Revealing key determinants of clonal variation in transgene expression in recombinant CHO cells using targeted genome editing

Lee, Jae Seong; Park, Jin Hyoung; Kwang Ha, Tae; Samoudi, Mojtaba; Lewis, Nathan E.; Palsson, Bernhard O.; Kildegaard, Helene Faustrup; Min Lee, Gyun

Published in:
A C S Synthetic Biology

Link to article, DOI:
10.1021/acssynbio.8b00290

Publication date:
2018

Document Version
Peer reviewed version

Published in final edited form as:

ACS Synth Biol. 2018 December 21; 7(12): 2867–2878. doi:10.1021/acssynbio.8b00290.

Revealing key determinants of clonal variation in transgene expression in recombinant CHO cells using targeted genome editing

Jae Seong Lee[1,2,*], **Jin Hyoung Park**[3], **Tae Kwang Ha**[1], **Mojtaba Samoudi**[4,5], **Nathan E. Lewis**[4,5,6], **Bernhard O. Palsson**[1,4,6], **Helene Faustrup Kildegaard**[1], and **Gyun Min Lee**[1,3,*]

[1]The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

[2]Department of Molecular Science and Technology, Ajou University, Suwon 16499, Republic of Korea

[3]Department of Biological Sciences, KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea

[4]Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA

[5]The Novo Nordisk Foundation Center for Biosustainability at the University of California, San Diego School of Medicine, CA 92093, USA

[6]Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093, USA

Abstract

Generation of recombinant Chinese hamster ovary (rCHO) cell lines is critical for the production of therapeutic proteins. However, the high degree of phenotypic heterogeneity among generated clones, referred to as clonal variation, makes the rCHO cell line development process inefficient and unpredictable. Here, we investigated the major genomic causes of clonal variation. We found: (1) consistent with previous studies, a strong variation in rCHO clones in response to hypothermia (33 vs 37°C) after random transgene integration; (2) altered DNA sequence of randomly integrated cassettes, which occurred during the integration process, affecting the transgene expression level in response to hypothermia; (3) contrary to random integration, targeted integration of the same expression cassette, without any DNA alteration, into three identified integration sites showed the similar response of transgene expression in response to hypothermia, irrespective of integration

*Correspondence: Jae Seong Lee (jaeseonglee@ajou.ac.kr) and Gyun Min Lee (gyunminlee@kaist.ac.kr).

Author contributions

J.S.L., B.O.P and G.M.L. conceived experiments. J.S.L., H.F.K., B.O.P and G.M.L. designed experiments and analyzed the data. J.S.L. and J.H.P. generated CHO-K1 GFP clones. J.S.L. and T.K.H. performed CRISPR/Cas9-mediated genome editing. J.S.L., M.S. and N.E.L. performed TLA analysis for transgene and integration site sequence. J.S.L and G.M.L. supervised the research and wrote the manuscript with input from all co-authors.

Supporting Information

Relationship between the cell number and total GFP intensity; Overview of TLA; Sequencing coverage profiles across the plasmid generated with TLA primer set 1 and 2; Relative *COSMC* expression levels in CHO-K1 and CHO-S host cells; Validation of targeting clones by junction and out-out PCR; *De novo* integration of transgenes harboring CMV-*GFP* into the HT2 site in CHO-S host cell; *De novo* integration of transgenes harboring EF1α-*mCherry* into the HT2 site in CHO-S host cell; *De novo* integration of transgenes harboring CMV-*GFP* into the HT2 site in an opposite orientation; Plasmids used in this study; sgRNA genomic target sequences; Primer sequences;

site; (4) switching the promoter from CMV to EF1α eliminated the hypothermia response; and (5) deleting the enhancer part of the CMV promoter altered the hypothermia response. Thus, we have revealed the effects of integration methods and cassette design on transgene expression levels, implying that rCHO cell line generation can be standardized through detailed genomic understanding. Further elucidation of such understanding is likely to have a broad impact on diverse fields that use transgene integration, from gene therapy to generation of production cell lines.

Keywords

Chinese hamster ovary cells; Clonal variation; Integration site; Targeted integration; Vector configuration

Therapeutic protein production has been the major revenue source in the biopharmaceutical industry[1,2]. Mammalian-based expression systems, particularly CHO cell lines, are predominately used for the manufacture of high-value biopharmaceuticals that require complex post-translational modifications such as protein folding, assembly, and glycosylation. To meet the increasing demand for therapeutic proteins, there have been tremendous advancements in cell line development, media composition optimization, and process optimization[2]. Recombinant mammalian cell lines have been developed by selecting clones with high productivity and target protein quality and properties suitable for the downstream process.

Despite the successful attainment of high titers for therapeutic proteins through high-throughput clone screening techniques[3], the high degree of phenotypic heterogeneity among clones renders the clone screening process laborious and expensive. In particular, considerable screening efforts are often required to cope with the heterogeneity in specific productivity (q_p)[4-6] and variable responses to different culture conditions[7,8]. This phenomenon, referred to as *clonal variation*, describes phenotypic and genetic variability among cells within a clonal population of either host or recombinant cells. Clonal variation can originate from host cell selection, generation of parental recombinant clonal cells, gene amplification, and maintenance of recombinant clonal cells being subjected to aging[4,5,9-13].

Among a number of variables underlying clonal variation, the transgene integration site is believed to be key to clonal variation in transgene expression when establishing recombinant cell lines based on random integration of transgenes[14-17]. Upon transfection of expression vectors encoding target genes of interest (GOI), a transgene is randomly integrated into the host cell chromosomes. It has been suggested that genomic integration sites with different genomic structural and regulatory backgrounds may influence transgene expression level and sustainability during long-term cultivation[18]. There have been few efforts to understand impacts of genomic environment, including chromosome structure and epigenetic regulation, on the expression and stability of transgenes[13-15,19-21]. Thus, the roles of the transgene integration site and configuration thereof in clonal variation have not been fully elucidated.

In this study, we investigated the issue of clonal variation particularly in differential transgene expression profiles in relation to transgene integration in rCHO cells. Specifically,

we focused on how transgene integration site and transgene configuration impact protein production and phenotypic responses to hypothermia. We found that random integration of the GOI in rCHO cell clones exhibiting different responses to hypothermia was associated with varying levels of damage to vector integrity at different integration sites. CRISPR/Cas9-mediated targeted integration was used to insert an intact GOI into the different genomic integration sites, but this did not lead to transgene expression patterns similar to those observed in the parental clones. Rather, promoter dissection and replacement support the crucial role of promoter elements in the differential transgene expression patterns at the identical genomic site. Thus, we demonstrated that clonal variation is not simply attributed to the transgene integration site, and that genetic rearrangement of vector regulatory elements may bring about distinct phenotypic outcomes.

Results

CHO-K1 GFP clones exhibit variable response of GFP expression levels to hypothermia.

We isolated clones expressing a GFP reporter protein using a single limiting dilution step and measured GFP intensity with a multiwell plate reader. We screened 210 individual clones with respect to cell growth and GFP expression at two temperatures, 37 and 33°C. Since the GFP intensity was directly proportional to cell number (Supporting Figure S1), the specific GFP expression level for each clone was calculated by dividing total GFP intensity by total cell number. The specific GFP expression levels of the 210 clones varied widely, ranging from 1.2 to 431.3 (arbitrary unit) at normal culture temperature (37°C) (Figure 1a) and from 2.1 to 2,272.0 at 33°C (Figure 1b). Upon shifting culture temperature from 37 to 33°C, cell growth was suppressed, and GFP was more highly expressed on average (Table 1). The fold change in specific GFP expression levels for individual clones was compared directly by normalization to the data at 37°C. At a lower culture temperature, the clones exhibited a 0.4- to 16.4-fold change in specific GFP expression levels, demonstrating that clones vary substantially in their response to culture temperature (Figure 1c and Table 1). Although the same expression vector was transfected into the same CHO-K1 host cells, the variability of both absolute GFP expression and responses to culture condition was significant among the clones. These results demonstrate clonal variation in rCHO cell lines following random integration of a transgene.

Targeted sequencing identifies random integration sites of transgenes in the selected CHO-K1 GFP clones.

To elucidate the clonal variation of rCHO cell lines with an emphasis on different responses to hypothermia, we localized and sequenced the transgene integration sites in the CHO genome. To perform this sequencing, we employed targeted locus amplification (TLA)-based sequencing of the transgenes[22].

We looked at transgene expression to identify clones for TLA sequencing. Among 210 analyzed clones, three CHO-K1 GFP clones were selected for analysis based on their different levels of response to hypothermia – strong positive response (HT3), moderate positive response (HT2), and no response (HT1) (Figure 2a). The absolute GFP expression levels of the three selected clones were comparable at 37°C, while the selection marker

NeoR gene was expressed five-fold higher in the HT1 clone than the HT2 and HT3 clones (Figure 2b). The fold change of *GFP* expression in selected clones at 33°C relative to expression at 37°C (hereafter referred to as the response level) indicates that the *GFP* response level was distinctly different from one-fold to three-fold among the selected clones, HT1 to HT3. (Figure 2c). The change in *GFP* production in response to hypothermia did not appear to be related to the absolute *GFP* expression level of the clones (Figure 2b,c). The production change for *NeoR* in the HT3 clone showed more than four-fold increase with the temperature shift, while expression in the HT1 and HT2 clones varied less than two-fold. Thus, the HT3 clone has a higher increase in transgene expression for *GFP* and *NeoR*, but the response of transgenes to hypothermia does not correlate with their absolute expression levels.

Targeted sequencing by proximity ligation uncovered specific integration sites of transgenes in the selected clones (Figure 2d, Supporting Figure S2). For this, two different positions on the *GFP* coding sequence were used as transgene specific sites (the anchor sequences) for amplification. Sequencing coverage profiles across the GFP expression vector showed that the complete vector sequence was covered by sequencing reads (Supporting Figure S3). Further analysis of the breakpoints between the plasmid sequence and the CHO-K1 genome sequence allowed us to identify integration site location and if the integration spanned different contigs. These results demonstrated that rCHO cell lines with different response levels of the transgene to hypothermia harbored their transgenes at different genomic sites.

Targeted integration of transgenes into three identified integration sites reduces clonal variation in response to hypothermia.

To determine whether the different responses of the rCHO clones to hypothermia were due to different transgene integration sites, rCHO cell lines expressing GFP were constructed by CRISPR/Cas9-mediated targeted integration[23] of transgenes into the integration sites that we had identified. Cas9 cleavable sgRNA target sites were selected as close as possible to the identified genome-plasmid breakpoints, which determined the genomic sequences of homology arms flanking target GOI in the donor plasmid (Figure 3a). The target GOI included GFP and NeoR expression cassettes. They were identical in sequence and configuration to the original GFP expression vector except for remaining vector parts that included the gene expression cassette for plasmid replication in bacteria. In addition to the three identified genomic sites, one control locus, *C1galt1c1* (*COSMC*), was included for the targeting experiment, resulting in a total of four different sites. Since no significant increase in *COSMC* mRNA expression was observed in CHO cells upon hypothermia, the *COSMC* locus was chosen as an endogenous ‘no response’ control site (Supporting Figure S4).

A 5’/3’ junction PCR was performed to evaluate correct integration of transgenes into HT1, HT2, and HT3 target sites. For the *COSMC* locus out-out PCR was applied, which spanned both genomic sites encompassing the whole integration unit (Supporting Figure S5). Because CHO cells only have one copy of the *COSMC* gene[24], the targeting event can be detected directly through amplification of the targeted genomic region using the out-out PCR. The 5’/3’ junction PCR or out-out PCR positive-targeted integrants at four sites were cultivated under hypothermic conditions as used for the original CHO-K1 GFP clones, and

then, the GFP response was measured by comparing green fluorescence levels between 37 and 33°C (Figure 3b). Surprisingly, the targeted integrants showed similarly increased GFP expression under hypothermia, regardless of the integration site used. They all showed enhanced GFP expression up to four- or five-fold under hypothermia. The *GFP* mRNA expression level at 37°C was comparable among targeted integrants, which was consistent with that in the original clones (Figure 3c). However, the variation in *NeoR* mRNA expression at 37°C was also within a narrow range of 0.6- to 1.2-fold, whereas an exceptionally high level expression of *NeoR* was observed in the original HT1 clone (Figure 2b and Figure 3c). The *GFP* response level of mRNA confirmed the fluorescence-based measurement of the GFP response level, but at lower levels, two- to three-fold, which may be due to higher stability of GFP protein than that of *GFP* mRNA (Figure 3d). The more sensitive mRNA analysis indicates a distinct difference in the *GFP* response level depending on the integration sites (HT3 and HT1 sites: ~ two fold versus HT2 and *COSMC* sites: ~ three fold), with a similar tendency. The *NeoR* response levels in the targeted integrants were similar to those in the original clones, except for the HT3 clone. In contrast to the apparent high response of the transgene in the original HT3 clone, *de novo* targeting into the genomic site originating from HT3 clone could not induce such a high response.

The reduced clonal variation in the hypothermic responses suggests that the target locus has a limited impact on both transgene expression and its response to hypothermia. Thus, other factors may be involved in differential transgene expression patterns. Additionally, the positive response to hypothermia at the *COSMC* locus implies that the response of the independent exogenous transcription unit did not correlate with the natural response of the endogenous gene in an identical genomic context.

Random integration of transgenes induces rearrangements to the expression vector.

Since integration site location seemed not to contribute substantially to clonal variation, we investigated additional factors affecting clonal variation. We analyzed the integrated transgene sequences in the original clones to assess if structural changes may have occurred upon random integration. The expression vector contains the *GFP* and *NeoR* genes, bounded by the CMV promoter / SV40 polyadenylation signal and SV40 promoter / TK polyadenylation signal, respectively. The two expression cassettes were also juxtaposed with their transcription oriented tandemly (Figure 4a).

We analyzed the TLA sequencing data, by evaluating the genome-plasmid breakpoints. This analysis provided information on structural changes including deletions, relative orientation of the expression vector, and position in the transgene expression vector at which the breakpoints occurred (Figure 4b-d; left and right). We also looked at the fusion reads within the transgene that were extracted from the total reads. Sequence analysis of these transgene-transgene breakpoints allowed for detection of transgene concatemerization and rearrangements of transgenes (i.e., where partial copies of the transgene were fused in different orientations: head to head, head to tail, and tail to tail organizations) (Figure 4b-d; middle). Several types of structural changes were seen, which included deletions in the coding region of target *GFP* and *NeoR* genes, different orientation of vector elements, and discontiguous arrangement of transcription units. In particular, the most evident observations

at both breakpoints in the entire clones were deletions in the CMV promoter (Figure 4b-d). Out of approximately 0.6 kb of the annotated CMV promoter region, 10-90% were deleted in the original clones. These data demonstrate that random integration of the transgene results in a complete loss of transgene integrity, generating unpredictable combinations of rearranged vector elements.

Transgene expression response to hypothermia is affected by promoter selection.

Transgene expression depends on both promoter and target locus[25]. To determine whether promoter choice applies to the response of transgene expression in hypothermia, the CMV promoter within the GFP expression cassette was replaced with an EF1α promoter, another widely used constitutive promoter for high level production of transgenes in rCHO cells. This was done in the clone in which *GFP* was integrated into the genomic site identified from the HT2 clone.

Simultaneous generation of two double-stranded breaks (DSBs) in the genomic region harboring the CMV promoter can cleave off the intervening segment in an HT2 site-targeted integrant while integrating an EF1α promoter by using a donor plasmid as a template (Figure 5a). Junction PCR analysis verified the successful targeting of the EF1α promoter by amplifying genome-EF1α promoter boundaries (Figure 5b). The EF1α promoter resulted in a 2.5-fold higher *GFP* expression level compared to the CMV promoter driven *GFP* expression at 37°C (Figure 5c). However, it eliminated the hypothermia-enhanced expression seen in the parental clone driven by the CMV promoter (Figure 5d). This result indicates no hypothermic response of transgene expression driven by EF1α promoter, despite its localization at the identical locus. The promoter exchange appeared to influence expression patterns of the independent *NeoR* transcription unit that is juxtaposed with *GFP* transcription unit. It resulted in a 50% increase in the absolute *NeoR* mRNA expression level and a 30% decrease in the *NeoR* response level compared to the patterns in the original HT2 site-targeted integrant (Figure 5c,d). The replacement of the coding sequence of *GFP* by *mCherry* did not change the propensity for the positive response to hypothermia (Figure 5e,f). The use of the CMV promoter led to positive responses regardless of transgene, but with different levels (*GFP*: 3.1 versus *mCherry*: 4.7; Figure 5g) when expressed at the identical locus, which supports the impact of promoter on transgene expression patterns.

Reproducible results were acquired in a different CHO host cell, CHO-S, at the HT2 site (Supporting Figure S6). CMV promoter-driven *GFP* expression showed a response level of five to six (Supporting Figure S6b,c), whereas EF1α promoter-driven *mCherry* expression did not respond to hypothermia (Supporting Figure S7). These data indicate that the choice of promoter played pivotal roles not only in expression of transgenes but also in their different responses to the culture condition.

Deletion of a promoter or its elements alters the transgene expression pattern.

We observed substantial structural changes in the promoter regions of the parental clones (Figure 4) and promoter dependence of transgene expression patterns (Figure 5 and Supporting Figure S6 and S7). To further examine the role of promoters in response to hypothermia, the CMV promoter was modified in the HT2 site-targeted integrant. Targeted

integration of an identical target GOI, but in the opposite orientation, into the HT2 site showed a minimal effect of orientation on the levels of both absolute *GFP* expression (1.4-fold compared to $HT2^t$ clone in Figure 3) and response (0.9-fold compared to $HT2^t$ clone in Figure 3; Supporting Figure S8). Thus, the focus was shifted to certain parts of the CMV promoter that might be involved in the differing response levels. To dissect the CMV promoter, either an enhancer part or the entire CMV promoter was deleted from the HT2 site-targeted integrant using CRISPR/Cas9-mediated generation of two DSBs (Figure 6a). Targeted clonal cells harboring deletion of the CMV promoter were isolated from a transient pool of cells` following junction PCR analysis and amplicon sequencing (Figure 6b,c). Concurrent DSBs led to formation of insertion/deletion mutations at two CRISPR/Cas9 target sites, resulting in the excision of the CMV promoter (Figure 6c).

As expected, deletion of the whole CMV promoter reduced the absolute *GFP* expression level significantly by 85%, compared to the parental clone, while increasing the absolute *NeoR* expression level by 27% (Figure 6d). Interestingly, deletion of the whole CMV promoter caused a drop in *GFP* response level by 50% from three-fold to 1.5-fold, but the *NeoR* response level appeared unaffected (Figure 6e). Deletion of the CMV enhancer, on the other hand, showed enhanced expression levels of *GFP* and *NeoR* up to 1.5 to 1.6 at 37°C. These data suggest that a residual core part of the CMV promoter can not only support expression of target GOI, but also have positive effects on the expression of both its direct target and downstream independently regulated transcription unit (Figure 6d). With regard to the response level, a lack of the enhancer part in the CMV promoter resulted in reduction in the levels of both *GFP* and *NeoR*, but to a lesser extent than the effect of whole CMV promoter deletion on the *GFP* response level, by approximately 35% (Figure 6e). Co-downregulation of *GFP* and *NeoR* expression by the deletion of CMV enhancer implies a possibility of interactions of upstream transcriptional regulators with downstream transcription machinery. The *in situ* deletion of promoter elements resulted in variegated expression of the *GFP* and *NeoR* genes at 37 and 33°C. Thus, structural rearrangement in the CMV promoter of parental clones contributes to the variable response to hypothermia.

Discussion

Over the last three decades, clonal variation has been one of the most puzzling aspects of rCHO cell line generation. Here, we aimed to understand the clonal variation in rCHO cells using a hypothermic condition as an example phenotype. Temperature, which is modulated easily in a cell culture process, is a key parameter that affects cell growth and recombinant protein productivity[7]. Mammalian cells, including rCHO cells, are cultivated at 37°C to simulate body temperature. Although lowering culture temperature below 37°C decreases the specific growth rate (μ), many studies reveal beneficial effects. These include slowing of metabolism such as glucose and glutamine consumption[26], oxygen uptake, and ammonium production, thus maintaining high viability for a longer period[27]. Furthermore, resistance to shear stress and apoptosis is improved[28]. Lowering the culture temperature may also increase the specific productivity (q_p) despite its heterogeneous effect on q_p among different rCHO cell lines[7]. In many reports, the q_p of rCHO cells was enhanced by lowering the culture temperature[29-33]. Other studies report that the q_p of rCHO cells was not increased at low temperature[7,34,35]. Thus, the beneficial effect of lowering culture temperature on

recombinant protein production varies considerably across cell lines and target proteins. GFP expression under hypothermia showed distinct differences among CHO-K1 GFP clones in this study, which is line with previous observations (Figure 1).

The issue of clonal variation has been extensively discussed not only for process development, but also for upstream cell line development in CHO cell culture. In general, MTX-mediated gene amplification increases gene copy number, and results in overexpression of mRNA and increased mAb productivity[36]. However, gene amplification does not always induce high protein expression, and the transcription and translation rate exhibits extensive variation among clones[37]. Our sub-culture of recombinant CHO-K1 clonal cells with the *GFP* gene inserted showed that the rate of proliferation and expression of GFP was variable even without gene amplification (Figure 1a). Although clonal variation is one of the most important aspects for recombinant therapeutic protein production, it has not been comprehensively investigated to date. Thus, better understanding of clonal variation concerning both transgene expression levels under normal and modified culture conditions may focus efforts toward screening the clones that are most suitable for subsequent processes.

This heterogeneity is generally attributed to the integration site of the foreign gene in the CHO chromosomes[14,15,18]. To test this hypothesis, the experiments described here were designed to directly assess specific transgene integration sites and their effects on clonal variation of transgene expression profiles, which was observed in parental rCHO cells. TLA sequencing revealed differences in the integration sites of the foreign *GFP* gene among clonally derived cell lines with different response levels of GFP to hypothermia (Figure 2d). However, the direct use of identified integration sites as transgene landing pads did not lead to analogous expression patterns (Figure 3). Rather, identification of several rearrangements, particularly in the CMV promoter regions in the integrated transgene of parental rCHO cells, suggested interrupted function of transgene regulatory elements as another feasible cause of clonal variation (Figure 4). The exchange of promoters from CMV to EF1α and deletions of CMV promoter elements resulted in a change in transgene expression patterns including absolute *GFP* expression and response to hypothermia (Figure 5 and 6). These observations support the aforementioned cause of clonal variation proposed in this study. The present study did not fully investigate the effect of other types of structural changes including deletions in the coding regions and polyadenylation signals, and did not rule out the possibility that vector sequences other than GFP and NeoR expression cassettes affect the expression patterns of the parental clones. Nevertheless, the results demonstrate the complex nature of clonal variation in rCHO cells that should encompass a concept of vector elements and their rearrangement upon random integration beyond the sole genomic integration sites.

The effect of target loci on transgene expression has been extensively addressed in studies of positional effects on either retroviral DNA integration[38] or randomly integrated heterologous reporter genes[39]. Depending on genomic location, more than three to four orders of magnitude of variation are reported in HIV[38] or reporter expression levels[39]. The previous high-throughput genome-wide studies provide sufficient data to determine the causal relationships between transgene expression and certain genomic context such as endogenous enhancers[38,39] or lamina-associated domains[39]. Characterization of different transgene

expression cassettes targeted into the adeno-associated virus site 1 (*AAVS1*) and C-C chemokine receptor type 5 (*CCR5*) loci of human cells revealed a strong dependence of transgene expression on both the promoter and the target locus[25].

In our model system, the three selected clones (HT1, HT2, and HT3) showed comparable *GFP* expression level at 37°C (Figure 2b). Targeted integration of the GFP cassette into identified target loci also resulted in similar *GFP* expression levels (Figure 3c). The expression profile of the hypothermic response, however, was not reproduced upon targeted integration (Figure 3b,d). Moreover, EF1α promoter-driven *GFP* expression exhibited 2.5-fold higher *GFP* expression than the CMV cassette at 37°C, but no hypothermic response compared to 3.1-fold higher *GFP* expression by the CMV promoter under hypothermia (Figure 5d). This result emphasizes the large impact of the promoter on the response and also indicates that the increased *GFP* transcript abundance at a lower temperature was not simply derived from enhanced mRNA stability but rather from transcriptional regulation. The positive response to hypothermia at the endogenous 'no response' *COSMC* locus further supports a much stronger effect of an exogenous promoter in the expression cassette than genomic environment including an endogenous promoter at the target locus on the environmental response (Figure 3c). Recent studies found that genes involved in transcription regulation, including transcription factors, were upregulated under hypothermia[40]. Given the importance of transcription factors and their cognate binding sites within promoters as the other side of the coin of transcriptional regulation[41], further analysis pertaining to global expression profiles of transcriptional factors and their interactions will help understand the distinct expression profiles driven by different promoters.

The GFP and NeoR expression cassettes were juxtaposed and tandemly arranged within the target GOI. Each transcription unit contained the regulatory elements required for independent transcription. According to the results of promoter exchange and deletion, *NeoR* expression patterns appeared to be affected by the upstream *GFP* transcription unit (Figure 5 and 6). This observation may be partly addressed by the term, transcriptional interference (TI), which refers to the direct impact of one transcriptional unit on the second unit linked in *cis*[42,43]. According to the potential mechanisms of TI, deletion of the CMV promoter may relieve competition for *cis*- or *trans*-acting factors or promoter occlusion[42,43], leading to increased expression of downstream *NeoR* (Figure 6d). Given the general definition of TI as the suppressive influence of one transcriptional process, it may not explain the *NeoR* expression patterns shown in the promoter exchange and deletion of the CMV enhancer. For future studies, it would be interesting to explore how the aforementioned different transcriptional factor profiles and regulatory DNA interactions work together to couple the expression pattern of one gene by the promoter of another, which could provide practical implications for the architecture of transgene constructs.

Targeted integration of transgene cassettes has been shown to remodel the epigenetic features of the target loci, such as the removal of repressive chromatin marks around the integration site towards active transcription[25]. Hence, the exogenous promoter appears to act independently of endogenous transcriptional control or to perturb preexisting regulation of transcription, which may lead to desired outcomes. This apparent independence of exogenous cassettes from endogenous control suggests limitations of the predictive value of

regulatory element screening, which is typically based on transcriptome/epigenome data of endogenous genes for cell engineering purposes. Likewise, application of endogenous promoters in heterologous reporter constructs may not be able to fully mimic the activity of a sequence originally seen in its endogenous genomic context[44]. Assessment of gene regulatory crosstalk between exogenous cassettes and genomic environment at certain endogenous loci would further elucidate the clonal variation and improve the predictive power of element screening.

Vector fragmentation is well studied and known to occur at a significant level during transfection and random integration in traditional CHO cell line development[45]. Random rearrangement of the transfected vector makes it difficult to predict its effect on transgene expression and also complicates interpretation of mutual relationships between integration sites and exogenous regulatory elements. Targeted integration, mediated by the homology-directed repair (HDR) pathway, on the other hand, could reduce complexity through the screening process that isolates clones with site-specific integration and intact transgene. This approach may provide a useful tool not only for the analysis of interactions between exogenous transcription units or between transcription unit and genomic context, but also for obtaining more predictable transgene expression levels between clones.

In conclusion, the present report provides insight into the long-standing and fundamental question of clonal variation in rCHO cell lines. It was found that clonal variation is not solely caused by genomic integration sites, but rather vector regulatory elements and their structural rearrangement appear to be stronger determinants of differential transgene expression. Further characterization of interactions between integration sites and vector regulatory elements together with controlled integration of transgenes could uncover underlying features of clonal variation more clearly, and could drive the tailored control of recombinant gene expression in rCHO cells while minimizing clonal variation.

Methods

Plasmids.

The GFP expression plasmid (pEGFP-C1, Clontech) was used to generate parental recombinant CHO-K1 cells expressing GFP. For CRISPR/Cas9-mediated targeted integration of the transgenes, CHO codon optimized Cas9 expression vectors[23], target site-specific sgRNA expression vectors, and target site-specific donor plasmids were used (summarized in Supporting Table S1). Target site-specific sgRNA and donor plasmids were constructed via a uracil-specific excision reagent (USER) cloning method, as described previously[23]. The CRISPR guide RNA selection tool, CRISPOR[46], was applied for generating sgRNA target sequences (Supporting Table S2). To flank each sgRNA target sequence, 5’ and 3’ homology arm regions were selected, and the length of each arm was ~750 bp. Since the target GOI, *GFP*, expressed green fluorescence, an mCherry expression cassette was placed outside the homology arms as an indicator of random integration. All constructs were verified with sequencing and purified using NucleoBond Xtra Midi EF (Macherey-Nagel, Düren, Germany) according to the manufacturer’s instructions.

Cell lines and cell culture.

The CHO-K1 host cells (ATCC CCL-61) were maintained in Dulbecco's modified Eagle's medium (DMEM, Thermo Fisher Scientific) supplemented with 10% (v/v) fetal bovine serum (FBS, Gibco). Cells were cultivated as monolayer cultures in 25 cm^2 T-flasks (Nunc) with a working volume of 5 mL. CHO-S cells (Thermo Fisher Scientific) were maintained in CD CHO medium supplemented with 8 mM L-Glutamine (Thermo Fisher Scientific) and cultivated as suspension cultures in 125 mL Erlenmeyer flasks (Sigma-Aldrich) with a working volume of 30 mL. All cells were incubated in a humidified 5% CO_2 incubator at 37°C without shaking (CHO-K1) or with 120 rpm shaking (CHO-S).

Recombinant cell line generation.

Parental recombinant CHO-K1 cells expressing GFP were made by transfecting pEGFP-C1 into CHO-K1 host cells using Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's instructions. DMEM supplemented with 7% FBS and 550 μg/mL G418 (Gibco) was used as a selection medium for CHO-K1-GFP cell line selection. After approximately 2 weeks of the drug selection process, a limiting dilution step was followed for single clone isolation. Targeted integrants expressing GFP at specific target sites were constructed by transfecting Cas9 and sgRNA expression vectors and donor plasmid at a ratio of 1:1:1 (w:w:w), followed by G418 selection process and clone generation. Lipofectamine 2000 and FreeStyle™ MAX transfection reagent (Thermo Fisher Scientific) were used for CHO-K1 and CHO-S cells, respectively, following the standard protocol. Stable pools of cells were subjected to limiting dilution or single cell sorting using a BD FACSJazz cell sorter (BD Biosciences) for clonal generation, and GFP positive/mCherry negative clones were isolated for further analysis. In promoter/coding sequence replacement and deletion experiments, two sgRNA expression vectors and Cas9 expression vector with or without corresponding donor plasmid were transfected by the Nucleofector 2b device using the Amaxa Cell Line Nucleofector Kit V (Lonza, Switzerland) according to the manufacturer's recommendations (program U-023). Clone generation was followed two or three days post transfection via limiting dilution or FACS.

Targeted locus amplification (TLA) based transgene and integration site sequencing.

The transgene integration sites were identified using Targeted Locus Amplification (TLA)[22]. Overview of the TLA analysis is described in Supporting Figure S2. Two independent primer pairs were designed for the transgenes for further validation and reliability of the break sites and/or occurrence of partial integrations. The primer sets, which are listed in Supporting Table S3, were used in individual TLA amplifications. PCR products were purified, and library preparation was done using the Illumina NexteraXT protocol. Libraries were subsequently sequenced using an Illumina Miseq sequencer. TLA sequencing data were analyzed using a custom pipeline. We leveraged BWA-MEM for read mapping[47], thus allowing partial mapping for the optimal identification of break-spanning reads. SAMtools[48] was also used for manipulating and post-processing alignments. Briefly, TLA reads were mapped to the genome assembly for the Chinese hamster using default parameters. Contigs in which integrations occur were identified. TLA reads were also mapped to the transgene plasmids to identify reads that spanned the genome and integrated plasmid. To identify the

precise breakpoints in the genome and the transgenes, the break spanning fusion reads, which had been mapped to both the genome and the transgene, were extracted and remapped to the hamster genome. Hard clipped alignments representing break sites were called using the findPeaks program from HOMER tools[49]. All breakpoints were visualized by IGV for further validation. Although paired-end sequencing was performed, the paired-end data were treated as single-end in the mapping, since the TLA protocol leads to large amounts of shuffling of the sequence.

Fluorescence level analyses.

Exponentially growing cells were plated into 96 well tissue culture plates (Nunc) at a concentration of 1.0×10^5 cells/mL with a working volume of 200 μL to measure fluorescence intensity of CHO-K1 GFP cell lines. The cells were incubated for 3 days in a humidified 5% CO_2 incubator at two different temperatures, 37 or 33°C. After 3 days of cultivation, the medium was replaced with PBS (200 μL/well) because the growth medium (DMEM) exhibited a strong auto-fluorescence. The fluorescence intensity of GFP was measured at an excitation wavelength of 483 nm and an emission wavelength of 535 nm using multiwell plate reader (Infinite 200 PRO Tecan™, TECAN, Switzerland). To measure viable cell concentration, cells were stained by PrestoBlue® Cell Viability Reagent (Thermo Fisher Scientific) according to the manufacturer's protocol, and cell concentration was determined by measuring the resulting fluorescent signal using Infinite 200 PRO Tecan™. For CHO-K1 and CHO-S based targeted integrants expressing GFP or mCherry, Celigo Imaging Cell Cytometer (Nexcelom Bioscience) was used to measure fluorescence applying the mask + target1 application as described previously[23]. The mask represented individual cells that were stained with blue fluorescent NucBlue™ Live ReadyProbes™ Reagent (Thermo Fisher Scientific), and the target1 represented the green or red fluorescence channel.

Genomic DNA extraction and PCR amplification of target regions.

Genomic DNA was extracted from the cell pellets using QuickExtract™ DNA extraction solution (Epicentre, Illumina) according to the manufacturer's instructions. Junction PCR and out-out PCR used 1-2 μL of genomic DNA mixture. PCR was carried out using 2x Phusion Master Mix (Thermo Fisher Scientific) by touchdown PCR following cycling instructions: 98°C for 30 s; 10x: 98°C for 10 s, 68-58°C (−1°C/cycle) for 30 s, 72°C for 30 s – 1 min 30 s depending on amplicon size; 30x: 98°C for 10 s, 58°C for 30 s, 72°C for 30 s – 1 min 30 s depending on amplicon size; 72°C for 10 min. DreamTaq DNA polymerase (Thermo Fisher Scientific) based PCR conditions[23] were used for *COSMC*-targeting clones. PCR primers are listed in Supporting Table S3.

Quantitative real time PCR (qRT-PCR).

qRT-PCR was carried out on genomic DNA and RNA samples to determine relative copy number and relative RNA expression levels of transgenes, respectively. Genomic DNA was extracted from approximately 5.0×10^6 cells using GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific). RNA was extracted from a minimum of 1.0×10^6 cells using TRIzol™ Reagent (Thermo Fisher Scientific), followed by DNase treatment to remove contaminating DNA (TURBO DNA-free™ DNase Treatment and Removal Reagents,

Thermo Fisher Scientific). cDNAs were synthesized from 1 μg of total RNAs using Maxima First Strand cDNA Synthesis Kit for RT-qPCR (Thermo Fisher Scientific). The qRT-PCR was performed in an Mx3005P qPCR System (Agilent Technologies) using Brilliant III Ultra-Fast SYBR® Green QPCR Master Mix (Agilent Technologies) as described previously[23]. Target regions were amplified using primers listed in Supporting Table S3 that were validated by melting curve analysis and agarose gel electrophoresis. Fold changes were calculated using the ΔΔCT method. *Vinculin* was used as normalizer for genomic DNA quantification[23] and *β-actin* (*ACTB*) was used for RNA quantification since the expression of *β-actin* RNA/protein was considered constant and was not affected by temperature in CHO cells[50].

Statistical analysis.

No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment. An unpaired two-tailed *t*-test, one-way analysis of variance followed by the Tukey/Bonferroni post-hoc test, or Welch test followed by the Games-Howell post-hoc test were performed to determine significance for comparisons of mRNA and protein expression levels. *P* values were calculated with SPSS (Statistical Package for the Social Sciences, IBM Corporation), and $P < 0.05$ was considered significant.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported in part by the Novo Nordisk Foundation (NNF10CC1016517), the NRF funded by the Korean government (2016R1A2B4014133 and 2018R1C1B6001423), NIH (R35 GM119850), and the new faculty research fund of Ajou University (S2017G000100097). The authors declare no competing financial interests.

References

1. Walsh G (2014) Biopharmaceutical benchmarks 2014. Nat. Biotechnol. 32, 992–1000. [PubMed: 25299917]
2. Zhu J (2012) Mammalian cell protein expression for biopharmaceutical production. Biotechnol. Adv. 30, 1158–1170. [PubMed: 21968146]
3. Browne SM, and Al-Rubeai M (2007) Selection methods for high-producing mammalian cell lines. Trends Biotechnol. 25, 425–432. [PubMed: 17659798]
4. Kim SJ, Kim NS, Ryu CJ, Hong HJ, and Lee GM (1998) Characterization of chimeric antibody producing CHO cells in the course of dihydrofolate reductase-mediated gene amplification and their stability in the absence of selective pressure. Biotechnol. Bioeng. 58, 73–84. [PubMed: 10099263]
5. Barnes LM, Bentley CM, and Dickson AJ (2001) Characterization of the stability of recombinant protein production in the GS-NS0 expression system. Biotechnol. Bioeng. 73, 261–270. [PubMed: 11283909]
6. Misaghi S, Shaw D, Louie S, Nava A, Simmons L, Snedecor B, Poon C, Paw JS, Gilmour-Appling L, and Cupp JE (2016) Slashing the timelines: Opting to generate high-titer clonal lines faster via viability-based single cell sorting. Biotechnol. Prog. 32, 198–207. [PubMed: 26587808]
7. Yoon SK, Hwang SO, and Lee GM (2004) Enhancing effect of low culture temperature on specific antibody productivity of recombinant Chinese hamster ovary cells: clonal variation. Biotechnol. Prog. 20, 1683–1688. [PubMed: 15575699]

8. Ryu JS, Kim TK, Chung JY, and Lee GM (2000) Osmoprotective effect of glycine betaine on foreign protein production in hyperosmotic recombinant chinese hamster ovary cell cultures differs among cell lines. Biotechnol. Bioeng. 70, 167–175. [PubMed: 10972928]

9. Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, Chen W, Xie M, Wang W, Hammond S, Andersen MR, Neff N, Passarelli B, Koh W, Fan HC, Wang J, Gui Y, Lee KH, Betenbaugh MJ, Quake SR, Famili I, Palsson BO, and Wang J (2011) The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nat. Biotechnol. 29, 735–741. [PubMed: 21804562]

10. Lewis NE, Liu X, Li Y, Nagarajan H, Yerganian G, O'Brien E, Bordbar A, Roth AM, Rosenbloom J, Bian C, Xie M, Chen W, Li N, Baycin-Hizal D, Latif H, Forster J, Betenbaugh MJ, Famili I, Xu X, Wang J, and Palsson BO (2013) Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. Nat. Biotechnol. 31, 759–765. [PubMed: 23873082]

11. Davies SL, Lovelady CS, Grainger RK, Racher AJ, Young RJ, and James DC (2013) Functional heterogeneity and heritability in CHO cell populations. Biotechnol. Bioeng. 110, 260–274. [PubMed: 22833427]

12. O'Callaghan PM, Berthelot ME, Young RJ, Graham JW, Racher AJ, and Aldana D (2015) Diversity in host clone performance within a Chinese hamster ovary cell line. Biotechnol. Prog. 31, 1187–1200. [PubMed: 25918883]

13. O'Callaghan PM, and Racher AJ (2015) Building a Cell Culture Process with Stable Foundations: Searching for Certainty in an Uncertain World In Animal Cell Culture (Al-Rubeai M, Ed.) pp 373–406, Springer, Cham.

14. Lattenmayer C, Loeschel M, Steinfellner W, Trummer E, Mueller D, Schriebl K, Vorauer-Uhl K, Katinger H, and Kunert R (2006) Identification of transgene integration loci of different highly expressing recombinant CHO cell lines by FISH. Cytotechnology 51, 171–182. [PubMed: 19002887]

15. Li S, Gao X, Peng R, Zhang S, Fu W, and Zou F (2016) FISH-Based Analysis of Clonally Derived CHO Cell Populations Reveals High Probability for Transgene Integration in a Terminal Region of Chromosome 1 (1q13). PLoS One 11, e0163893. [PubMed: 27684722]

16. Migliaccio AR, Bengra C, Ling J, Pi W, Li C, Zeng S, Keskintepe M, Whitney B, Sanchez M, Migliaccio G, and Tuan D (2000) Stable and unstable transgene integration sites in the human genome: extinction of the Green Fluorescent Protein transgene in K562 cells. Gene 256, 197–214. [PubMed: 11054549]

17. Cheng JK, Lewis AM, Kim do S, Dyess T, and Alper HS (2016) Identifying and retargeting transcriptional hot spots in the human genome. Biotechnol. J. 11, 1100–1109. [PubMed: 27311394]

18. Dickson AJ (2009) Importance of Genetic Environment for Recombinant Gene Expression In Cell Line Development (Al-Rubeai M, Ed.) pp 83–96, Springer, Dordrecht.

19. Kim SJ, and Lee GM (1999) Cytogenetic analysis of chimeric antibody-producing CHO cells in the course of dihydrofolate reductase-mediated gene amplification and their stability in the absence of selective pressure. Biotechnol. Bioeng. 64, 741–749. [PubMed: 10417224]

20. Baik JY, and Lee KH (2017) A framework to quantify karyotype variation associated with CHO cell line instability at a single-cell level. Biotechnol. Bioeng. 114, 1045–1053. [PubMed: 27922175]

21. Kim M, O'Callaghan PM, Droms KA, and James DC (2011) A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. Biotechnol. Bioeng. 108, 2434–2446. [PubMed: 21538334]

22. de Vree PJ, de Wit E, Yilmaz M, van de Heijning M, Klous P, Verstegen MJ, Wan Y, Teunissen H, Krijger PH, Geeven G, Eijk PP, Sie D, Ylstra B, Hulsman LO, van Dooren MF, van Zutven LJ, van den Ouweland A, Verbeek S, van Dijk KW, Cornelissen M, Das AT, Berkhout B, Sikkema-Raddatz B, van den Berg E, van der Vlies P, Weening D, den Dunnen JT, Matusiak M, Lamkanfi M, Ligtenberg MJ, ter Brugge P, Jonkers J, Foekens JA, Martens JW, van der Luijt R, van Amstel HK, van Min M, Splinter E, and de Laat W (2014) Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. Nat. Biotechnol. 32, 1019–1025. [PubMed: 25129690]

23. Lee JS, Kallehauge TB, Pedersen LE, and Kildegaard HF (2015) Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. Sci. Rep. 5, 8572. [PubMed: 25712033]
24. Yang Z, Halim A, Narimatsu Y, Jitendra Joshi H, Steentoft C, Schjoldager KT, Alder Schulz M, Sealover NR, Kayser KJ, Paul Bennett E, Levery SB, Vakhrushev SY, and Clausen H (2014) The GalNAc-type O-Glycoproteome of CHO cells characterized by the SimpleCell strategy. Mol. Cell Proteomics 13, 3224–3235. [PubMed: 25092905]
25. Lombardo A, Cesana D, Genovese P, Di Stefano B, Provasi E, Colombo DF, Neri M, Magnani Z, Cantore A, Lo Riso P, Damo M, Pello OM, Holmes MC, Gregory PD, Gritti A, Broccoli V, Bonini C, and Naldini L (2011) Site-specific integration and tailoring of cassette design for sustainable gene transfer. Nat. Methods 8, 861–869. [PubMed: 21857672]
26. Ryll T, Dutina G, Reyes A, Gunson J, Krummen L, and Etcheverry T (2000) Performance of small-scale CHO perfusion cultures using an acoustic cell filtration device for cell retention: characterization of separation efficiency and impact of perfusion on product quality. Biotechnol. Bioeng. 69, 440–449. [PubMed: 10862682]
27. Furukawa K, and Ohsuye K (1998) Effect of culture temperature on a recombinant CHO cell line producing a C-terminal alpha-amidating enzyme. Cytotechnology 26, 153–164. [PubMed: 22358553]
28. Ludwig A, Tomeczkowski J, and Kretzmer G (1992) Influence of the temperature on the shear stress sensitivity of adherent BHK 21 cells. Appl. Microbiol. Biotechnol. 38, 323–327. [PubMed: 1369159]
29. Fogolin MB, Wagner R, Etcheverrigaray M, and Kratje R (2004) Impact of temperature reduction and expression of yeast pyruvate carboxylase on hGM-CSF-producing CHO cells. J. Biotechnol. 109, 179–191. [PubMed: 15063626]
30. Chen ZL, Wu BC, Liu H, Liu XM, and Huang PT (2004) Temperature shift as a process optimization step for the production of pro-urokinase by a recombinant Chinese hamster ovary cell line in high-density perfusion culture. J. Biosci. Bioeng. 97, 239–243. [PubMed: 16233622]
31. Fox SR, Patel UA, Yap MG, and Wang DI (2004) Maximizing interferon-gamma production by Chinese hamster ovary cells through temperature shift optimization: experimental and modeling. Biotechnol. Bioeng. 85, 177–184. [PubMed: 14705000]
32. Tharmalingam T, Sunley K, and Butler M (2008) High yields of monomeric recombinant beta-interferon from macroporous microcarrier cultures under hypothermic conditions. Biotechnol. Prog. 24, 832–838. [PubMed: 19194894]
33. Yoon SK, Song JY, and Lee GM (2003) Effect of low culture temperature on specific productivity, transcription level, and heterogeneity of erythropoietin in Chinese hamster ovary cells. Biotechnol. Bioeng. 82, 289–298. [PubMed: 12599255]
34. Clark KJ, Chaplin FW, and Harcum SW (2004) Temperature effects on product-quality-related enzymes in batch CHO cell cultures producing recombinant tPA. Biotechnol. Prog. 20, 1888–1892. [PubMed: 15575729]
35. Weidemann R, Ludwig A, and Kretzmer G (1994) Low temperature cultivation--a step towards process optimisation. Cytotechnology 15, 111–116. [PubMed: 7765923]
36. Jiang Z, Huang Y, and Sharfstein ST (2006) Regulation of recombinant monoclonal antibody production in chinese hamster ovary cells: a comparative study of gene copy number, mRNA level, and protein expression. Biotechnol. Prog. 22, 313–318. [PubMed: 16454525]
37. Lattenmayer C, Trummer E, Schriebl K, Vorauer-Uhl K, Mueller D, Katinger H, and Kunert R (2007) Characterisation of recombinant CHO cell lines by investigation of protein productivities and genetic parameters. J. Biotechnol. 128, 716–725. [PubMed: 17324483]
38. Chen HC, Martinez JP, Zorita E, Meyerhans A, and Filion GJ (2017) Position effects influence HIV latency reversal. Nat. Struct. Mol. Biol. 24, 47–54. [PubMed: 27870832]
39. Akhtar W, de Jong J, Pindyurin AV, Pagie L, Meuleman W, de Ridder J, Berns A, Wessels LF, van Lohuizen M, and van Steensel B (2013) Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell 154, 914–927. [PubMed: 23953119]
40. Bedoya-López A, Estrada K, Sanchez-Flores A, Ramírez OT, Altamirano C, Segovia L, Miranda-Ríos J, Trujillo-Roldán MA, and Valdez-Cruz NA (2016) Effect of Temperature Downshift on the

Transcriptomic Responses of Chinese Hamster Ovary Cells Using Recombinant Human Tissue Plasminogen Activator Production Culture. PLoS One 11, e0151529. [PubMed: 26991106]

41. Brown AJ, and James DC (2016) Precision control of recombinant gene transcription for CHO cell synthetic biology. Biotechnol. Adv. 34, 492–503. [PubMed: 26721629]
42. Shearwin KE, Callen BP, and Egan JB (2005) Transcriptional interference--a crash course. Trends Genet. 21, 339–345. [PubMed: 15922833]
43. Eszterhas SK, Bouhassira EE, Martin DI, and Fiering S (2002) Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. Mol. Cell. Biol. 22, 469–479. [PubMed: 11756543]
44. Le H, Vishwanathan N, Kantardjieff A, Doo I, Srienc M, Zheng X, Somia N, and Hu WS (2013) Dynamic gene expression for metabolic engineering of mammalian cells in culture. Metab. Eng. 20, 212–220. [PubMed: 24055788]
45. Ng SK, Lin W, Sachdeva R, Wang DI, and Yap MG (2010) Vector fragmentation: characterizing vector integrity in transfected clones by Southern blotting. Biotechnol. Prog. 26, 11–20. [PubMed: 19847885]
46. Haeussler M, Schönig K, Eckert H, Eschstruth A, Mianné J, Renaud JB, Schneider-Maunoury S, Shkumatava A, Teboul L, Kent J, Joly JS, and Concordet JP (2016) Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biol. 17, 148. [PubMed: 27380939]
47. Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv:1303.3997v2 [q-bio.GN].
48. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, and Durbin R; 1000 Genome Project Data Processing Subgroup. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079. [PubMed: 19505943]
49. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, and Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589. [PubMed: 20513432]
50. Baik JY, Lee MS, An SR, Yoon SK, Joo EJ, Kim YH, Park HW, and Lee GM (2006) Initial transcriptome and proteome analyses of low culture temperature-induced expression in CHO cells producing erythropoietin. Biotechnol. Bioeng. 93, 361–371. [PubMed: 16187333]

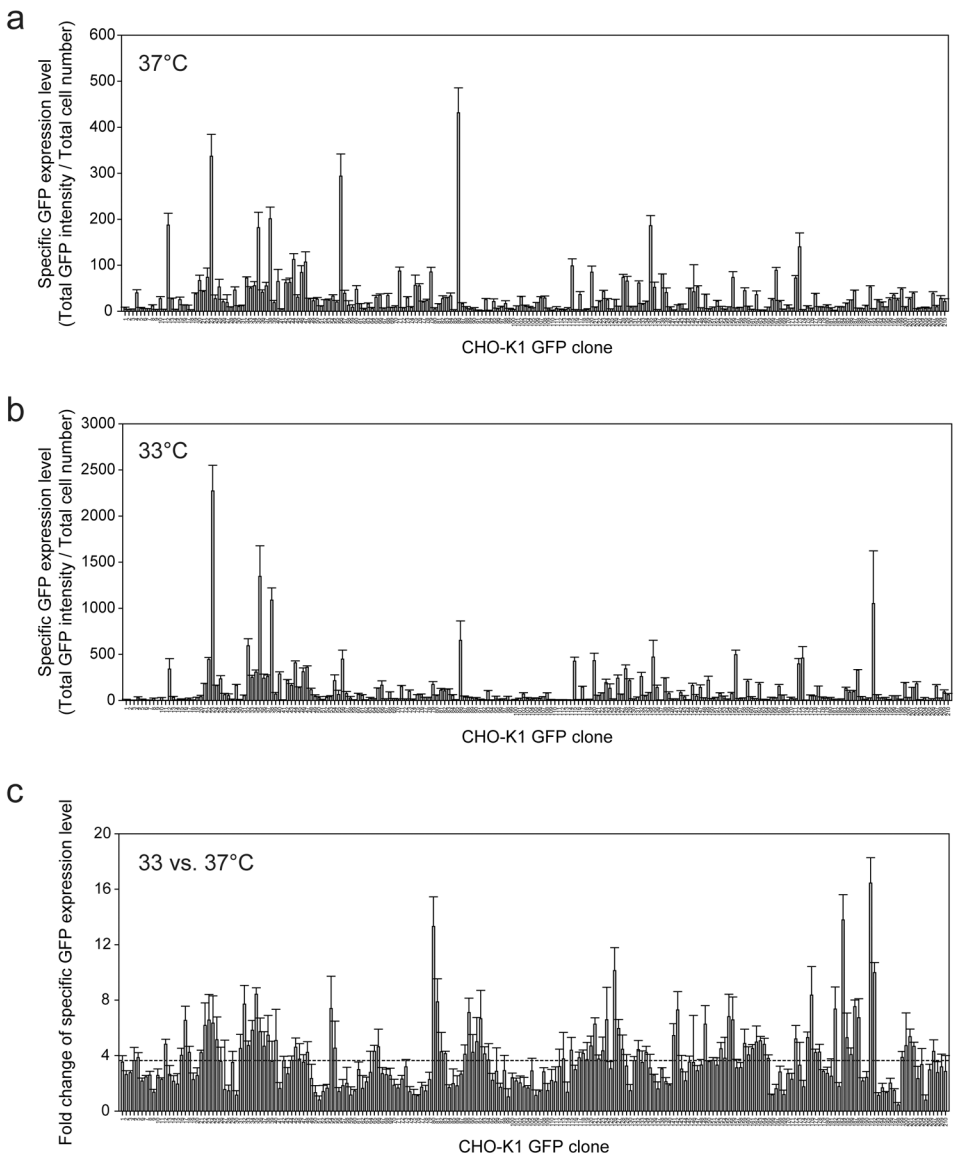

Figure 1. Clonal variation in transgene expression levels in traditional cell line development.

CHO-K1 GFP clones were constructed by the typical process of recombinant cell line development including the introduction of the GFP expression vector into CHO-K1 host cell, followed by clone selection and screening. Quantification of specific GFP expression levels of CHO-K1 GFP clones at **(a)** normal culture temperature, 37°C or **(b)** lowered culture temperature, 33°C. The specific GFP expression levels were calculated by dividing total GFP intensity by total cell number, and normalized to the lowest value at 37°C. **(c)** Fold change of the specific GFP expression levels of CHO-K1 GFP clones (33 versus 37°C). A dotted line indicates an average value of fold change. Cellular GFP contents of 210 isolated clonal cells were measured after 3-day cultivation period. The error bars represent the standard deviations calculated from data obtained in triplicate experiments.

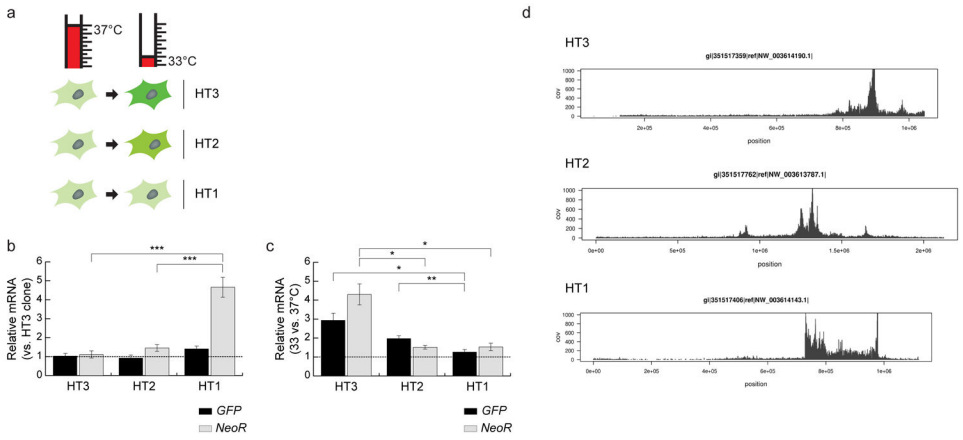

Figure 2. Identification of transgene integration sites of representative clones with positive response in hypothermic cultivation.

(a) Overview of three clones showing different responses to hypothermia from a strong positive response clone (HT3) to a no response clone (HT1). **(b)** Relative RNA expression levels of *GFP* and *NeoR* transgenes in selected clones at 37°C compared to the HT3 clone. The data were analyzed by one-way analysis of variance followed by the Tukey post-hoc test for significance, ***$P < 0.001$ **(c)** Fold change in transgene RNA expression in selected clones at 33°C relative to expression at 37°C. Results are shown as the average values ± standard error (SE) in three and four independent cultivation experiments in b and c, respectively. The data were analyzed by the Welch test followed by the Games-Howell post-hoc test for significance, *$P < 0.05$, **$P < 0.01$ **(d)** TLA-based transgene integration site sequencing in selected clones. Remapping of the breakpoints spanning fusion reads from TLA data identified the precise breakpoints between the transgene and the genome. All breakpoints that were visualized by IGV were represented. Position indicates genomic position on the dedicated scaffold (bp).

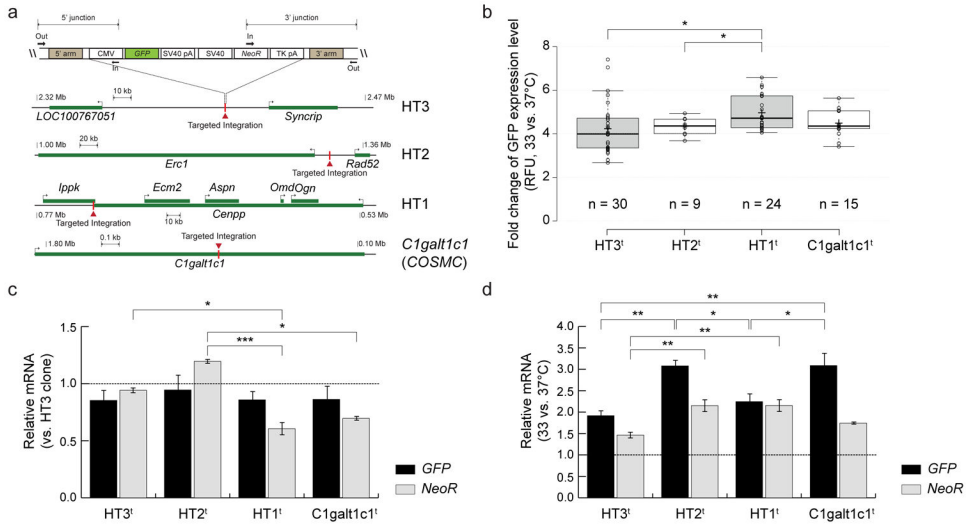

Figure 3. ***De novo*** **integration of transgenes into identified integration sites.**

(a) Representation of genomic regions across targeting sites including HT3 to HT1 sites, and one control site, *C1galt1c1* (*COSMC*) where no significant increase in RNA was observed under hypothermia. Neighboring genes (green bar) and their direction of transcription (arrow) are shown. Red arrow heads indicate sgRNA target sites where GFP and NeoR expression cassettes in donor plasmids are inserted through HDR-mediated repair upon generation of DSBs. Arrows indicate primer positions for PCR analysis to identify targeted integration. **(b)** Hypothermic cultivation of targeting clones. Targeting clones were denoted by a superscript t. The fold change in GFP expression levels (Relative fluorescence unit, RFU) is shown. Mean RFU values were used to calculate the fold change. Center lines show the medians; box limits indicate the 25th and 75th percentiles as determined by R software; whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles, outliers are represented by dots; crosses represent sample means; data points are plotted as open circles. Data points are from three independent cultivation experiments of 10, 3, 8, and 5 targeting clones at the HT3, HT2, HT1, and *C1galt1c1* sites, respectively. The data were analyzed by the Welch test followed by the Games-Howell post-hoc test for significance, $*P < 0.05$. **(c)** The relative RNA expression levels of *GFP* and *NeoR* transgenes in targeting clones at 37°C compared to those in a targeting clone at the HT3 site. **(d)** Fold change in transgene RNA expression in targeting clones at 33°C relative to expression at 37°C. In c and d, results are shown as the average values ± SE from four targeting clones at the HT3, HT2, and *C1galt1c1* sites and five targeting clones at the HT1 site. In c and d, the data were analyzed by one-way analysis of variance followed by the Bonferroni post-hoc test for significance, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$.

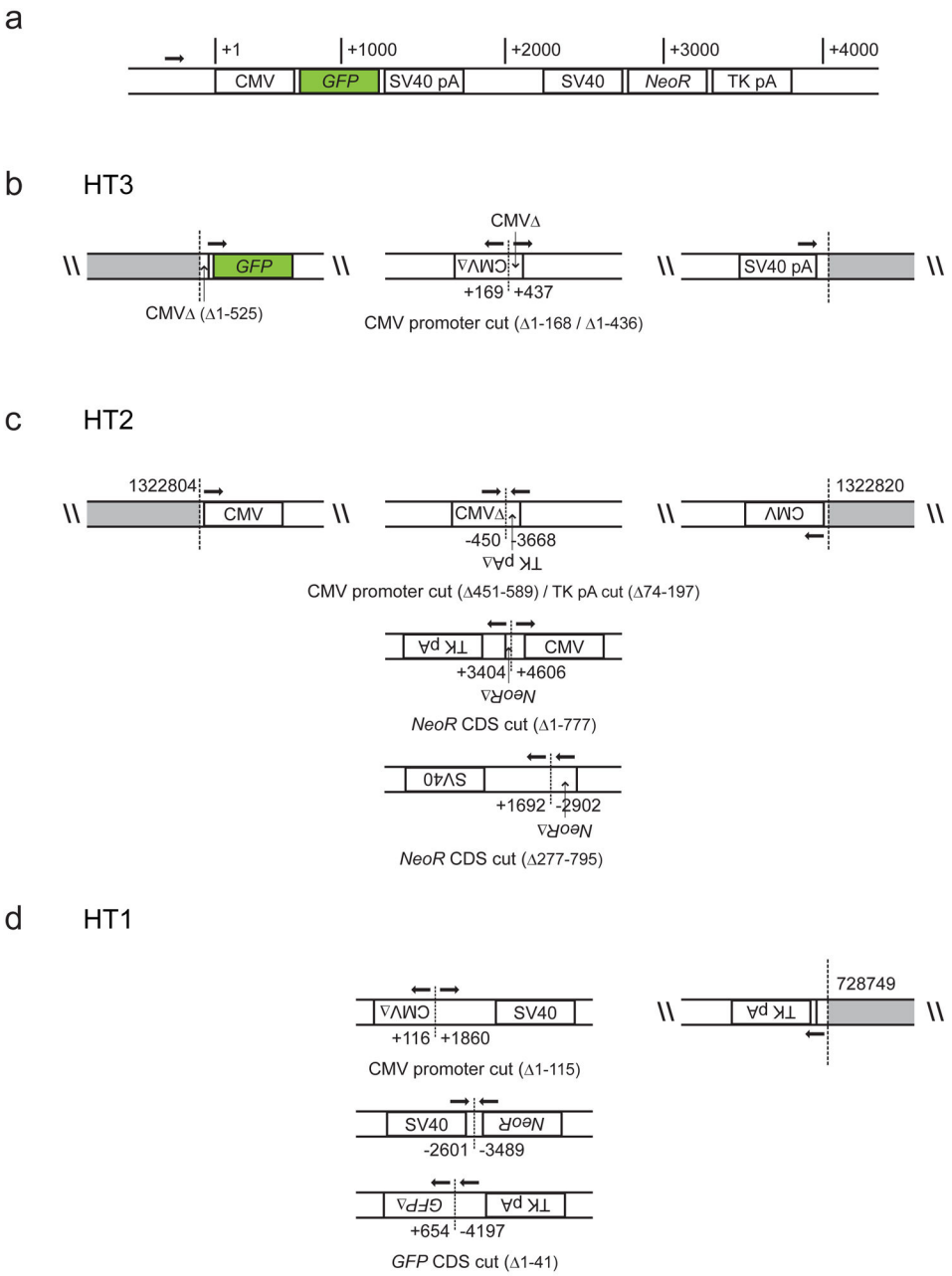

Figure 4. Random integration induces structural rearrangements in transgene sequences.

(a) Representation of the reference plasmid map. Schematic representation of breakpoints (left and right sides) and rearranged vector elements in the parental **(b)** HT3, **(c)** HT2, and **(d)** HT1 clones.

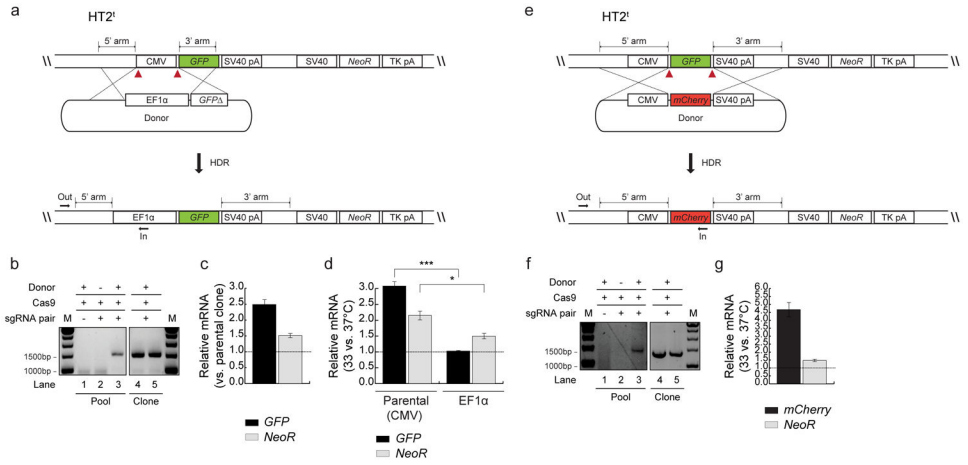

Figure 5. Effect of the promoter on transgene expression.

(a) Schematic representation of a targeting approach to replace the CMV promoter with an EF1α promoter in the targeted integrant at the HT2 site (HT2^t). Red arrow heads indicate sgRNA target sites where the EF1α promoter in a donor plasmid are inserted while cleaving off the intervening CMV promoter segment upon generation of DSBs. Homologous regions are depicted as dotted lines. Arrows indicate primer positions for junction PCR analysis. **(b)** Validation of promoter targeting clones by junction PCR. Agarose gel results of junction PCR on stable cell pools and representative targeting positive clonal cells are shown. Promoter replacement events were detected only in the presence of Cas9, sgRNA pair and donor plasmids. **(c)** Relative RNA expression levels of *GFP* and *NeoR* transgenes in EF1α driven clones at 37°C compared to the expression level of the parental HT2 site-targeted integrant. **(d)** Fold change in transgene RNA expression in EF1α driven clones at 33°C relative to expression at 37°C. Relative RNA expression values observed in the parental clone are plotted as reference. $*P < 0.05$, $***P < 0.001$ by 2-tailed, unpaired *t*-test. **(e)** Schematic representation of a targeting approach to replace *GFP* with the *mCherry* coding sequence in the targeted integrant at the HT2 site (HT2^t), as described in (a). **(f)** Validation of the coding region targeting clones by junction PCR method, as described in b. **(g)** Fold change in transgene RNA expression in CMV-*mCherry* clones at 33°C relative to expression at 37°C. In c, d, and g, results are shown as the average values ± SE from two targeting clones in two independent cultivation experiments.

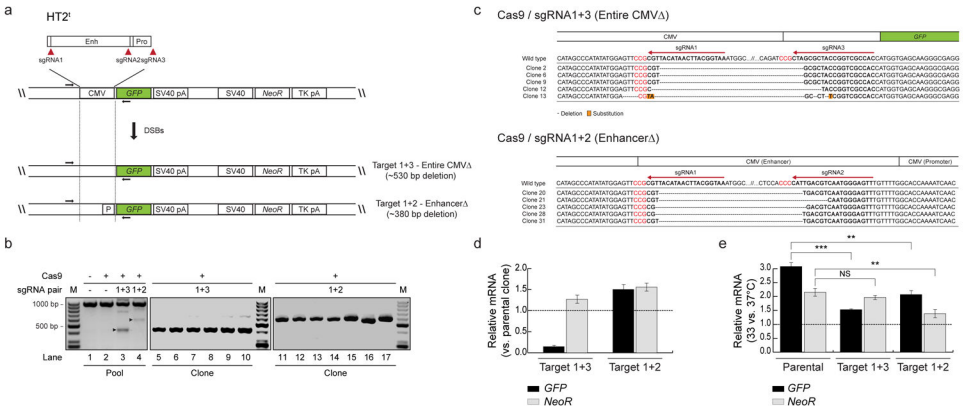

Figure 6. Dissection of the CMV promoter.

(a) Schematic representation of CMV promoter dissection in the targeted integrant at the HT2 site (HT2^t). Three sgRNA target sites cleaved by Cas9 are shown (red triangles), which results in excision of a whole part of the CMV promoter by targeting site 1 and 3 or an enhancer part by targeting sites 1 and 2. Arrows indicate primer positions for PCR analysis. **(b)** Genomic deletions detected in stable cell pools and representative deletion positive clonal cells. Arrow heads indicate deletion-specific PCR amplicons shown on stable cell pools together with wild type amplicons. **(c)** Validation of deletion clones by amplicon sequencing. Deletion-specific PCR amplicons of selected deletion clones were exposed to Sanger sequencing. Representation of reference plasmid map magnifying the CMV promoter region are shown above DNA sequences. Target site PAM sequences are shown in red, and sgRNA-matching sequences are shown in bold type. **(d)** Relative RNA expression levels of *GFP* and *NeoR* transgenes in promoter deletion clones at 37°C compared to the expression level of the parental HT2 site-targeted integrant. **(e)** Fold change in transgene RNA expression in promoter deletion clones at 33°C relative to expression at 37°C. In d and e, results are shown as the average values ± SE from five targeting clones. Relative RNA expression values observed in the parental clone were plotted as reference. $^{**}P < 0.01$, $^{***}P < 0.001$ by 2-tailed, unpaired *t*-test; NS, not significant.

Table 1.

Characteristics of CHO-K1 GFP clones under different environmental culture conditions.

	Control (37°C)	Hypothermia (33°C)
Specific GFP expression level variation (Arbitrary unit)	1.2 – 431.3	2.1 – 2,272.0
Average of specific GFP expression levels *	32.5 ± 10.0	118.2 ± 58.1
Fold change variation	-	0.4 - 16.5
Average of fold change *	-	3.6 ± 0.9
Clone number	210	210

* Pooled average values ± pooled standard deviation (SD_{pooled}) are shown.