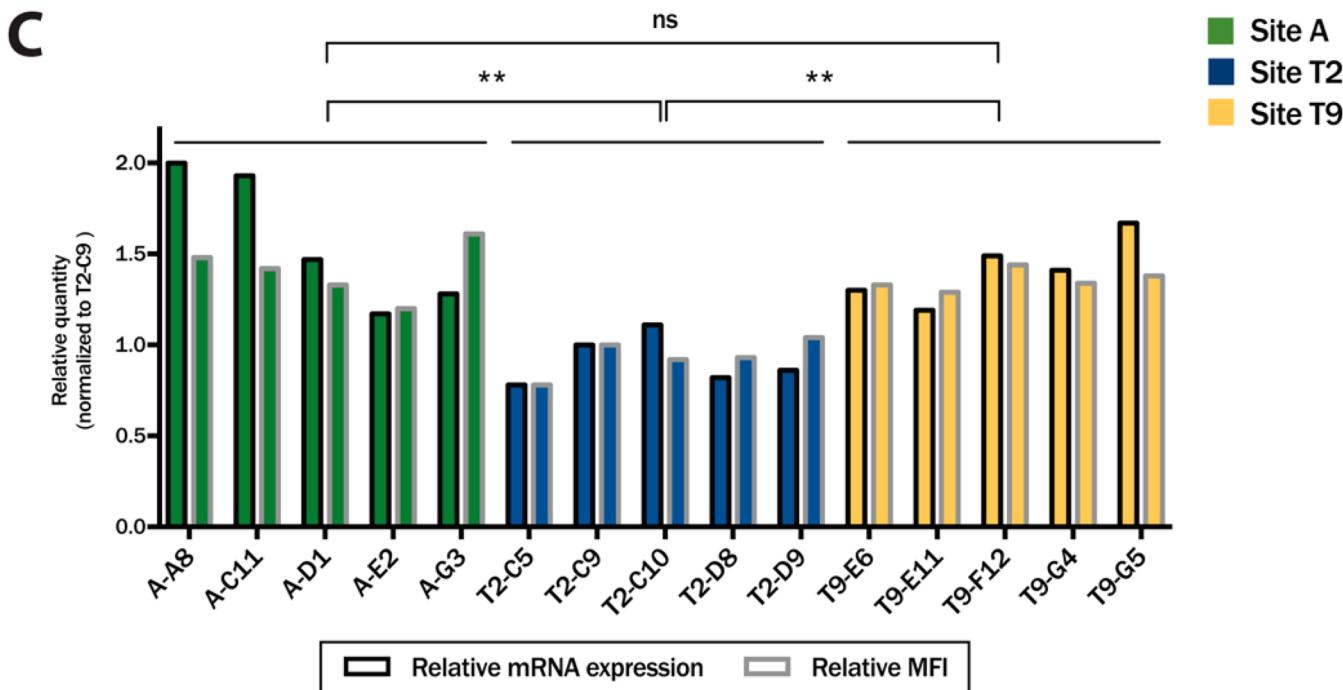


Figure 1. Generation of mCherry platform cell lines with a poly(A) trap-based recombinase-mediated cassette exchange (RMCE) landing pad in putative safe harbors yields homogeneous mCherry expression patterns across different parental clones. (A) Five independent mCherry platform cell lines (PCLs) in each of the three putative safe harbors (A, T2, and T9) were generated with CRISPR/Cas9-based targeted integration *via* homology-directed repair (HDR) using site-specific landing pad (LP) donor and sgRNA plasmid, together with a Cas9 vector. LP donor vector consisted of mCherry and selection marker (NeoR) expression cassettes, flanked by 750 bp long 5' and 3' arms, homologous to the site of interest, and ZsGreen1-DR expression cassette outside of homology arms to exclude random integrants. After HDR-based targeted integration, a 14-day G418-based selection was performed to enrich for cells with a successful integration of the LP donor. This was followed by single-cell cloning of double mCherry positive/ZsGreen1-DR negative cells by FACS to isolate monoclonal cell lines with targeted integration of the LP donor. The LP donors contained *loxP* and *lox2272* recombination sites, flanking EF1 α -mCherry, providing a poly(A)-containing LP (*i.e.* poly(A) trapping) for recombinase-mediated cassette exchange (RMCE). EF1 α -mCherry was upon cotransfection of Cre recombinase and RMCE donor exchanged with a range of different expression cassettes. RMCE donors consisted of different genes of interest (GOIs) under different 5' proximal regulatory elements (*i.e.* promoters and Kozak/translation initiation sequences (TIS)). Two weeks post-recombination, successfully recombined cells were detected by FACS and bulk sorted (mCherry negative cells for EPO and double mCherry negative/GFP positive cells for eGFP) to isolate an RMCE-generated cell pool comprising targeted integrants. To assess site-specific expression strength resulting from specific regulatory elements, the pools were further analyzed by measuring the GOI's mRNA (by RT-qPCR) and protein levels (by FACS or biolayer interferometry for eGFP and EPO, respectively). Further, eGFP expression profiles at the mRNA and protein level were measured during two month long-term cultivation to assess the expression stability of both, the integration site and the promoter of interest. (B) mCherry fluorescence levels of the generated mCherry-encoding PCLs ($n = 5$ clones/site) were measured in mid to late exponential phase in batch cultures. Coefficient of variation (CV) and median fluorescence intensity (MFI) are shown for each clone, by analyzing 20 000–40 000 cells per clone. (C) mCherry mRNA levels were also assessed in mid to late exponential phase in batch cultures in the generated mCherry-encoding PCLs. mCherry mRNA levels were normalized to T2-C9 clone. Statistical significance between relative mRNA expression levels of five clones with targeted site A, T2, and T9 was calculated using unpaired *t* test. A linear regression-based correlation analysis of relative mCherry MFIs from Figure 1b (normalized to mCherry MFI of T2-C9 clone) and relative mCherry mRNA expression levels was conducted.

levels of T2 and T9 neighboring genes across the 40 RNA-Seq samples are provided in the Supporting Figure S3, demonstrating that T2 and T9 neighboring genes are consistently among the highest expressed genes in this cell line. The integration loci information is available in the Supporting Table S1, together with the corresponding Chinese hamster-derived chromosome numbers and the surrounding epigenetic marks. The features of these three integration sites can in the future help toward defining the general characteristics of safe harbors in CHO cells.

Modular Toolbox for Generation of mCherry-Encoding Platform Cell Lines with a Poly(A) Trap-Based RMCE Landing Pad. Having identified three sites of interest, a modular CRISPR/Cas9-based toolbox was developed allowing for systematic quantification of recombinant gene expression levels in these sites modulated by different 5' proximal regulatory elements (Figure 1A). In the first part of the toolbox, platform cell lines (PCLs) with landing pads flanked by unique

lox sites are generated for each integration site of interest *via* HDR and G418 selection, as previously described.²¹ The PCLs are designed with EF1 α -mCherry inside the landing pad and poly(A) outside of the landing pad (poly(A) trapping). By employing a panel of parental clones for each integration site, a number of individual PCLs are generated. This allows for investigating the effects on gene expression originating from the cell host and the integration site, respectively. In the second part of the toolbox, generated PCLs with the poly(A) trap can be subjected to RMCE to replace EF1 α -mCherry with desired 5' proximal regulatory elements followed by the gene of interest (GOI). The RMCE facilitates fast integration of the desired expression cassettes into the same PCL, allowing for investigation of the effects originating from the vector elements and type of recombinant protein integrated.

In conclusion, this platform was designed to interrogate site-specific expression levels of any recombinant protein with any 5'

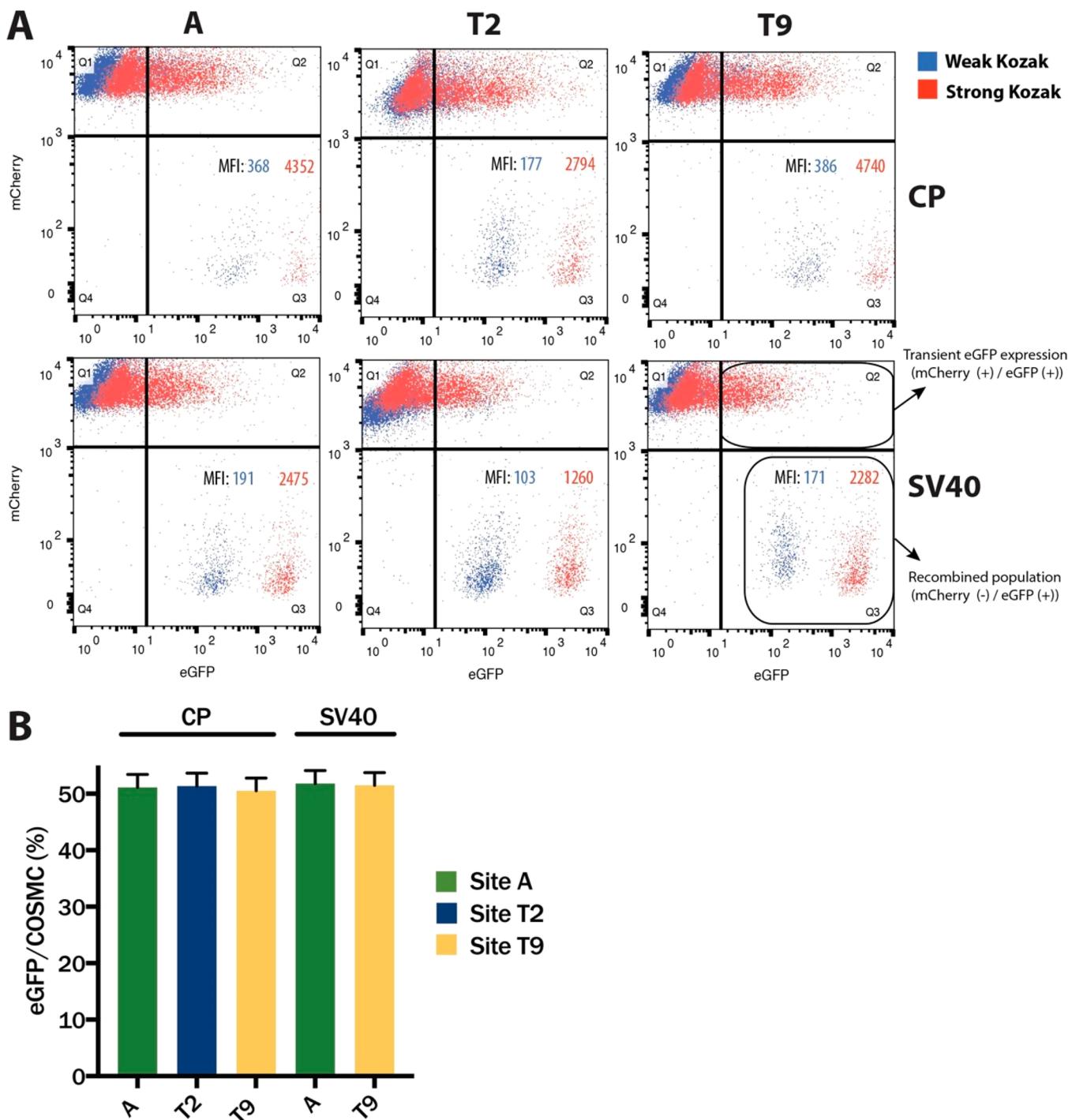


Figure 2. Proof-of-principle RMCE experiment with different eGFP constructs confirms functionality of poly(A) trap and generates single-copy populations with tunable expression patterns. (A) Three representative PCLs (A-D1, T2-C9 and T9-G4) were randomly selected for site A, T2, and T9, respectively, and subjected to RMCE with four different eGFP constructs with composite promoter (CP) or SV40 promoter, preceded by either a weak or strong Kozak sequence. One week post-RMCE, mCherry and eGFP fluorescence were measured by FACS (100 000–200 000 cells) with MFIs of successfully recombined populations (*i.e.* mCherry negative/eGFP positive cells) reported. (B) Recombined populations with eGFP under CP or SV40 promoter and strong Kozak only were bulk sorted and their eGFP copy number measured using digital PCR (dPCR). Using COSMC as an internal one-copy control gene, obtained eGFP/COSMC reads (%) are reported for five bulk-sorted populations. The error bars represent 95% confidence intervals, averaged between the technical replicates ($n = 2$).

proximal regulatory element in a controlled and streamlined manner. Moreover, due to its design, it can be used for direct comparison and benchmarking of different expression contexts and different integration sites of interest.

Validation of Generated mCherry-Encoding Platform

Cell Lines. Following the workflow of the toolbox, we generated PCLs with mCherry integrated in site A, T2, and T9. To account for clonal variation originating from the heterogeneous polyclonal pool, five independent PCLs were generated for

each of the three sites. 5' and 3' junction PCRs were performed for confirmation of specific integration into the three integration loci (Supporting Figure S4). The frequency of clones with correctly targeted integration events post-G418 selection ranged from 21% for site A, 25% for site T2 to 30% for site T9. Following junction PCRs, the PCLs were confirmed to harbor a single copy of the landing pad expression cassette by copy number analysis using qPCR (Supporting Figure S5). Furthermore, $\geq 98\%$ of the total cell population was mCherry positive for all 15 PCLs, indicating that the mCherry expression is relatively homogeneous, despite the generated PCLs being in different growth phases (Supporting Figure S6).

Using the established CRISPR/Cas9-based platform, we successfully generated a panel of PCLs with the mCherry expression cassette in the three selected integration sites. These PCLs were selected as cell line models for studying both clonal variation and interactions between integration site and vector regulatory elements for different recombinant proteins.

Minimal Clonal Variation at Defined Integration Sites.

To characterize the effect of the cell host background on site-specific gene expression levels, the 15 generated PCLs (5 PCLs per integration site) were analyzed in batch cultures. Since the expression cassette in the landing pad was identical in all clones (EF1 α -mCherry), we were able to examine the two factors contributing to the differences in expression levels: clonal background and the site of integration.

The five PCLs for each integration site demonstrated different growth profiles (see Supporting Figure S7), probably because they were derived from a genetically heterogeneous cell pool of host cells. Despite different growth profiles, mCherry median fluorescence intensities (MFIs) in the mid exponential phase were uniform between the PCLs of one site as evaluated by the small coefficient of variation (CV) (site A: CV = 10.9%; site T2: CV = 10.9%; site T9: CV = 4.1%) (Figure 1B). When comparing the CVs in mCherry MFI within each PCL, we observed much larger values ranging between 42% and 70%. Large CVs observed in mCherry MFI within the clone indicate the presence of clone-specific fluctuations in mCherry expression on account of transcriptional and translational bursts. However, we observed no differences in clone-specific fluctuations between the three sites, as evaluated by the mean CV of each site. Furthermore, mCherry mRNA expression levels correlated with the MFIs, suggesting that mCherry fluorescence can be used as a proxy for transcription and expression levels ($R^2 = 0.62$, $p = 0.0005$) (Figure 1C). The variation in mCherry mRNA expression levels observed between the five independent PCLs of each site was relatively small, as seen with the MFI values (site A: CV = 24%; site T2: CV = 15%; site T9: CV = 13%). When comparing mCherry mRNA expression levels between the integration sites, we observed significant differences in how the EF1 α promoter responds to the genetic context ($p = 0.0015$ for A vs T2 and $p = 0.0059$ for T2 vs T9). Sites A and T9 yielded approximately 1.5-fold higher mCherry expression levels than site T2 (Figure 1B and 1C), indicating the local chromatin environment and possible neighboring enhancers of site A and T9 provide an increased transcriptional output from EF1 α promoter.

Overall, these results demonstrated that when controlled, targeted integration into putative safe harbor loci is employed, the generated clones display minimal variation in recombinant gene expression levels, despite having different growth profiles. Moreover, it was reconfirmed that the promoter activity is governed by its spatial positioning (*i.e.* position effect) within a

defined genomic landscape. On the basis of these results, one representative PCL for each integration site was randomly selected from the generated panel for the following recombination experiments.

Proof-of-Principle Recombination into Poly(A) Trap Elicited Tunable eGFP Expression Patterns.

To compare expression from different vector elements in different integration sites, we needed to confirm that EF1 α -mCherry can be successfully recombined with another 5' proximal regulatory element-GOI and can be functionally transcribed from the poly(A) trap. To this end, the EF1 α -mCherry cassettes in the landing pads of the three representative PCLs were exchanged with four different eGFP constructs. The eGFP constructs were designed with either a composite promoter (CP; comprised of mCMV enhancer, hEF1 α and 5'UTR HTLV) or the viral SV40 promoter together with a strong or weak Kozak sequence.³¹

One week post-recombination, double mCherry negative/eGFP positive populations were observed, indicating successful recombination and eGFP transcriptional output from the poly(A) trap (Figure 2A), thereby confirming the functionality of our platform. When comparing MFIs of the recombined cells, an approximately 13-fold difference was observed between strong and weak Kozak sequences across all three sites and both promoters (Figure 2A). However, the same constructs recombined in different sites yielded approximately 2-fold higher eGFP fluorescence levels in site A and T9 in comparison to T2, for both promoters with either a weak or strong Kozak sequence. This confirmed the trend observed with the 15 PCLs where site A and T9 gave rise to increased mCherry expression levels with the EF1 α promoter.

Next, we wanted to investigate whether the observed double mCherry negative/eGFP positive populations were a result of specific recombination of mCherry with a single copy of eGFP. Only then, these double mCherry negative/eGFP positive populations could be used for analyzing the effect of integration sites and vector constructs on expression levels. Hence, we bulk-sorted five double mCherry negative/eGFP positive populations with strong Kozak sequences and conducted copy number analysis using digital PCR (dPCR). All bulk-sorted populations displayed a 1:1 target-to-reference ratio, confirming that on average these populations harbored a single copy of eGFP (Figure 2B). A 1:1 ratio was also found in a confirmed single-copy monoclonal RMCE cell line producing EPO, generated in our previous study (Supporting Figure S8).²¹ Furthermore, a control titration experiment was performed where incrementally increasing amounts of gDNA from A-CP-strong Kozak-eGFP bulk-sorted population were spiked into CHO-S wt gDNA. A linear relationship was confirmed between the eGFP spike-ins and the target-to-reference ratio ($R^2 = 0.98$, $p = 0.0002$) (Supporting Figure S9).

Overall, our proof-of-principle experiment confirmed three important objectives, needed for the following systematic evaluation of site-specific expression patterns. First, poly(A) trapping enabled functional output postrecombination. Second, recombined populations were sufficiently homogeneous (*i.e.* single copy) to be reliably used for comparative analysis without the need to establish monoclonal cell lines. Third, different 5' proximal regulatory elements (promoters and Kozak sequences) elicited tunable and context-dependent gene expression patterns demonstrating that stable recombinant gene expression is both construct-dependent and site-dependent.

Observation of Site-Dependent and Construct-Dependent eGFP Expression Patterns.

Following the

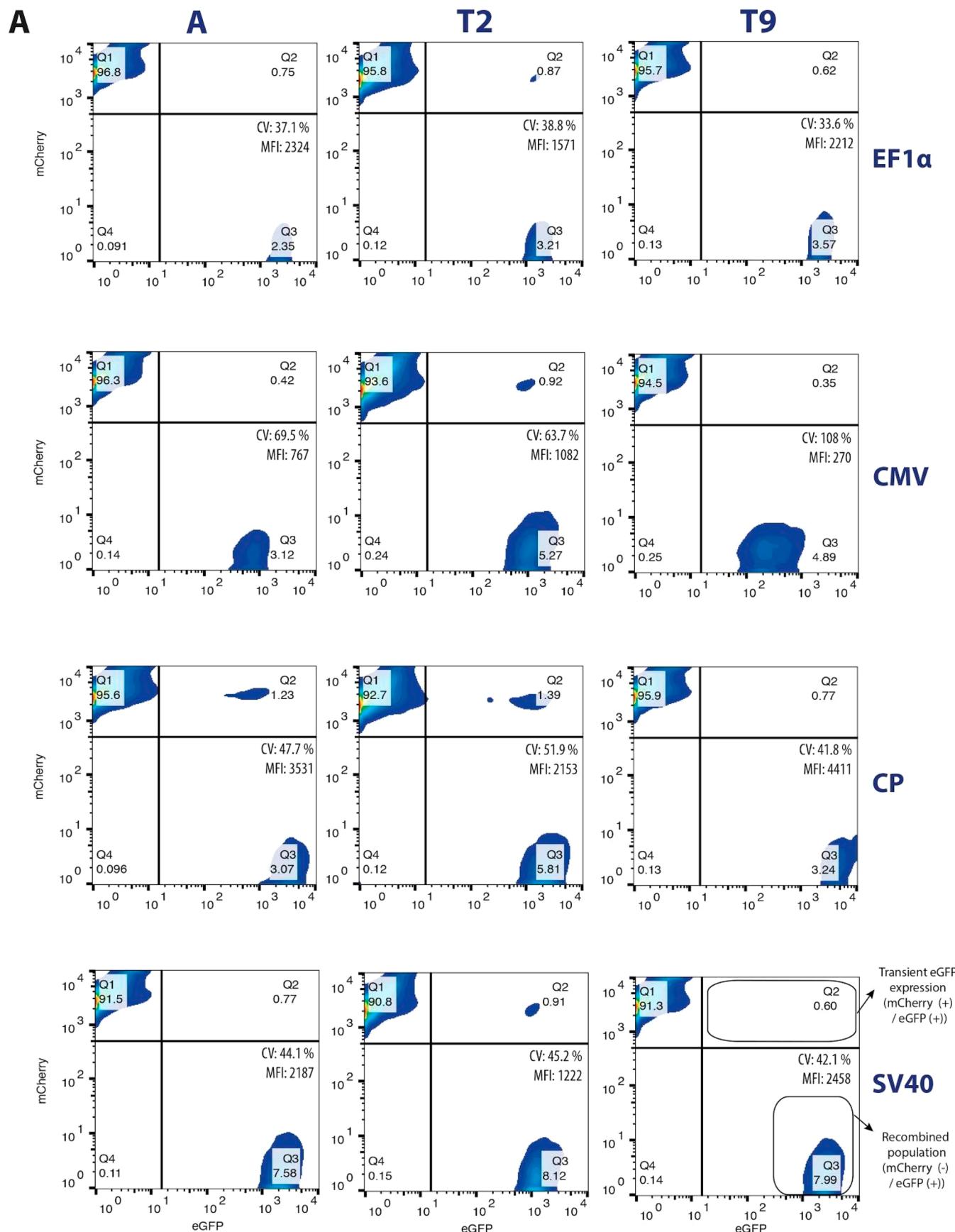


Figure 3. continued

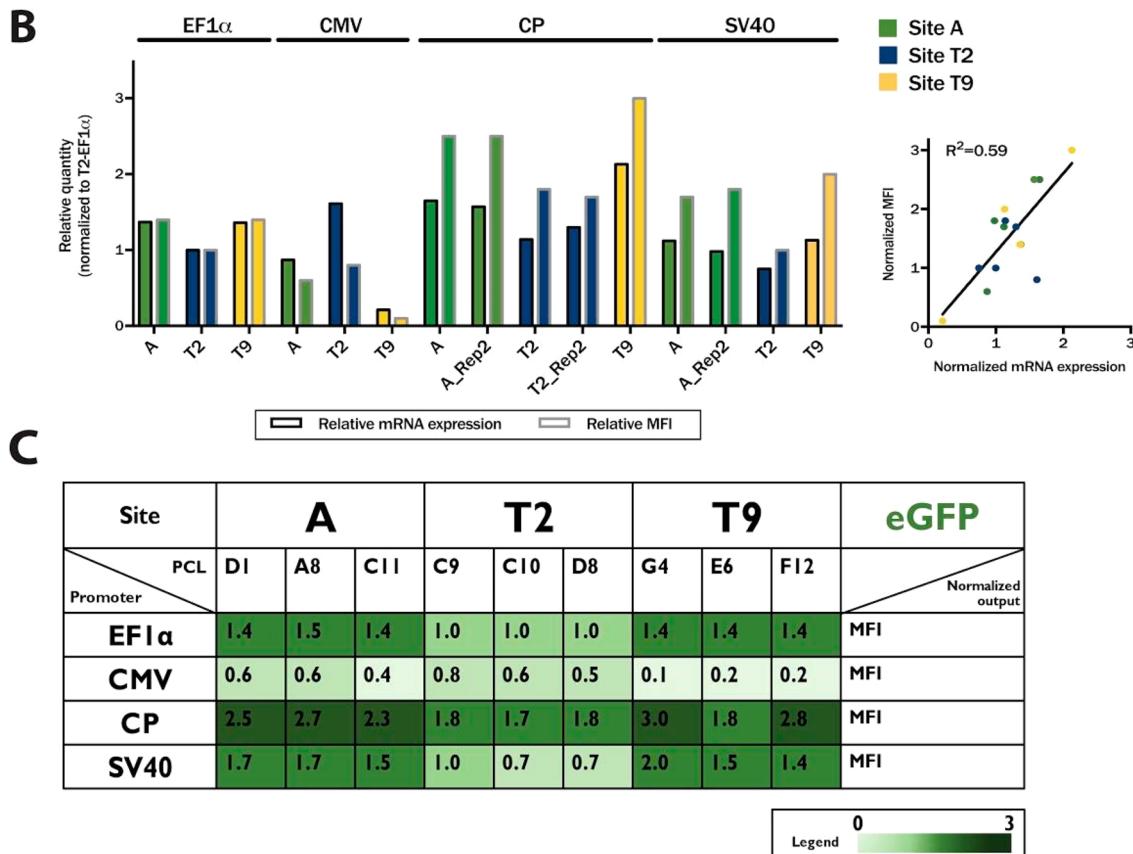


Figure 3. Screening of the promoter-eGFP panel generates site-dependent and construct-dependent expression patterns with large dynamic range and minimal clonal variation. (A) A panel of eGFP constructs with four different promoters (EF1 α , CMV, CP and SV40 promoter) preceded by a strong Kozak sequence was recombined into A-D1, T2-C9 and T9-G4 PCLs. Two weeks post-RMCE, FACS analysis (100 000–200 000 cells) for all 12 conditions was performed assessing MFI and CV of successfully recombined eGFP populations (*i.e.* mCherry negative/eGFP positive cells). (B) All 12 recombined populations were bulk sorted and their eGFP fluorescence and mRNA levels measured. eGFP MFIs and mRNA levels were normalized to the values of T2-EF1 α population (internal calibrator). A linear regression-based correlation analysis of relative eGFP MFIs and relative eGFP mRNA expression levels is shown. RMCE, bulk sorting as well as eGFP fluorescence and mRNA levels measurements were independently replicated for three PCL-promoter combinations (Rep2 samples). (C) Heatmap with normalized MFIs for 36 eGFP recombined populations, measured with FACS (100 000–200 000 cells). The panel of eGFP constructs with four promoters and a strong Kozak sequence was recombined into three different parental PCLs for each site. eGFP MFIs of all populations were normalized to the value of T2-C9-EF1 α population.

successful proof-of-principle experiment, we set out to conduct a systematic evaluation of site-specific gene expression patterns with all 5' proximal regulatory elements in our panel. We therefore generated eGFP constructs with four different promoters—two viral promoters (CMV and SV40), a constitutive promoter (EF1 α) and a composite promoter (CP), all followed by a strong Kozak sequence. By recombining them into the three representative PCLs, we were able to quantify the combined effect of these promoters and the three identified genomic loci on the expression levels from stably integrated eGFP.

Two weeks post-recombination, the recombined population (*i.e.* mCherry negative/eGFP positive population) was observed in all 12 samples, while samples were largely devoid of the transient eGFP expression (*i.e.* mCherry positive/eGFP positive population) (Figure 3A). The percentage of successfully recombined cells varied between 2% to 8%. By comparing the MFI of these recombined populations, the promoter giving rise to the highest expression level was CP in all three sites, followed by the EF1 α and SV40 promoters. For all three promoters, sites A and T9 generated approximately a 2-fold higher output than site T2, consistent with the results seen in the analysis of the mCherry expression levels in 15 PCLs (Figure 1C). While CP

and SV40 promoter demonstrated higher cell-to-cell variation (wider distribution of the fluorescence levels, *i.e.* CV), constitutive EF1 α promoter displayed lower intrinsic noise with a relatively tight and unimodal expression pattern (site A: CV = 37.1%; site T2: CV = 38.8%; site T9: CV = 33.6%). Surprisingly, CMV, generally thought of as one of the strongest exogenous promoters,³⁴ generated the lowest levels of fluorescence in all three sites. While CMV in combination with site T2 generated the strongest MFI when compared to the other two sites, this promoter performed particularly poorly in site T9 with a 16-fold lower MFI compared to CP. Overall, CMV generated the highest cell-to-cell variation in all three sites (site A: CV = 69.5%; site T2: CV = 63.7%; site T9: CV = 108%), indicating that the variation primarily originates from the promoter itself and not the integration site.

To confirm the results seen in the initial fluorescence analysis, where the cell population consisted of a small portion of the recombined cells and a large pool of non-recombined (mCherry positive) cells (Figure 3A), bulk sorting of the recombined populations was performed. After recovery, copy number analysis was conducted, confirming the presence of a single eGFP copy in all recombined populations (Supporting Figure S10). Next, eGFP expression levels were determined at the

Table 1. Long-Term Stability of eGFP Expression Is Ensured in All Three Putative Safe Harbors Regardless of the Expression Cassette Used^a

Site and clone	A-D1			T2-C9			T9-G4			eGFP
	Week	1	4	8	1	4	8	1	4	8
Promoter										Normalized output
EF1 α	1.4	1.4	1.3	1.0	1.0	1.0	1.4	1.3	1.3	MFI
	1.4	1.1	1.2	1.0	0.9	1.0	1.3	1.2	1.3	mRNA expression
CMV	0.6	0.4	0.4	0.8	0.5	0.5	0.1	0.1	0.1	MFI
	0.8	1.0	1.0	1.4	1.4	1.3	0.2	0.3	0.4	mRNA expression
CP	2.5	2.2	1.9	1.8	1.4	1.3	3.0	2.8	2.6	MFI
	1.5	1.5	1.5	1.1	1.0	1.2	2.2	1.6	2.1	mRNA expression
SV40	1.7	1.4	1.3	1.0	0.8	0.8	2.0	1.5	1.5	MFI
	1.0	0.9	1.0	0.7	0.8	0.8	1.1	0.8	1.2	mRNA expression



^aHeatmap with relative MFIs and relative mRNA expression levels for 12 eGFP recombined populations (from Figure 3a,b) at week 1, week 4, and week 8 from the start of long-term cultivation. Three bulk-sorted populations (namely A-CP, A-SV40, and T2-CP) were generated twice, independently (for more details refer to subsection Subpopulation Generation by Recombinase-Mediated Cassette Exchange (RMCE) and Bulk Sorting in the Methods section), and the corresponding MFIs are reported as relative mean MFIs between the two biological replicates. eGFP MFIs and mRNA levels were normalized to the values of T2-C9-EF1 α population at the first time point (week 1; internal calibrator).

mRNA and protein level (Figure 3B). A clear agreement was observed between the fluorescence measurements (Figure 3A and 3B), confirming that the initial MFI measurements prior to bulk sorting give a good proxy for the eGFP expression level. Consistent with the analysis of mCherry expression levels (Figure 1C), a significant linear correlation was observed between eGFP fluorescence and mRNA levels (Figure 3B, $R^2 = 0.59$; $p = 0.0008$). Despite normalizing the data, the discrepancies that we observe between generally lower mRNA levels and higher corresponding protein levels might be on account of slow degradation rates of long-lived proteins such as eGFP that can limit the ability to track rapid mRNA fluctuations. Both the analyses (at mRNA and protein level) confirmed the strength of the promoters in the three sites, observed at the time of bulk sorting (Figure 3A). Moreover, similar gene expression patterns were generated when RMCE and bulk sorting were repeated (independent replicates named A-CP_Rep2, T2-CP_Rep2, A-SV40_Rep2), indicating that data generated by the poly(A) trap platform is reproducible and robust.

In conclusion, these results facilitated ranking of the promoter strength in a site-specific manner. A wide range of site- and promoter-specific expression levels was observed whereby elements have been identified for predictable and tunable gene expression patterns. Finally, it has been shown that commonly used promoters, such as EF1 α and CMV, function in a position-dependent mechanism.

Similar Clone-Specific eGFP Expression Patterns Across All Three Integration Sites. To explore if the findings mentioned above are only specific to the particular PCL used, RMCE was performed with the same set of eGFP constructs into two additional PCLs per site. FACS analysis was performed 2 weeks post-recombination and the generated MFIs were closely following the trend obtained with the original three PCLs (Figure 3C). These results further supported our observations with the mCherry-encoding PCLs, that gene expression patterns are prevalently site- and promoter-specific when controlled

targeted integration is performed. The cell host background, despite additional round of RMCE-based generation of a subpopulation, did not seem to contribute substantially to the differences in gene expression levels in the three selected putative safe harbor sites. Due to robust site-specific expression levels observed across independent replicates (Figure 3b) and different clonal backgrounds (Figure 3c), we postulated that for an initial, quick screening of different expression cassettes in a bulk-sorted population, no additional replicates are necessary.

Long-Term Stability of eGFP Expression Achieved in All Three Integration Sites. To explore the long-term stability of recombinant gene expression under the selected set of promoters when integrated in site A, T2, and T9, a two-month cultivation experiment of eGFP-encoding bulk-sorted populations was performed. eGFP fluorescence and mRNA levels were measured in the beginning (week 1), in the middle (week 4) and at the end (week 8) of the cultivation. None of the bulk-sorted populations displayed a demonstrable loss of the eGFP expression during 8 weeks of culturing (Table 1), retaining the initial values observed in Figure 3B.

These results indicate that the three integration sites characterized in this study provide robust gene expression patterns over long-term culturing. Therefore, these sites seem to constitute safe harbors in the CHO-S genome for integration of transgenes, with varying strength of expression levels dependent on the construct integrated.

Site-Specific Gene Expression Characterization for a Secreted Biotherapeutic. To demonstrate the usefulness of our toolbox for recombinant protein production in mammalian cell lines, we aimed to characterize the expression levels of a secreted biopharmaceutical, EPO. The RMCE was performed in the three representative PCLs with EPO constructs comprising three different promoters (CMV, EF1 α and CP) and the previously described strong Kozak sequence. Two weeks post-recombination, bulk sorting was performed on mCherry negative cells for all 9 conditions (Figure 4A). The mCherry

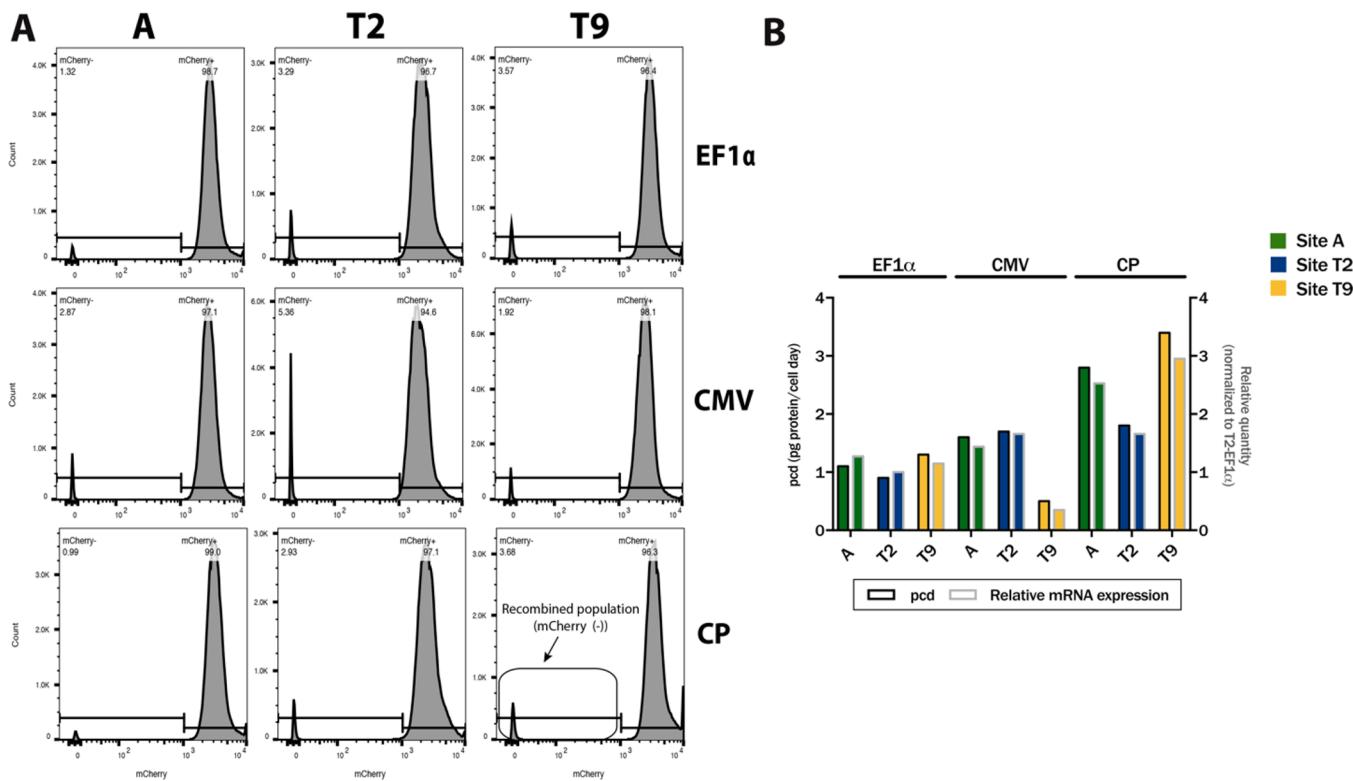


Figure 4. Expression level characterization for a secreted biotherapeutic (EPO) demonstrates protein-specific patterns under different expression cassettes. (A) A panel of erythropoietin (EPO) constructs with three different promoters ($\text{EF1}\alpha$, CMV, and CP) preceded by a strong Kozak sequence was recombined into A-D1, T2-C9, and T9-G4 PCLs. Two weeks post-RMCE, FACS analysis (100 000–200 000 cells) for all 9 conditions was performed assessing percentage of successfully recombined EPO populations (*i.e.* mCherry negative cells). (B) All 9 recombined populations were bulk-sorted and EPO protein and mRNA levels measured. Specific productivities of EPO (pcd; pg protein/cell day) were correlated with relative EPO mRNA expression levels using linear regression. EPO mRNA levels were normalized to the value of T2- $\text{EF1}\alpha$ population.

negative cells were deemed as correctly recombined cells expressing EPO and these populations constituted 1% to 5% of all cells. Bulk-sorted cells were recovered and using dPCR analysis the presence of a single copy of EPO on average was confirmed in these populations (Supporting Figure S8). This was followed by a three-day batch culture in 125 mL shake flasks where specific EPO productivity and mRNA expression levels were measured in the mid-exponential phase of the culture. Specific EPO productivity and EPO mRNA levels displayed a high level of correlation (Figure 4B; $R^2 = 0.98$). The ranking of the promoters for the expression of the secreted EPO was similar to the intracellular eGFP. The strongest promoter in all integration sites was CP, reaching 3 pcd in sites T9 and A. The $\text{EF1}\alpha$ promoter construct generated a 3-fold lower output and approximately the same amount of protein in all three sites (1 pcd). Interestingly, the CMV promoter gave rise to higher EPO mRNA levels and specific productivity values in sites A and T2 compared to the levels provided in these sites for eGFP, but remained the weakest promoter in site T9 (7-fold lower pcd than T9-CP-EPO).

These results demonstrate that our toolbox can be successfully used to modulate the expression levels of a secreted biotherapeutic. Furthermore, we were able to show that promoters can generate protein-specific gene expression patterns at characterized genomic loci, advocating for testing of the desired vector constructs for both, the site of interest and the protein of interest.

DISCUSSION

Tight control of recombinant gene expression is a major quest in different fields of biology, ranging from gene therapy, functional genetics studies in basic research, biopharmaceutical production, to synthetic biology. In order to adapt to ever changing environmental stimuli, eukaryotic cells have complex systems in place for dynamic and flexible regulation of their transcriptional and translational processes. Thus, heterologous gene networks with predictable and desirable outputs are difficult to install in eukaryotic cells when random integration is applied. Molecular tools for site-specific integration are therefore crucial to move this field forward; however, increased specificity and efficiency of these tools have to be achieved for routine applications.³⁵ We and others have developed CRISPR/Cas9-based targeted integration protocols for *in vitro* gene editing in mammalian cells.^{5,10–12,19–21} In this study we wanted to further expand the CRISPR/Cas9 applications by developing a CRISPR/Cas9-based modular toolbox for systematic evaluation of site-specific recombinant expression patterns modulated by different 5' proximal regulatory elements. Here, we describe a systematic and streamlined analysis of stable gene expression patterns resulting from the interplay between the four contributing layers: the cell host, the integration locus, the inserted vector elements and the type of recombinant protein being expressed. By precisely quantifying the resulting expression profile, the objective of this toolbox is to pave the way toward generation of mammalian cell lines with fine-tuned and user-defined gene expression patterns.

First, we characterized three candidate safe harbor integration sites in the CHO-S genome, one in an intragenic, specifically intronic region (site A), and two in intergenic regions (site T2 and T9) (Supporting Table S1) with the objective of further expanding the growing repository of available sites for programmable mammalian cell engineering.^{12,20,24} These sites were selected either on the basis of experimental data with sites showing stable expression or *in silico* analysis identifying sites with high and stable expression profiles of neighboring genes. All three sites identified here are positioned in open chromatin areas that are generally transcriptionally active (Supporting Table S1). High expression levels of transgenes when integrated into regions with high transcriptional activity and chromatin accessibility were recently reported.³⁶ Being in transcriptionally active regions of the genome, these newly proposed regions are possibly less noisy, exhibiting lower cell-to-cell variability in gene expression patterns. Our study indeed unveiled that all three sites generated relatively robust gene expression patterns with minimal clonal variation and approximately 2-fold higher transcriptional output in sites A and T9 in comparison to T2. We have previously shown that minimal clonal variation can be achieved when site-specific integration is applied and that a more homogeneous and predictive transgene expression is observed in targeted clones in comparison to clones generated by random integration.^{5,21} Moreover, since the chromosome regions surrounding the integration sites seem to support relatively stable long-term transgene expression, the regions around site A, T2, and T9 might be less prone to genomic rearrangements from transcriptionally active regions of euchromatin into closed heterochromatin. We therefore believe that these three sites can be considered safe harbors for programmable transgene expression in CHO-S cells. However, further studies are warranted to confirm whether these sites give rise to gene expression patterns as robust as seen in other established safe harbors, such as ROSA26.

Second, we developed a toolbox for generation of PCLs using a novel approach with poly(A) trap-based RMCE landing pads. Poly(A) traps have previously been used, however, in a different setting for gene trapping purposes.³⁷ The toolbox allows users to generate PCLs with their own integration site of interest and use them for rapid evaluation of site-specific gene expression patterns resulting from any desired 5' proximal regulatory elements. After PCLs with the sites of interest are established, the timeline to optimize site-specific expression profiles of any given recombinant protein is two to four weeks from transfection (as measured by fluorescence or protein levels, respectively). This brings the timeline of expression optimization closer to that of transient expression-based experiments.

Third, employing the developed toolbox, we analyzed site-specific expression profiles of genes under transcriptional control of some of the commonly used promoters in mammalian cell engineering. Using eGFP and EPO expression cassettes, we observed that the gene expression is heavily influenced by the interplay between the integration site and the promoter of interest, generating large dynamic range in protein expression. All the promoters tested here were dependent on the site of integration featuring highly genetic context-specific transcriptional regulation. Viral promoters, such as CMV, have previously been shown to be prone to the local epigenetic regulation and silencing.^{38–42} Our study showed that the otherwise potent CMV promoter gave rise to the overall lowest and the most stochastic transgene expression levels of the four promoters studied here. As the transgene copy number is low in our studied

populations (*i.e.* one copy), bursts in transcription/translation can lead to easily detectable differences between otherwise genetically identical cells.⁴³ Therefore, the lower the amount of transcript (*e.g.* CMV at site T9), the higher the intrinsic noise, since low copy number together with low transcription rates can limit the precision of gene regulation.⁴⁴ On the other hand, we observed that the SV40 viral promoter allowed for relatively high and stable transgene expression levels in our integration loci, suggesting the genomic context of these sites is favorable toward high transcription rates under the SV40 promoter. The SV40 promoter has previously been reported to be more resistant to transcriptional silencing than CMV⁴⁵ and is usually included in vector systems to drive the selection marker expression.²⁶ For the purpose of finding high-producing clones it is therefore of utmost importance to carefully devise the expression cassettes of both the GOI and the selection marker by considering the strength, but also general responsiveness of the promoter to the local environment. The strongest promoters in our study were CP (composed of the mCMV enhancer, hEF1 α promoter and 5'UTR HTLV) and EF1 α promoter, making them desirable for the applications where high expression levels are needed. However, for multigene engineering strategies, their usage might be limited due to a few reasons. First, both CP and EF1 α are large promoters that would in turn generate very large multigene vectors that might not be efficiently transfected or integrated. Moreover, if either of these promoters would be used successively or if used together (high similarity between the two promoters) for multiple genes, there might be a higher risk of promoter interference when competing for the same endogenous transcription factors.⁴⁶ This constitutes an important consideration for mammalian synthetic biology, where the more attractive solution might be the integration of a completely synthetic system (*i.e.* a pair of nonmammalian transcription factor and regulatory elements with its cognate binding sites) with minimal host cell interactions.⁴⁷ All in all, these findings stress the importance of having a toolbox to empirically interrogate site-specific gene expression patterns modulated by a range of desired regulatory elements. The toolbox facilitates site- and application-specific design of expression cassettes, preventing the risk of unpredictable loss of promoter functionality, even from increasingly more utilized built-for-purpose synthetic promoters.^{46,48,49}

Numerous studies have been conducted to improve the levels of stable recombinant expression by optimization of expression constructs.^{34,50–56} However, the performance of different vector designs cannot be reliably compared between these studies, since the analyses were conducted: (a) in different cell hosts; (b) in randomly or semirandomly integrated, stable cell pools or clones with high variability in gene copy numbers; and (c) under antibiotic selection and in overall different culture conditions. Only a few studies so far have explored different vector constructs in a site-specific manner,^{14,27,28} confirming that the strength of the expression cassette is defined by its chromosomal context. However, to the best of our knowledge, no systematic evaluation has been performed wherein the performance of different 5' proximal regulatory elements were benchmarked against each other in a controlled manner. In our system, a single copy of 5' proximal regulatory element-GOI is integrated in a defined, retargeted genomic site. Further, comparisons are done without the presence of antibiotic selection, thereby considerably reducing variability in cell-extrinsic factors causing dynamic transcriptional regulation.⁵⁷ In fact, by aligning all three main sources of transcriptional regulation (cell-extrinsic,

cell type-specific and genomic-loci intrinsic⁵⁸), the contribution of any 5' proximal regulatory element to the final transgene expression level can be precisely determined.

Further studies are warranted to explore how the local genomic environment in the three presented sites interacts with the exogenous promoters, to decipher how the regulatory elements affect the transgene expression in these sites. Similarly, it remains to be explored whether any perturbations have been inflicted upon host cell transcriptome after targeted integration in the three putative safe harbors. Generally, transgene integration into a safe harbor should cause little to no transcriptional alterations to the endogenous genes and should not disrupt the host genome in the proximity of integration (*i.e.* genomic rearrangements). Moreover, the preferred promoters for predictable transgene expression should not be inducing changes in the host cell transcriptome by titrating away the endogenous transcription factors from endogenous genes⁵⁹ and should not be as prone to position effects and corresponding epigenetic mechanisms.⁴⁶

There are many potential implications of this toolbox, overall enabling identification of parameters determining robust gene expression patterns and fidelity over long-term culturing. By using this toolbox for mammalian cell line engineering, the inherent stochasticity of gene expression can be minimized and predictable transgene expression profiles can be achieved. Predictable and tunable transgene expression profiles are a requirement for programming more complex cell behaviors, such as expressing entire metabolic pathways with dynamic expression patterns and high temporal precision,⁶⁰ building reliable synthetic circuits,⁶¹ or engineering the optimal effector gene expression levels.⁶² We employed this toolbox to increase the expression levels of a secreted biotherapeutic (EPO), demonstrating the applicability of our workflow for CHO cell line engineering. Due to protein-specific expression patterns, optimization of expression for other biotherapeutics will be necessary and feasible using this toolbox, being particularly useful for molecules such as monoclonal antibodies where the ratio between the two transcriptional units has to be optimally regulated. Similar approaches could be undertaken to optimize the expression context of other recombinant proteins in any integration site of interest across different expression platforms, including a variety of human cells and cell lines.

METHODS

Identification of Integration Sites. Site A was identified in a monoclonal cell line with randomly integrated Rituximab (C6_2), generated with an in-house developed clone screening platform.²⁹ For more information regarding the C6_2 cell line generation, refer to the *Supporting Methods*. The C6_2 cell line was outsourced to Cergentis B.V. (Utrecht, The Netherlands), where integration site A was first identified using the proprietary “Targeted Locus Amplification” (TLA) technology.³⁰ Site A was reanalyzed with an in-house targeted MiSeq deep-sequencing protocol. The integrated construct and flanking genomic regions were amplified from genomic DNA (gDNA) obtained from the C6_2 cell line and was prepared for Illumina MiSeq DNA-sequencing using the Nextera XT v2 set A kit (Illumina) according to the manufacturer's instructions. Sequencing was performed in a pooled run (2 × 150 bp paired-end). The resulting reads were mapped to a draft reference sequence encompassing the inserted heterologous plasmid sequence using CLC Genomics Workbench (version 8.5). Analysis of broken mapped reads was used to manually reconstruct the sequence

junctions and thereby the anticipated sequence of the integrated construct and genomic context. Sites T2 and T9 were predicted based on transcriptomic data and model predictions. Briefly, regions between genes with consistently high expression profiles in a previously published RNA-Seq data set from CHO-S during exponential and stationary phase³² were identified. A region between two highly expressed genes (site T2) and between two highly expressed genes, one of which (*Rrm1*) is essential (site T9), were chosen as integration sites, with the assumption that the surrounding chromatin would remain open, and so it would be less likely that transgenes integrated at those sites would be silenced. The assumption of essentiality was based on literature³³ and *in silico* predicted lethality when simulating a *Rrm1* knockout in the genome-scale model of CHO cell metabolism.³² We utilized flux balance analysis⁶³ to predict essential genes. For this, model simulations used the default uptake and secretion constraints provided previously.³² To account for the possibility that the cells could change which metabolites are being catabolized or secreted, we did not force secretion or uptake of any metabolites in order to exclude false positive predictions of essentiality. The percentile ranks from expression levels of T2 and T9 neighboring genes across the 40 RNA-Seq samples are provided in the *Supporting Figure S3*. Integration regions and the accompanying chromosomal and epigenetic information for all three sites, obtained from the new Chinese hamster reference genome⁶⁴ and the CHO-Epigenome database,⁶⁵ respectively, are available in the *Supporting Table S1*.

Plasmid Design and Construction. The CHO codon optimized Cas9 and sgRNA expression vectors with the same design as previously described⁶⁶ were used. sgRNA target sequences (available in *Supporting Table S2*) with the shortest possible distance to the identified genome-plasmid breakpoints (site A) or to the identified high-expressing locus (site T2 and T9) were designed either using the CRISPy tool⁶⁶ (site A) or manually (site T2 and T9). 5' and 3' homology arms were selected to flank each sgRNA target sequence with the length of 750 bp each, and were amplified from CHO-S genomic DNA. The sequences of both homology arms for each site are provided in the *Supporting Table S2*. Donor plasmids with the landing pad and donor recombinase-mediated cassette exchange (RMCE) plasmids were constructed *via* the uracil-specific excision reagent (USER) cloning method as previously described²¹ with a minor adjustment. Donor landing pad plasmids were designed with mCherry coding sequence together with EF1 α promoter being flanked by *loxP* at the 5' end and *lox2272* at the 3' end (*Figure 1a*). Consequently, donor RMCE plasmids contained both the GOI and promoter of interest between *loxP* at the 5' end and *lox2272* at the 3' end. Plasmids used to PCR amplify expression cassettes of mCherry, eGFP and CHO codon-optimized erythropoietin (EPO) have previously been described.^{21,31,67} Cre recombinase was expressed from the commercial PSF-CMV-CRE recombinase expression vector (Sigma-Aldrich; OGS591). The weak and strong Kozak sequences used here have previously been described.³¹ All the plasmids used in the present study are depicted in the *Supporting Figure S11*, and all the primers used are in the *Supporting Table S3*. The sequences of the GOIs, promoters, terminators and Kozaks used in this study are listed in the *Supporting Table S4*. Assembled PCR fragments were transformed into *E. coli* Mach1 competent cells (Life Technologies). All constructs were verified by sequencing of the expression cassettes (Eurofins Scientific) and purified using NucleoBond

Xtra Midi EF (Macherey-Nagel) according to manufacturer's instructions.

Cell Cultivation. CHO-S cells (Thermo Fisher Scientific) were maintained in CD CHO medium (Gibco) supplemented with 8 mM L-glutamine (Thermo Fisher Scientific) and 2 μ L/mL antyclumping agent (Life Technologies) and cultivated in 125 mL Erlenmeyer flasks. Cells were incubated at 37 °C, 5% CO₂ at 120 rpm (25 mm shaking amplitude) and passaged every 2–3 days, unless otherwise stated. Viable cell density (VCD) and viability were monitored using the NucleoCounter NC-200 Cell Counter (ChemoMetec, Denmark), using Vial-Cassettes and the "Viability and Cell Count Method 2" assay.

Platform Cell Line Generation Using CRISPR/Cas9. CHO-S cells were transfected with the sgRNA plasmid and landing pad donor plasmid (EF1 α -mCherry) for the corresponding integration sites, together with a Cas9 plasmid, according to a previously described protocol.²¹ Stable cell pools were established with G418 selection (500 μ g/mL) as previously described.²¹ To establish monoclonal platform cell lines (PCLs), G418-selected pools were subjected to single-cell cloning using BD FACSJazz cell sorter (BD Biosciences). Double ZsGreen1-DR negative/mCherry positive single cells (*i.e.*, cells with targeted integration only) were seeded in flat-bottom 384-well plates (Corning, Sigma-Aldrich) with 30 μ L of CD CHO medium supplemented with 8 mM L-glutamine, 1.5% HEPES (Gibco) and 1× Antibiotic-Antimycotic (Gibco). Two weeks after the sorting, surviving single-cell derived colonies were transferred to flat-bottom 96-well plates (Corning) with 200 μ L of CD CHO medium supplemented with 8 mM L-glutamine and 1× Antibiotic-Antimycotic using an epMotion 5070 liquid handler (Eppendorf). Subsequently, cells were expanded in suspension and verified by mCherry fluorescent analysis, 5' and 3' junction PCR and copy number analysis by qPCR.

Fluorescent Level Analysis. The generated mCherry PCLs were analyzed using an Xcyto 10 image cytometer (ChemoMetec), applying unstained suspension cells on the slide and examining mCherry fluorescence intensity in 20 000–40 000 individual cells of the population. Only PCLs with homogeneous mCherry expression (\geq 98% of mCherry positive cells) were selected.

Confirmation of Landing Pad Integration. For 5' and 3' junction PCRs, crude gDNA was extracted from the cell pellet using QuickExtract DNA extraction solution (Epicenter, Illumina) according to the manufacturer's instructions. 1–2 μ L of gDNA mixture was used as template for touchdown PCR using 2× Phusion Master Mix (Thermo Fisher Scientific) and the following conditions: 98 °C for 30 s; 98 °C for 10 s; 68–58 °C [−1°/cycle] for 30 s; 72 °C for 2 min; 30×: 98 °C for 10 s; 58 °C for 30 s; 72 °C for 2 min; 72 °C for 10 min. PCR primers for 5'/3' junction PCRs are listed in the [Supporting Table S3](#) and were designed as shown previously.⁵ PCR products were visualized on 1% agarose gels.

Copy Number Analysis Using qPCR. qPCR was conducted to determine the relative copy numbers of the integrated mCherry gene and the Neomycin selection marker (NeoR). GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific) was used for extraction of gDNA according to manufacturer's instruction. qPCR was run on a QuantStudio 5 Real-Time PCR System (Agilent Technologies). Amplification was performed under the following conditions: 50 °C for 2 min, 95 °C for 10 min; 40×: 95 °C for 15 s, 60 °C for 1 min. Copy numbers of mCherry and NeoR were determined with

C1GALT1C1 (COSMC) as an internal control gene for normalization⁶⁸ using TaqMan Gene Expression Master Mix and custom-made TaqMan probes for all three genes (Thermo Fisher Scientific). Primers and probes are listed in the [Supporting Table S5](#) and were validated by melting curve analysis and primer efficiency test. A delta–delta threshold cycle ($\Delta\Delta Ct$) method was applied to calculate the copy number of mCherry/NeoR integrated using two previously generated PCLs with a single copy of mCherry landing pad integrated within COSMC gene and T2 site (single-copy calibrators).^{5,31} Each experiment included no template control and negative control (CHO-S wt as a template) with technical triplicates. Only PCLs with a single copy of both, mCherry and NeoR were expanded and banked.

Batch Cultivation with Single-Copy PCL Clones. Cells were seeded at density 1×10^5 cells/mL in 40 mL volume of CD CHO medium supplemented with 8 mM L-glutamine and 2 μ L/mL antyclumping agent in 125 mL flasks. VCD and viability were measured daily, as described above. 2×10^6 cells were harvested on day three in the mid to late exponential phase by centrifugation (200g, 5 min, RT). Cell pellets were stored at −80 °C. On day three, mCherry fluorescent analysis was performed using Xcyto 10 image cytometer, as described above. Cultures were discontinued when viability dropped below 60%.

Subpopulation Generation by Recombinase-Mediated Cassette Exchange (RMCE) and Bulk Sorting. mCherry PCL clones A-D1, T2-C9 and T9-G4 were randomly selected from the panel of single-copy monoclonal PCLs for site A, T2, and T9, respectively. These three clones were used for recombination experiments, unless otherwise stated. PCLs with a VCD of 1×10^6 cells/mL were transfected with an RMCE donor (with desired 5' proximal regulatory element-GOI) and a Cre recombinase vector, as previously described.²¹ After transfection, cells were passaged for two weeks in order to minimize transient transgene expression, followed by bulk sorting of the recombined cell population. When RMCE donor encoded eGFP, double mCherry negative/eGFP positive cells were considered recombined and were bulk-sorted. When RMCE donor encoded EPO, mCherry negative cells were considered recombined and were bulk-sorted. Cells ($0.5\text{--}1} \times 10^5$ cells/well) were sorted in flat-bottom 96-well plates using a BD FACSJazz cell sorter. Nonrecombined PCLs were used for gating the mCherry negative cells. Bulk-sorted cells were expanded in suspension (to 125 mL flasks) and further used either in a long-term cultivation experiment (eGFP populations) or short-term productivity assay (EPO populations). Independent set of experiments comprising recombination, bulk sorting and long-term cultivation were performed twice in the same original PCL (*i.e.* for A-CP-eGFP, A-SV40-eGFP and T2-CP-eGFP) accounting for two independent replicates for these three conditions.

Copy Number Analysis on Bulk-Sorted Populations Using dPCR. To ensure that bulk-sorted populations on average harbor a single copy of the recombined GOI (eGFP or EPO), relative copy number analysis with digital PCR (dPCR) was performed using a QuantStudio 3D Digital PCR System (Applied Biosystems). Copy numbers of eGFP and EPO were determined using COSMC as an internal control gene for normalization, with custom-made TaqMan probes used for all three genes (Thermo Fisher Scientific). Primers and probes are listed in the [Supporting Table S5](#) and were validated by melting curve analysis and primer efficiency test. gDNA of bulk-sorted populations, isolated with GeneJET Genomic DNA Purification

Kit, together with the QuantStudio 3D Digital PCR Master Mix v2 was loaded onto QuantStudio 3D PCR Chips. Amplification was performed under the following conditions: 96 °C for 10 min; 39X: 60 °C for 2 min, 98 °C for 30 s; 60 °C for 2 min. A previously generated monoclonal PCL with a single copy of EPO integrated within T9 site was used as a single-copy EPO control.²¹ Additionally, a titration experiment was performed to ensure the reliability of the obtained results. gDNA of a bulk-sorted sample (A-CP-strong Kozak-eGFP) was spiked into CHO-S wt gDNA comprising 0%, 10%, 20%, 50%, and 80% of the total gDNA. The corresponding target-to-reference ratio was subsequently determined for each titration mixture. Each experiment included a no template control. The final analysis of the target-to-reference ratio was performed with the QuantStudio 3D Analysis Suite Cloud Software.

Relative mRNA Expression Level Using RT-qPCR. Total RNA was extracted from 1–2 × 10⁶ cells using the RNeasy Plus Mini Kit (Qiagen) according to the manufacturer's instructions. RNA concentration was measured with Qubit fluorometric analysis (Life technologies) and the purity/quality was assessed with Nanodrop (Thermo Fisher Scientific) and 1% agarose gel visualization. cDNA was synthesized from 1.5 to 3 µg of total RNA using the Maxima First Strand cDNA Synthesis Kit for RT-qPCR with dsDNase treatment (ThermoFisher Scientific). RT-qPCR analyses were performed on the QuantStudio 5 Real-Time PCR System using TaqMan Multiplex Master Mix (Thermo Fisher Scientific) in triplexes (GOI and two normalization genes) using the following amplification conditions: 50 °C for 2 min, 95 °C for 10 min; 40X: 95 °C for 15 s, 60 °C for 1 min. Custom-made Taqman assays were used for mCherry, eGFP and EPO, as well as normalization genes *Gnb1* and *Fkbp1a*.⁶⁹ All the primers and probes are listed in the Supporting Table S5 and were validated by melting curve analysis and primer efficiency test. Using the ΔΔC_t method, the relative expression levels of GOIs were calculated by normalization to the geometric mean of expression levels of the two normalization genes. Each experiment included controls with no template and was performed using technical triplicates.

Long-Term Cultivation. eGFP bulk-sorted populations were recovered after cell sorting and expanded in suspension (6–9 passages), and transferred to 6-well plates (Corning) after which long-term cultivations were initiated in complete media containing 1X Antibiotic-Antimycotic. Cells were passaged every 2–3 days and kept in culture for two months. Cells from passage 2 (week 1), passage 11 (week 4) and passage 23 (week 8), as counted from the start of the long-term cultivation, were analyzed for eGFP fluorescence levels using FACS and relative eGFP mRNA expression levels using RT-qPCR. The BD FACSJazz cell sorter was calibrated each time using Rainbow Calibration Particles (8 peaks) 3.0–3.4 µm (BD Biosciences), followed by the use of constant photomultiplier settings for eGFP and mCherry detection. A total of 100 000–200 000 events were collected in the analysis. Data analysis was performed with FACSDiva Software (BD Biosciences) and FlowJo 10 (FlowJo LLC). Median fluorescence intensities (MFIs) of bulk-sorted populations at each time point were normalized to MFI of the T2-EF1 α sample for the ease of representation. For the analysis of eGFP mRNA expression levels, 1–2 × 10⁶ cells/mL cells at late exponential phase were centrifuged (200g, 5 min, RT) and cell pellets were stored at –80 °C. RNA samples were prepared as described above. eGFP mRNA expression levels of bulk-sorted populations at each time

point were normalized to the levels of the T2-EF1 α sample at the first time point (week 1; internal calibrator).

Short-Term Productivity Assay and Titer Measurements. A three-day batch culture was performed to determine specific productivity of EPO in mid-exponential phase for different bulk-sorted populations. Cells were seeded at density 5 × 10⁵ cells/mL in 25 mL volume of CD CHO medium supplemented with 8 mM L-glutamine and 2 µL/mL antyclumping agent in 125 mL flasks. VCD and viability were measured daily, as described above. On day three, EPO titers were determined in supernatants by biolayer interferometry (ForteBio, Pall, Menlo Park, CA) using Streptavidin biosensors (ForteBio 18–5021, Pall), functionalized with the anti-EPO V_HH biotin conjugate (Life Technologies, cat. no.: 7103372100), as previously described.⁶⁷ Absolute EPO titers were calculated using a 7-point calibration curve generated from a dilution series of commercially available EPO (Genscript, Piscataway, NJ, USA). Specific productivity of EPO was calculated using titer and integral viable cell density values as described elsewhere.⁷⁰ On day three 2 × 10⁶ cells were harvested for mRNA analysis by centrifugation at 200g for 5 min at RT and cell pellets were stored at –80 °C, and mCherry fluorescence analysis was performed using Xcyto 10 image cytometer, as described above. EPO mRNA expression levels of bulk-sorted populations were normalized to the levels of the T2-EF1 α sample (internal calibrator).

Statistical Analyses. All statistical analyses were performed using GraphPad Prism 7 software (GraphPad Prism Software Inc., La Jolla, CA). Statistical significance of log2-transformed mCherry mRNA level differences between the integration sites was calculated using an ordinary one-way analysis of variance (ANOVA) with Turkey's multiple comparisons test ($n = 5$ clones/site) where * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, and **** $p < 0.0001$. To analyze the correlation between the two variables, linear regression was used and R^2 together with p -value reported.

ASSOCIATED CONTENT

S Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acssynbio.8b00453](https://doi.org/10.1021/acssynbio.8b00453).

Generation of C6_2 cell line; Residual productivity and copy numbers of Rituximab; Targeted MiSeq deep-sequencing; Neighboring gene expression levels; Integration site information; 5' and 3' junction PCRs; mCherry and NeoR copy number analysis; mCherry fluorescence homogeneity analysis; Growth curves of selected PCLs; Target-to-reference ratios for titration experiment using dPCR; Target-to-reference ratios bulk-sorted samples using dPCR; sgRNA and homology arm target sequences; Depictions of plasmids; Cloning primer sequences; Gene, promoter, terminator and Kozak sequences; qPCR primer and probe sequences (PDF)

AUTHOR INFORMATION

Corresponding Authors

*E-mail: gyunminlee@kaist.ac.kr.

*E-mail: hef@biosustain.dtu.dk.

ORCID

Nuša Pristovšek: [0000-0002-5492-7085](https://orcid.org/0000-0002-5492-7085)

Peter Røgbjerg: [0000-0003-2561-5063](https://orcid.org/0000-0003-2561-5063)

Helene Fastrup Kildegaard: [0000-0003-3727-5721](https://orcid.org/0000-0003-3727-5721)

Author Contributions

N.P. designed and conducted majority of the experiments, analyzed the data, and composed the manuscript. S.N. performed the cloning and helped with the generation of platform cell lines, recombination experiments, and preparation of the Supporting Information. L.M.G. helped with the batch cultivation and provided guidance for the platform cell line generation. H.H. and N.E.L. performed T2 and T9 integration site discovery. P.R. performed MiSeq deep-sequencing for confirmation of site A. H.F.K., M.R.A., H.G.H., and G.M.L. guided the project and supported the experimental designs and data analysis. All authors read and provided critical input for the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Novo Nordisk Foundation (NNF10CC1016517). N.P., M.R.A., and H.F.K. are receiving funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 642663. The authors thank Nachon Charanyanonda Petersen for her assistance with FACS, Mads Valdemar Anderson for designing the sgRNA for site A, Sara Petersen Bjørn for cloning the sgRNA plasmid for site A, Mojtaba Samoudi for the reanalysis of TLA sequencing, Daria Sergeeva for the introduction to digital PCR, and Heena Dhiman for her assistance with acquiring the epigenetic data from the CHO-Epigenome database.

ABBREVIATIONS

Cas9, CRISPR-associated protein 9; CHO, Chinese hamster ovary; CP, composite promoter; CRISPR, clustered regularly interspaced short palindromic repeats; CV, coefficient of variation; EPO, erythropoietin; GOI, gene of interest; HDR, homology-directed repair; MFI, median fluorescence intensity; PCL, platform cell line; RMCE, recombinase-mediated cassette exchange; VCD, viable cell density.

REFERENCES

- Wilson, C., Bellen, H. J., and Gehring, W. J. (1990) Position Effects on Eukaryotic Gene Expression. *Annu. Rev. Cell Biol.* 6, 679–714.
- Daboussi, F., Zaslavskiy, M., Poirot, L., Loperfido, M., Gouble, A., Guyot, V., Leduc, S., Galetto, R., Grizot, S., Oficjalska, D., et al. (2012) Chromosomal Context and Epigenetic Mechanisms Control the Efficacy of Genome Editing by Rare-Cutting Designer Endonucleases. *Nucleic Acids Res.* 40 (13), 6367–6379.
- Wurm, F. M. (2004) Production of Recombinant Protein Therapeutics in Cultivated Mammalian Cells. *Nat. Biotechnol.* 22, 1393.
- Noh, S. M., Sathyamurthy, M., and Lee, G. M. (2013) Development of Recombinant Chinese Hamster Ovary Cell Lines for Therapeutic Protein Production. *Curr. Opin. Chem. Eng.* 2 (4), 391–397.
- Lee, J. S., Kallehauge, T. B., Pedersen, L. E., and Kildegaard, H. F. (2015) Site-Specific Integration in CHO Cells Mediated by CRISPR/Cas9 and Homology-Directed DNA Repair Pathway. *Sci. Rep.* 5, 8572.
- Doudna, J. A., and Charpentier, E. (2014) Genome Editing. The New Frontier of Genome Engineering with CRISPR-Cas9. *Science* 346 (6213), 1258096.
- Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013) Genome Engineering Using the CRISPR-Cas9 System. *Nat. Protoc.* 8 (11), 2281–2308.
- Hsu, P. D., Lander, E. S., and Zhang, F. (2014) Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell* 157 (6), 1262–1278.
- Bandaranayake, A. D., and Almo, S. C. (2014) Recent Advances in Mammalian Protein Production. *FEBS Lett.* 588 (2), 253–260.
- Lee, J. S., Grav, L. M., Pedersen, L. E., Lee, G. M., and Kildegaard, H. F. (2016) Accelerated Homology-Directed Targeted Integration of Transgenes in Chinese Hamster Ovary Cells via CRISPR/Cas9 and Fluorescent Enrichment. *Biotechnol. Bioeng.* 113 (11), 2518–2523.
- Scarcelli, J. J., Shang, T. Q., Iskra, T., Allen, M. J., and Zhang, L. (2017) Strategic Deployment of CHO Expression Platforms to Deliver Pfizer's Monoclonal Antibody Portfolio. *Biotechnol. Prog.* 33 (6), 1463–1467.
- Zhao, M., Wang, J., Luo, M., Luo, H., Zhao, M., Han, L., Zhang, M., Yang, H., Xie, Y., Jiang, H., et al. (2018) Rapid Development of Stable Transgene CHO Cell Lines by CRISPR/Cas9-Mediated Site-Specific Integration into C12orf35. *Appl. Microbiol. Biotechnol.* 102 (14), 6105–6117.
- Kim, M. S., and Lee, G. M. (2008) Use of Flp-Mediated Cassette Exchange in the Development of a CHO Cell Line Stably Producing Erythropoietin. *J. Microbiol. Biotechnol.* 18 (7), 1342–1351.
- Nehlsen, K., Schucht, R., da Gama-Norton, L., Krömer, W., Baer, A., Cayli, A., Hauser, H., and Wirth, D. (2009) Recombinant Protein Expression by Targeting Pre-Selected Chromosomal Loci. *BMC Biotechnol.* 9, 100.
- Zhou, H., Liu, Z.-G., Sun, Z.-W., Huang, Y., and Yu, W.-Y. (2010) Generation of Stable Cell Lines by Site-Specific Integration of Transgenes into Engineered Chinese Hamster Ovary Strains Using an FLP-FRT System. *J. Biotechnol.* 147 (2), 122–129.
- Mayrhofer, P., Kratzer, B., Sommeregger, W., Steinfellner, W., Reinhart, D., Mader, A., Turan, S., Qiao, J., Bode, J., and Kunert, R. (2014) Accurate Comparison of Antibody Expression Levels by Reproducible Transgene Targeting in Engineered Recombination-Competent CHO Cells. *Appl. Microbiol. Biotechnol.* 98 (23), 9723–9733.
- Zhang, L., Inniss, M. C., Han, S., Moffat, M., Jones, H., Zhang, B., Cox, W. L., Rance, J. R., and Young, R. J. (2015) Recombinase-Mediated Cassette Exchange (RMCE) for Monoclonal Antibody Expression in the Commercially Relevant CHOK1SV Cell Line. *Biotechnol. Prog.* 31 (6), 1645–1656.
- Baumann, M., Gludovacz, E., Sealover, N., Bahr, S., George, H., Lin, N., Kayser, K., and Borth, N. (2017) Preselection of Recombinant Gene Integration Sites Enabling High Transcription Rates in CHO Cells Using Alternate Start Codons and Recombinase Mediated Cassette Exchange. *Biotechnol. Bioeng.* 114 (11), 2616–2627.
- Inniss, M. C., Bandara, K., Jusiak, B., Lu, T. K., Weiss, R., Wroblewska, L., and Zhang, L. (2017) A Novel Bxb1 Integrase RMCE System for High Fidelity Site-Specific Integration of mAb Expression Cassette in CHO Cells. *Biotechnol. Bioeng.* 114 (8), 1837–1846.
- Gaidukov, L., Wroblewska, L., Teague, B., Nelson, T., Zhang, X., Liu, Y., Jagtap, K., Mamo, S., Tseng, W. A., Lowe, A., et al. (2018) A Multi-Landing Pad DNA Integration Platform for Mammalian Cell Engineering. *Nucleic Acids Res.* 46 (8), 4072–4086.
- Grav, L. M., Sergeeva, D., Lee, J. S., Marín de Mas, I., Lewis, N. E., Andersen, M. R., Nielsen, L. K., Lee, G. M., and Kildegaard, H. F. (2018) Minimizing Clonal Variation during Mammalian Cell Line Engineering for Improved Systems Biology Data Generation. *ACS Synth. Biol.* 7, 2148.
- Papapetrou, E. P., and Schambach, A. (2016) Gene Insertion Into Genomic Safe Harbors for Human Gene Therapy. *Mol. Ther.* 24 (4), 678–684.
- Sadelain, M., Papapetrou, E. P., and Bushman, F. D. (2012) Safe Harbours for the Integration of New DNA in the Human Genome. *Nat. Rev. Cancer* 12 (1), 51–58.
- Cheng, J. K., Lewis, A. M., Kim, D. S., Dyess, T., and Alper, H. S. (2016) Identifying and Retargeting Transcriptional Hot Spots in the Human Genome. *Biotechnol. J.* 11 (8), 1100–1109.

- (25) Kim, J. Y., Kim, Y.-G., and Lee, G. M. (2012) CHO Cells in Biotechnology for Production of Recombinant Proteins: Current State and Further Potential. *Appl. Microbiol. Biotechnol.* 93 (3), 917–930.
- (26) Romanova, N., and Noll, T. (2018) Engineered and Natural Promoters and Chromatin-Modifying Elements for Recombinant Protein Expression in CHO Cells. *Biotechnol. J.* 13 (3), 1700232.
- (27) Lombardo, A., Cesana, D., Genovese, P., Di Stefano, B., Provasi, E., Colombo, D. F., Neri, M., Magnani, Z., Cantore, A., Lo Riso, P., et al. (2011) Site-Specific Integration and Tailoring of Cassette Design for Sustainable Gene Transfer. *Nat. Methods* 8 (10), 861–869.
- (28) Chen, C.-M., Krohn, J., Bhattacharya, S., and Davies, B. (2011) A Comparison of Exogenous Promoter Activity at the ROSA26 Locus Using a PhiC31 Integrase Mediated Cassette Exchange Approach in Mouse ES Cells. *PLoS One* 6 (8), e23376.
- (29) Pristovsek, N., Hansen, H. G., Sergeeva, D., Borth, N., Lee, G. M., Andersen, M. R., and Kildegard, H. F. (2018) Using Titer and Titer Normalized to Confluence Are Complementary Strategies for Obtaining Chinese Hamster Ovary Cell Lines with High Volumetric Productivity of Etanercept. *Biotechnol. J.* 13 (3), e1700216.
- (30) de Vree, P. J. P., de Wit, E., Yilmaz, M., van de Heijning, M., Klous, P., Verstegen, M. J. A. M., Wan, Y., Teunissen, H., Krijger, P. H. L., Geelen, G., et al. (2014) Targeted Sequencing by Proximity Ligation for Comprehensive Variant Detection and Local Haplotyping. *Nat. Biotechnol.* 32 (10), 1019–1025.
- (31) Petersen, S. D., Zhang, J., Lee, J. S., Jakociunas, T., Grav, L. M., Kildegard, H. F., Keasling, J. D., and Jensen, M. K. (2018) Modular 5'-UTR Hexamers for Context-Independent Tuning of Protein Expression in Eukaryotes. *Nucleic Acids Res.*, DOI: 10.1093/nar/gky734.
- (32) Hefzi, H., Ang, K. S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C. A., Baycin-Hizal, D., Huang, Y., Ley, D., et al. (2016) A Consensus Genome-Scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Syst.* 3 (5), 434–443.e8.
- (33) Cerqueira, N. M. F. S. A., Fernandes, P. A., and Ramos, M. J. (2007) Ribonucleotide Reductase: A Critical Enzyme for Cancer Chemotherapy and Antiviral Agents. *Recent Pat. Anti-Cancer Drug Discovery* 2 (1), 11–29.
- (34) Ho, S. C. L., Mariati, Yeo, J. H. M., Fang, S. G., and Yang, Y. (2015) Impact of Using Different Promoters and Matrix Attachment Regions on Recombinant Protein Expression Level and Stability in Stably Transfected CHO Cells. *Mol. Biotechnol.* 57 (2), 138–144.
- (35) Weber, W., and Fussenegger, M. (2010) Synthetic Gene Networks in Mammalian Cells. *Curr. Opin. Biotechnol.* 21 (5), 690–696.
- (36) O'Brien, S. A., Lee, K., Fu, H.-Y., Lee, Z., Le, T. S., Stach, C. S., McCann, M. G., Zhang, A. Q., Smanski, M. J., Somia, N. V., et al. (2018) Single Copy Transgene Integration in a Transcriptionally Active Site for Recombinant Protein Synthesis. *Biotechnol. J.* 13, No. e1800226.
- (37) Niwa, H., Araki, K., Kimura, S., Taniguchi, S., Wakasugi, S., and Yamamura, K. (1993) An Efficient Gene-Trap Method Using Poly A Trap Vectors and Characterization of Gene-Trap Events. *J. Biochem.* 113 (3), 343–349.
- (38) Jun, S. C., Kim, M. S., Hong, H. J., and Lee, G. M. (2006) Limitations to the Development of Humanized Antibody Producing Chinese Hamster Ovary Cells Using Glutamine Synthetase-Mediated Gene Amplification. *Biotechnol. Prog.* 22 (3), 770–780.
- (39) Yang, Y., Mariati, Chusainow, J., and Yap, M. G. S. (2010) DNA Methylation Contributes to Loss in Productivity of Monoclonal Antibody-Producing CHO Cell Lines. *J. Biotechnol.* 147 (3–4), 180–185.
- (40) Kim, M., O'Callaghan, P. M., Droms, K. A., and James, D. C. (2011) A Mechanistic Understanding of Production Instability in CHO Cell Lines Expressing Recombinant Monoclonal Antibodies. *Biotechnol. Bioeng.* 108 (10), 2434–2446.
- (41) Osterlehner, A., Simmeth, S., and Göpfert, U. (2011) Promoter Methylation and Transgene Copy Numbers Predict Unstable Protein Production in Recombinant Chinese Hamster Ovary Cell Lines. *Biotechnol. Bioeng.* 108 (11), 2670–2681.
- (42) Ho, S. C. L., Koh, E. Y. C., Soo, B. P. C., Mariati, Chao, S.-H., and Yang, Y. (2016) Evaluating the Use of a CpG Free Promoter for Long-Term Recombinant Protein Expression Stability in Chinese Hamster Ovary Cells. *BMC Biotechnol.* 16 (1), 71.
- (43) Raj, A., and van Oudenaarden, A. (2008) Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell* 135 (2), 216–226.
- (44) Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002) Stochastic Gene Expression in a Single Cell. *Science* 297 (5584), 1183–1186.
- (45) Ho, S. C. L., and Yang, Y. (2014) Identifying and Engineering Promoters for High Level and Sustainable Therapeutic Recombinant Protein Production in Cultured Mammalian Cells. *Biotechnol. Lett.* 36 (8), 1569–1579.
- (46) Brown, A. J., and James, D. C. (2016) Precision Control of Recombinant Gene Transcription for CHO Cell Synthetic Biology. *Biotechnol. Adv.* 34 (5), 492–503.
- (47) Müller, K., Engesser, R., Schulz, S., Steinberg, T., Tomakidi, P., Weber, C. C., Ulm, R., Timmer, J., Zurbriggen, M. D., and Weber, W. (2013) Multi-Chromatic Control of Mammalian Gene Expression and Signaling. *Nucleic Acids Res.* 41 (12), e124.
- (48) Cheng, J. K., and Alper, H. S. (2016) Transcriptomics-Guided Design of Synthetic Promoters for a Mammalian System. *ACS Synth. Biol.* 5 (12), 1455–1465.
- (49) Brown, A. J., Gibson, S. J., Hatton, D., and James, D. C. (2017) In Silico Design of Context-Responsive Mammalian Promoters with User-Defined Functionality. *Nucleic Acids Res.* 45 (18), 10906–10919.
- (50) Li, J., Menzel, C., Meier, D., Zhang, C., Dübel, S., and Jostock, T. (2007) A Comparative Study of Different Vector Designs for the Mammalian Expression of Recombinant IgG Antibodies. *J. Immunol. Methods* 318 (1–2), 113–124.
- (51) Majocchi, S., Aritonovska, E., and Mermod, N. (2014) Epigenetic Regulatory Elements Associate with Specific Histone Modifications to Prevent Silencing of Telomeric Genes. *Nucleic Acids Res.* 42 (1), 193–204.
- (52) Saunders, F., Sweeney, B., Antoniou, M. N., Stephens, P., and Cain, K. (2015) Chromatin Function Modifying Elements in an Industrial Antibody Production Platform—Comparison of UCOE, MAR, STAR and cHS4 Elements. *PLoS One* 10 (4), No. e0120096.
- (53) Zboray, K., Sommeregger, W., Bogner, E., Gili, A., Sterovsky, T., Fauland, K., Grabner, B., Stiedl, P., Moll, H. P., Bauer, A., et al. (2015) Heterologous Protein Production Using Euchromatin-Containing Expression Vectors in Mammalian Cells. *Nucleic Acids Res.* 43 (16), e102.
- (54) Kang, S.-Y., Kim, Y.-G., Kang, S., Lee, H. W., and Lee, E. G. (2016) A Novel Regulatory Element (E77) Isolated from CHO-K1 Genomic DNA Enhances Stable Gene Expression in Chinese Hamster Ovary Cells. *Biotechnol. J.* 11 (5), 633–641.
- (55) Ebadat, S., Ahmadi, S., Ahmadi, M., Nematpour, F., Barkhordari, F., Mahdian, R., Davami, F., and Mahboudi, F. (2017) Evaluating the Efficiency of CHEF and CMV Promoter with IRES and Furin/2A Linker Sequences for Monoclonal Antibody Expression in CHO Cells. *PLoS One* 12 (10), No. e0185967.
- (56) Chai, Y.-R., Ge, M.-M., Wei, T.-T., Jia, Y.-L., Guo, X., and Wang, T.-Y. (2018) Human Rhinovirus Internal Ribosome Entry Site Element Enhances Transgene Expression in Transfected CHO-S Cells. *Sci. Rep.* 8 (1), 6661.
- (57) Kaufman, W. L., Kocman, I., Agrawal, V., Rahn, H.-P., Besser, D., and Gossen, M. (2008) Homogeneity and Persistence of Transgene Expression by Omitting Antibiotic Selection in Cell Line Isolation. *Nucleic Acids Res.* 36 (17), e111.
- (58) Voss, T. C., and Hager, G. L. (2014) Dynamic Regulation of Transcriptional States by Chromatin and Transcription Factors. *Nat. Rev. Genet.* 15, 69.
- (59) Brewster, R. C., Weinert, F. M., Garcia, H. G., Song, D., Rydenfelt, M., and Phillips, R. (2014) The Transcription Factor Titration Effect Dictates Level of Gene Expression. *Cell* 156 (6), 1312–1323.

- (60) Le, H., Vishwanathan, N., Kantardjieff, A., Doo, I., Srienc, M., Zheng, X., Somia, N., and Hu, W.-S. (2013) Dynamic Gene Expression for Metabolic Engineering of Mammalian Cells in Culture. *Metab. Eng.* 20, 212–220.
- (61) Xie, M., and Fussenegger, M. (2018) Designing Cell Function: Assembly of Synthetic Gene Circuits for Cell Biology Applications. *Nat. Rev. Mol. Cell Biol.* 19 (8), 507–525.
- (62) Hansen, H. G., Pristovšek, N., Kildegaard, H. F., and Lee, G. M. (2017) Improving the Secretory Capacity of Chinese Hamster Ovary Cells by Ectopic Expression of Effector Genes: Lessons Learned and Future Directions. *Biotechnol. Adv.* 35 (1), 64–76.
- (63) Gutierrez, J. M., and Lewis, N. E. (2015) Optimizing Eukaryotic Cell Hosts for Protein Production through Systems Biotechnology and Genome-Scale Modeling. *Biotechnol. J.* 10 (7), 939–949.
- (64) Rupp, O., MacDonald, M. L., Li, S., Dhiman, H., Polson, S., Griep, S., Heffner, K., Hernandez, I., Brinkrolf, K., Jadhav, V., et al. (2018) A Reference Genome of the Chinese Hamster Based on a Hybrid Assembly Strategy. *Biotechnol. Bioeng.* 115 (8), 2087–2100.
- (65) Feichtinger, J., Hernández, I., Fischer, C., Hanscho, M., Auer, N., Hackl, M., Jadhav, V., Baumann, M., Krempel, P. M., Schmidl, C., et al. (2016) Comprehensive Genome and Epigenome Characterization of CHO Cells in Response to Evolutionary Pressures and over Time. *Biotechnol. Bioeng.* 113 (10), 2241–2253.
- (66) Ronda, C., Pedersen, L. E., Hansen, H. G., Kallehauge, T. B., Betenbaugh, M. J., Nielsen, A. T., and Kildegaard, H. F. (2014) Accelerating Genome Editing in CHO Cells Using CRISPR Cas9 and CRISPy, a Web-Based Target Finding Tool. *Biotechnol. Bioeng.* 111 (8), 1604–1616.
- (67) Kol, S., Kallehauge, T. B., Adema, S., and Hermans, P. (2015) Development of a VHH-Based Erythropoietin Quantification Assay. *Mol. Biotechnol.* 57 (8), 692–700.
- (68) Yang, Z., Halim, A., Narimatsu, Y., Jitendra Joshi, H., Steentoft, C., Schjoldager, K. T.-B. G., Alder Schulz, M., Sealover, N. R., Kayser, K. J., Paul Bennett, E., et al. (2014) The GalNAc-Type O-Glycoproteome of CHO Cells Characterized by the SimpleCell Strategy. *Mol. Cell. Proteomics* 13 (12), 3224–3235.
- (69) Brown, A. J., Gibson, S., Hatton, D., and James, D. C. (2018) Transcriptome-Based Identification of the Optimal Reference CHO Genes for Normalisation of qPCR Data. *Biotechnol. J.* 13 (1), 1700259.
- (70) Pybus, L. P., Dean, G., West, N. R., Smith, A., Daramola, O., Field, R., Wilkinson, S. J., and James, D. C. (2014) Model-Directed Engineering of “Difficult-to-Express” Monoclonal Antibody Production by Chinese Hamster Ovary Cells. *Biotechnol. Bioeng.* 111 (2), 372–385.