A multi-landing pad DNA integration platform for mammalian cell engineering

Leonid Gaidukov[1,†], Liliana Wroblewska[2,*,†], Brian Teague[1], Tom Nelson[3], Xin Zhang[2], Yan Liu[2], Kalpana Jagtap[1], Selamawit Mamo[1], Wen Allen Tseng[1], Alexis Lowe[2], Jishnu Das[4,5], Kalpanie Bandara[3], Swetha Baijuraj[3], Nevin M. Summers[1], Timothy K. Lu[1,6], Lin Zhang[3] and Ron Weiss[1,*]

[1]Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, [2]Biomedicine Design, Pfizer Inc, Cambridge, MA 02139, USA, [3]Cell Line Development, Biotherapeutics Pharmaceutical Science, Pfizer Inc, Andover, MA 01810, USA, [4]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, [5]Ragon Institute of MGH, MIT & Harvard, Cambridge, MA 02139, USA and [6]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

ABSTRACT

Engineering mammalian cell lines that stably express many transgenes requires the precise insertion of large amounts of heterologous DNA into well-characterized genomic loci, but current methods are limited. To facilitate reliable large-scale engineering of CHO cells, we identified 21 novel genomic sites that supported stable long-term expression of transgenes, and then constructed cell lines containing one, two or three 'landing pad' recombination sites at selected loci. By using a highly efficient BxB1 recombinase along with different selection markers at each site, we directed recombinase-mediated insertion of heterologous DNA to selected sites, including targeting all three with a single transfection. We used this method to controllably integrate up to nine copies of a monoclonal antibody, representing about 100 kb of heterologous DNA in 21 transcriptional units. Because the integration was targeted to pre-validated loci, recombinant protein expression remained stable for weeks and additional copies of the antibody cassette in the integrated payload resulted in a linear increase in antibody expression. Overall, this multi-copy site-specific integration platform allows for controllable and reproducible insertion of large amounts of DNA into stable genomic sites, which has broad applications for mammalian synthetic biology, recombinant protein production and biomanufacturing.

INTRODUCTION

Mammalian cell lines that support reliable and predictable expression of large numbers of transgenes are an enabling technology for a wide range of scientific, industrial and therapeutic applications. In a biomanufacturing context, such cell lines could be used to improve production of recombinant proteins that can treat autoimmune disorders, cancer and other diseases (1,2). There is also an increasing interest in augmenting cell lines with entirely new synthetic gene networks that can dramatically change the cells' phenotype and behavior (3). These methods may one day form the basis for 'smart' cellular therapeutics that can sense disease biomarkers and respond appropriately, treating or curing currently intractable ailments (4).

Such large-scale engineering of a cell's genome requires the ability to precisely and efficiently integrate large amounts of heterologous DNA into genomic loci that support robust expression of transgenes, but current genome-engineering approaches fall short for this purpose. One class of methods involves random integration: for instance, heterologous DNA can be packaged in a retrovirus that inserts the DNA payload semi-randomly into the genome (5–9). Because multiple retroviral particles can infect each cell, transducing a culture with a large number of viruses can lead to multiple integrations and very high transgene expression levels. However, commonly used retroviral vectors can only package a modest amount of DNA, and the trans-

duced populations are highly heterogeneous which necessitates significant work to isolate a stable clonal population.

An alternate approach integrates payload DNA using the cell's native DNA repair machinery. By flanking a linear transgene with DNA that is homologous to a desired genomic insertion site, transfected cells can insert the transgene into the target site via homologous recombination with low frequency (10). The efficiency of this recombination process can be improved by using zinc-finger nucleases, TALE-effector nucleases and CRISPR/Cas systems to induce double-stranded breaks at defined locations (11,12). However, the frequency of homologous recombination decreases as the size of the inserted cassette increases (13), limiting the amount of heterologous DNA that can be inserted in a single integration.

A third class of techniques uses site-specific recombinases to insert DNA into the genomes of mammalian cells. First, a 'landing pad' (LP) containing a recombination site and a selectable marker is integrated into the genome. Then, a matching recombinase is used to insert a DNA payload specifically into that locus, allowing for reproducible integration at well-defined sites in the genome (14–16). Unfortunately, only a limited number of well-validated 'safe harbor' sites have been described, and current approaches only allow the integration of a single cassette. Cell lines harboring multiple well-characterized integration sites could allow for integration of different transgenes at different sites, or reproducible multiple integrations of a single cassette and correspondingly higher transgene expression levels. Such cell lines could serve as easily customized 'chassis', simplifying large-scale genome engineering for basic research and biotechnological applications (17–23).

Here, we describe the integration of multiple well-characterized LP sites into the genome of the CHO-K1 cell line, which has gained popularity for the production of recombinant protein therapeutics due to its human-like pattern of post-translational modification and its excellent safety and regulatory profile (24). First, we used a lentiviral integration screen to identify 21 stable integration loci and found that a majority supported long-term stable gene expression in the absence of selective pressure. Next, we inserted LPs at selected loci using a CRISPR/Cas9 genome editing approach and demonstrated that they retained the desirable stability of gene expression. Finally, we created cell lines bearing two and three LPs and demonstrated integration into up to three LP sites in a single transfection. We then demonstrated their utility by using LPs with different fluorescent reporters and antibiotic selection markers to target payload integration into selected LP sites from a multi-LP cell line. By combining a multi-LP cell line with expression cassettes bearing multiple copies of a monoclonal antibody, we demonstrated precise control over protein expression levels via controlled integration of between one and nine copies of the transgene. Importantly, because these LPs are located in pre-validated stable genomic sites, recombinant protein expression levels remained stable for weeks of continuous culture without selective pressure and recombinant protein levels increased linearly with increasing gene copy number. Overall, these results suggest that cell lines harboring LPs in multiple well-characterized sites enable precise, reproducible engineering of the CHO-K1 genome and allow for predictable expression of heterologous transgenes from those sites, which is poised to have a variety of applications in mammalian synthetic biology, cell line engineering and biomanufacturing.

MATERIALS AND METHODS

Vector construction

A lentiviral LP integration vector (pLV-LP) was constructed using Gateway cloning as previously described (15) with the following modifications: Gateway destination vectors contained the pFUGW (Addgene plasmid 14883, a gift from David Baltimore) (25) backbone and Gateway cassette (chloramphenicol resistance and ccdB genes flanked by attR4 and attR2 recombination sites). A multi-site LR reaction was performed between the destination vector, an entry vector carrying the human elongation factor 1 alpha (hEF1a) promoter followed by BxB1-attP site and flanked by attL4 and attR1 recombination sites, and an entry vector encoding a EYFP-P2A-Hygromycin cassette and flanked by attL1 and attL2 recombination sites. This generated an expression vector, pLV-hEF1a-BxB1-attP-EYFP-P2A-Hygromycin. The vector additionally contained the 5′ and 3′ LTRs necessary for lentiviral processing and integration and a WPRE element (Woodchuck Hepatitis Virus (WHP) Post-transcriptional Regulatory Element) for enhanced stability of the viral mRNA transcript (Figure 1A).

LP donor vectors for CRISPR-Cas9 integration were constructed by appending homology arm sequences (typically ∼0.5–1 kb long; Supplementary Table S5) to an LP expression cassette void of lentiviral elements. Homology arms were synthesized as a single gBlock (Integrated DNA Technologies) containing a PmeI restriction site between the left (LHA) and right (RHA) homology arms, and BsaI cleavage sites in 5′ and 3′ termini for Golden Gate cloning (26). Each gBlock was cloned into a pIDTsmart vector (Integrated DNA Technologies) modified to contain compatible BsaI cloning sites. The hEF1a-BxB1-attP-EYFP-P2A-Hygro LP cassette was polymerase chain reaction (PCR) amplified from the pLV-LP lentiviral vector and cloned into a PmeI linearized pIDTsmart backbone between the left and right homology arms using In-Fusion cloning (Clontech) to create the LP1 donor vector. The LP2 donor vector, containing a hEF1a-BxB1-attP-EBFP-P2A-Blasticidin cassette, was generated from LP1 using In-Fusion cloning.

Genetic constructs encoding light and heavy chains (LC and HC, respectively) of human monoclonal antibody (mAb) JUG444 (Pfizer) were constructed using modular Gateway/Gibson assembly (27). Promoter sequences of hEF1a and Cytomegalovirus (CMV) were cloned into a Gateway entry vector between attL4 and attR1 sites using In-Fusion cloning, and JUG444 light and heavy chain genes were cloned into pDONR221 using Gateway BP cloning (Life Technologies). Gateway LR reactions were performed between promoter and gene entry vectors and pZDonor_Seq(*n*)-GTW-Seq(*n+1*)_R4_R2 destination vectors to generate expression vectors (transcription units in different position vectors). Gibson reactions were performed using the Gibson Assembly Ultra Kit (SGI-DNA) using equimolar concentrations (∼40 fmols per 10 µl reaction) of column-purified expression vectors, adaptor vector

and carrier vector cleaved with I-SceI, XbaI and XhoI, and FseI, respectively. The 3x-hEF1a-mAb payload was constructed from the 2x-hEF1a-mAb vector using hierarchical assembly (27). Gibson reactions were transformed into E. cloni 10G electrocompetent cells (Lucigen) and grown at 30°C in LB media supplemented with ampicillin (100 μg/ml) and kanamycin (50 μg/ml). After verification by restriction mapping analysis and sequencing, correctly assembled constructs were expanded in Stbl3 cells (Invitrogen).

Lentiviral infection

The pLV-LP vector was packaged into lentiviral particles in HEK293FT cells and used to infect adherent CHO-K1 cells as described previously (15). Aiming at a single integration event per cell, we used a low multiplicity of infection (MOI) of 0.05 and 0.005 in two separate rounds of infection. The infection was followed by fluorescence-activated cell sorting (FACS) of EYFP expressing cell pools, cell recovery and propagation for 1–2 weeks to expand and enrich for stably expressing cells. Subsequent FACS was performed to select single clones from the top ∼10 and ∼15% of EYFP expressing cells from the two infection rounds, respectively. Clones were expanded and analyzed by flow cytometry and Droplet Digital PCR (ddPCR). Cell lines with robust and homogenous EYFP expression from a single copy of the EYFP and hygromycin genes were evaluated with genome walking analysis and tested for their ability to express a mAb (JUG-444 payload integration and titer assay). One of the first clones identified, mLP1, turned out to have two integration events, and was also included in further analysis.

Identification of lentiviral integration sites by genome walking

Lentiviral integration sites were identified by genome walking using the Lenti-X™ Integration Site Analysis Kit (Clontech) following manufacturer instructions. Briefly, genomic DNA was isolated from 5×10^6 cells, digested with restriction endonucleases (2 h, 37°C) and column purified. Four separate reactions were performed for each cell line: DraI, SspI, HpaI and a positive control that included human genomic DNA that was digested with DraI. GenomeWalker adapters were ligated to the blunt ended restriction digestion products by incubating 1.9 μl of 25 uM adapters, 0.8 μl of 10× ligation buffer and 0.5 μl of T4 DNA ligase (6 units/μl) at 16°C overnight. Reactions were stopped by the addition of 32 μl of TE (pH 7.5) and incubation at 70°C for 5 min. Two rounds of nested touchdown suppression PCR were performed to maximize the specificity of amplification of integrated lentiviral sequences. PCR products were separated on an agarose gel and major bands were excised, purified and sequenced using PCR primers. The resulting DNA sequence data was analyzed to determine the lentiviral insertion site. The portion of the sequence composing the flanking genomic sequence was searched against the Chinese hamster ovary (CHO) cell genome using BLAST (28), from which the integration loci were determined.

Identification of Rosa26 locus in CHO-K1

The murine Rosa26 locus has been described as a 'safe harbor' site for heterologous gene expression and used extensively for genetic engineering in mice and human cells (29,30). To identify the putative Rosa26 site in CHO we performed a BLAST search on the CHO-K1 genome with the murine Rosa26 locus (NC_000072, *Mus musculus* strain C57BL/6J, chromosome 6, GRCm38.p2). We identified a ∼9 kb homologous locus (∼63% nt sequence identity) in scaffold 2241 of CHO-K1 located downstream of the Thumpd3 gene in reverse orientation. In CHO-K1 this locus is part of a large (>2 Mb) unannotated region. CRISPR-Cas9 was used to target LP integration into the Rosa26 locus 8 kb downstream to the last exon (exon 10) of Thumpd3 gene in reverse orientation, resulting in the mono- and bi-allelic Rosa26 LP clones named sLP20 and biLP20, respectively.

Bioinformatics analysis of the integration loci

Enrichment of the newly identified integration loci across different KEGG pathways (31) was calculated using the WebGestalt toolkit (32). The significance of the over-representation was assessed using a hypergeometric test. mRNA expression levels were compared across three groups—integration sites corresponding to coding genes, housekeeping genes and other (non-housekeeping) genes—using published RNA-seq data from CHO-K1 (33) as well as newly generated data (manuscript in preparation). For RNA-seq analysis, clonally propagated CHO-K1 cells with random integration of a model monoclonal antibody JUG444 were grown in fed-batch bioreactors using commercially available CD CHO media (Thermo Fisher). Cell cultures were sampled and centrifuged into cell pellets of about 1×10^6 viable cells, and RNA extracted using Trizol reagent (Thermo Fisher) and flash frozen in a dry ice/ethanol bath. A total of 200 ng of total RNA was used for strand specific library construction using a TruSeq Stranded mRNA Sample Preparation Kit (Illumina). The resulting 160 bp insert library was quantified using qPCR and the insert size of the library was measured using an Agilent 2100 Bioanalyzer. Samples were sequenced using an Illumina HiSeq 2000, and the processed RPKM values were used for the analysis. A previously established compendium was used to define housekeeping genes (34). The significance of the difference in expression levels across the three different groups was evaluated using a Mann–Whitney U test.

Landing pads construction by homologous recombination with CRISPR/Cas9

Donor vector sequences for homologous recombination were derived from the public CHO-K1 genome database (35,36) based on the identified lentiviral integration sites. First, guide RNA sequences were designed using the CasFinder/CasValue algorithm or a CRISPR design tool such as http://crispr.mit.edu/ (37) typically within 1 kb of the lentiviral integration sites. Then, 0.5–1 kb of genomic sequences 5′ and 3′ of the gRNA target site were used as the homology arms flanking the LP sequence to

generate LP donor vectors. Targeted integrations were performed by co-transfecting the circular LP donor vector (100–500 ng) with Cas9 and gRNA using three different methods. Method 1 included the cloning of a gRNA sequence into the px330-U6-chimeric_BB-CBh-hSPCas9 vector (Addgene plasmid 42230, a gift from Feng Zhang) (38) using Golden Gate assembly (26) with BbsI. The resulting vector px330-gRNA, co-expressing a gRNA and human codon optimized *Streptococcus pyogenes* Cas9, was co-transfected at a range of concentrations (typically 10–50 ng) along with the LP donor vector. Method 2 included co-transfection of 50–200 ng px330-hSPCas9 expressing vector (Addgene plasmid 42230) (38) and 100–150 ng gRNA GeneArt DNA String (ThermoFisher). Method 3 included co-transfection of 0.5 µg Cas9 mRNA (Invitrogen) and 100–150 ng gRNA GeneArt DNA String (ThermoFisher). Out of 10 Cas9-constructed LPs, 5, 3 and 2 clones were constructed with method 1, 2 and 3, respectively, with no noticeable differences in the efficiency and specificity of the genomic modifications. About 10^5 cells were transfected in duplicates using a Neon electroporator (Invitrogen) and seeded in a 24-well plate. Three days post-transfection, cells were transferred to a 6-well plate and subjected to antibiotic selection with either hygromycin (600 µg/ml) or blasticidin (10 µg/ml) for 2 weeks followed by clonal sorting with FACS. Alternatively, cells were expanded and cultured for 8–10 days and directly subjected to clonal FACS based on EYFP expression. Single clones were verified with a number of genomic tests, including diagnostic PCR, ddPCR and Southern blot.

Cell culture and transfections

Adherent CHO-K1 cells (ATCC) were maintained in complete HAMS-F12 (cHAMS-F12) medium (ATCC) containing 10% fetal bovine serum (Sigma-Aldrich), 1% Hy-Clone non-essential amino acids (GE Healthcare Life Sciences) and 1% penicillin/streptomycin (Gibco). Cells were grown in a humidified 37°C incubator with 5% CO_2 and passaged every 2–3 days. Transfections were carried out with the Neon electroporation system (Invitrogen) using 10 µl Neon tips and 1×10^5 cells transfected in duplicate with 1560 V/5 ms/10 pulses, and transferred to a 24-well plate containing cHAMS-F12 without antibiotics. For CRISPR/Cas9 targeted integration of LPs, cells were co-transfected with the circular LP donor vector (100–500 ng) and either px330-hSPCas9-gRNA (10–50 ng; Addgene plasmid 42230 with the cloned gRNA sequence), or px330-hSPCas9 (50–200 ng; Addgene plasmid 42230) (38) and gRNA GeneArt DNA String (100–150 ng; ThermoFisher), or Cas9 mRNA (0.5 µg; Invitrogen) and gRNA GeneArt DNA String (100–150 ng; ThermoFisher). Three days post-transfection cells were transferred to a 6-well dish and the media was changed to cHAMS-F12 supplemented with either hygromycin (600 µg/ml) or blasticidin (10 µg/ml) depending on LP configuration. Cells were maintained under selection for 2 weeks, with antibiotic-containing media refreshed every 2–3 days. For integration of mAb expressing payloads, 1×10^5 cells were transfected in duplicate with 500 ng BxB1 expressing plasmid (pEXPR-CAG-BxB1) and 0.5–1 µg mAb vector under the same parameters as above,

and transferred to a 24-well plate containing cHAMS-F12 without antibiotics. Three days post-transfection cells were transferred to a 6-well dish and the media was changed to cHAMS-F12 supplemented with the following antibiotic combinations. Single LP integrants were selected with puromycin only (8 µg/ml), and multi-LP integrants were selected with puromycin (8 µg/ml or 20 µg/ml for mAb payload integrated into a single or multiple LPs, respectively), or puromycin and hygromycin, or puromycin and blasticidin to select cell pools with mAb payload integrated into different LPs. Following 2 weeks of selection, correct integration was confirmed by the complete disappearance of LP fluorescence (either EYFP or EBFP or both). Cells were maintained either as a pool of mAb integrants or sorted into clonal populations by FACS. Viability and cell density were monitored with a Vi-CELL automated cell viability analyzer (Beckman-Coulter).

Flow cytometry and single-cell cloning

Cell fluorescence was analyzed using an LSRFortessa flow cytometer (BD Biosciences), equipped with 405, 488 and 561 nm lasers. A total of 30,000 events were collected for analysis, using the 488 nm laser and a 530/30 nm band-pass filter for EYFP and the 405 nm laser and 450/50 filter for EBFP. SPHERO RCP-30-5A Rainbow Calibration Particles, eight peaks (Spherotech) were used for fluorescence normalization and conversion of the measured fluorescence intensities to absolute units of fluorescence (MEFL, Molecular Equivalents of Fluorosceine). Data analysis was performed with FACSDiva software (BD Biosciences) and FlowJo (FlowJo LLC). Cell sorting was performed on a FACSAria cell sorter. Untransfected CHO-K1 cells were used for setting morphological gates. After LP integration, EYFP and EBFP positive cells were sorted into 96-well plates. After mAb payload integration, different sorting schemes were applied, including sorting of single positive (EYFP+/EBFP− and EYFP−/EBFP+) and double-negative (EYFP−/EBFP−) cells to select multi-LP clones with mAb payload integrated into different LPs. In both cases, single cells were sorted into 96-well plates, and expanded to 24-well and then 6-well plates.

Landing pad gene expression stability studies

For stability studies, LP cells were maintained in 6-well plates in cHAMS-F12 for 1–3 months with no antibiotic (hygromycin or blasticidin) selection. Cells were passaged every 3–4 days and analyzed weekly with flow cytometry, from which mean fluorescence (EYFP or EBFP) and the percentage of positive cells were calculated. CHO-K1 untransfected cells were used for setting morphological gates. SPHERO RCP-30-5A Rainbow Calibration Particles, eight peaks (Spherotech) were routinely used during the stability study for fluorescence normalization and MEFL calculation. A doubling time of 18 h was used to calculate generation number of adherent CHO-K1 derived clones.

Genotypic analysis of landing pads and mAb integrants

Gene copy analyses were performed by ddPCR using the QX100 system (Bio-Rad Laboratories). Genomic DNA was

digested with CviQ1 and diluted 10-fold. Two target genes were quantified for each probe: EYFP and hygromycin or EBFP and blasticidin for LP clones, and LC and HC for mAb integrants. Gene copy numbers were normalized to the CHO endogenous housekeeping gene Cog1. Two to three runs were used for each sample.

For diagnostic PCR, genomic DNA was isolated with a QIAamp genomic DNA kit (Qiagen), and PCR amplified with PfuUltra II Fusion HS DNA Polymerase (Agilent). For LP verification, internal LP primers facing outwards were paired with external primers in the flanking genomic sequence (on-target), or with primers in the vector backbone (off-target). For mAb payload integration, a payload specific primer was paired with the LP-specific primer (on-target) or with the vector backbone specific primer (off-target).

For Southern blots, genomic DNA was quantified with a Qubit fluorometer and the Qubit dsDNA BR kit (Invitrogen). Probes were generated by PCR using appropriate primer sets, and were labeled with dCTP (α-^{32}P) (Perkin Elmer) using the Prime-It RmT Random Primer Labeling Kit (Stratagene), followed by removal of free nucleotides using Quick Spin Columns for radiolabeled DNA purification Sephadex G-50 (Roche). For LP clones, DNA was digested with BamHI, HindIII and EcoRV and probes for EYFP, EBFP, hygromycin and blasticidin were used, and for mAb integrants DNA was digested with SpeI and XbaI and probes for LC and HC were used.

Antibody expression from stably integrated landing pads

mAb payload vectors expressing the JUG444 human monoclonal antibody (Pfizer) were integrated into LP cell lines, selected for two weeks with puromycin (8 μg/ml or 20 μg/ml for mAb payload integrated into a single or multiple LPs, respectively) and maintained either as cell pools or as clonal populations following FACS. For mAb expression stability analysis, a 6-well master plate of cells was continuously propagated in cHAMS-F12 and used weekly for seeding cells for a mAb expression assay. 1.5×10^5 cells were seeded in 2–3 replicates in a 24-well plate with 0.5 ml cHAMS-F12 media and grown at 37°C. On day 4, conditioned media was collected and the amount of secreted JUG444 was measured in duplicates with an Octet RED96 system using anti-human IgG Fc Capture (AHC) or Protein A biosensors (ForteBio), with no difference in the measurements observed between the sensors. Purified JUG444 was used to generate a standard curve, from which mAb titers were derived.

RESULTS

Discovery of stable landing pad integration sites in the CHO-K1 genome

To discover stable integration sites in the CHO-K1 genome, we used a random lentiviral screen with an LP payload. Lentiviral integration was chosen for screening loci because it efficiently selects for actively transcribed genomic sites (6–9) and provides a rough means for controlling copy number through MOI. The LP cassette contained a constitutive promoter driving co-expression of a fluorescent reporter (EYFP) and a selection marker (hygromycin), as well as the attP phage attachment site for the BxB1 recombinase (15) (Figure 1A). Using an LP with a fluorescent reporter as a lentiviral payload allowed us to monitor the initial infection and select stable integrants with consistent long-term expression, then use BxB1-mediated DNA recombination for site-specific insertion of additional genes at the integration locus.

The LP vector was packaged in lentiviral particles and used to infect adherent CHO-K1 cells at low MOI (0.05 and 0.005), aiming at a single integration event per cell. Infection was followed by hygromycin selection and FACS of EYFP expressing cell pools. Pools of expressing cells were then propagated for ~2 weeks with no hygromycin selection in order to select for stably expressing cells that maintain robust LP expression in the absence of antibiotic selection. Stable cell pools were then re-sorted into clonal populations of the highest EYFP expressing cells using two expression thresholds: ~3% top clones and ~10–15% top clones from the two infection rounds (Figure 1B). Sorted clones were expanded and analyzed for EYFP expression and gene copy number to select stable clones with a single integration event and a strong, homogeneous EYFP expression profile (Figure 1B and C). We aimed to discover stable single integration clones with strong and consistent expression. Most high expressing clones (~3% top expressers) turned out to have multiple integrations and were discarded. In the group selected from ~10 to 15% top expressers, clones obtained using a higher MOI (0.05) contained ~50% single integrants, while infections at lower MOI (0.005) yielded only single integrants. A total of 18 monoclonal single integration clones were isolated from lentiviral screens (named mLP2 to mLP19; Figure 1C). A double-integration clone, mLP1, was also isolated for further analysis (Figure 1C). Selected clones were further analyzed by Southern blots, confirming ddPCR gene copy results (Supplementary Figure S1).

In addition to the random lentiviral screen, we also searched the CHO-K1 genome for regions homologous to the Rosa26 locus, which has been widely used for genetic engineering in mice and human cell lines (29,30). We performed a BLAST search of the murine Rosa26 locus (29) against the CHO-K1 genome scaffold and identified a ~9 kb homologous region in CHO-K1 (63% sequence identity) (Supplementary Figure S2). Similar to the murine Rosa26 locus, the homologous region in CHO-K1 is located downstream of the Thumpd3 gene in the reverse orientation. In CHO-K1 this locus is part of a large (>2 Mb) unannotated region. We used CRISPR-Cas9 to integrate an LP 8 kb downstream of the last exon of the Thumpd3 gene in reverse orientation, and thus obtained clones containing a single copy (mono-allelic) and a double copy (bi-allelic) of the Rosa26 LP which we named sLP20 and biLP20, respectively (Figure 2B and C; discussed below).

To assess the stability of gene expression from the newly discovered loci, we propagated 18 monoclonal single-integration lentiviral mLPs for 1.5–2 months (65–75 generations) in the absence of antibiotic selection (hygromycin) and measured their EYFP expression weekly using flow cytometry. Most clones exhibited very stable EYFP expression (typically, 96–100% positive cells and no drop in ex-
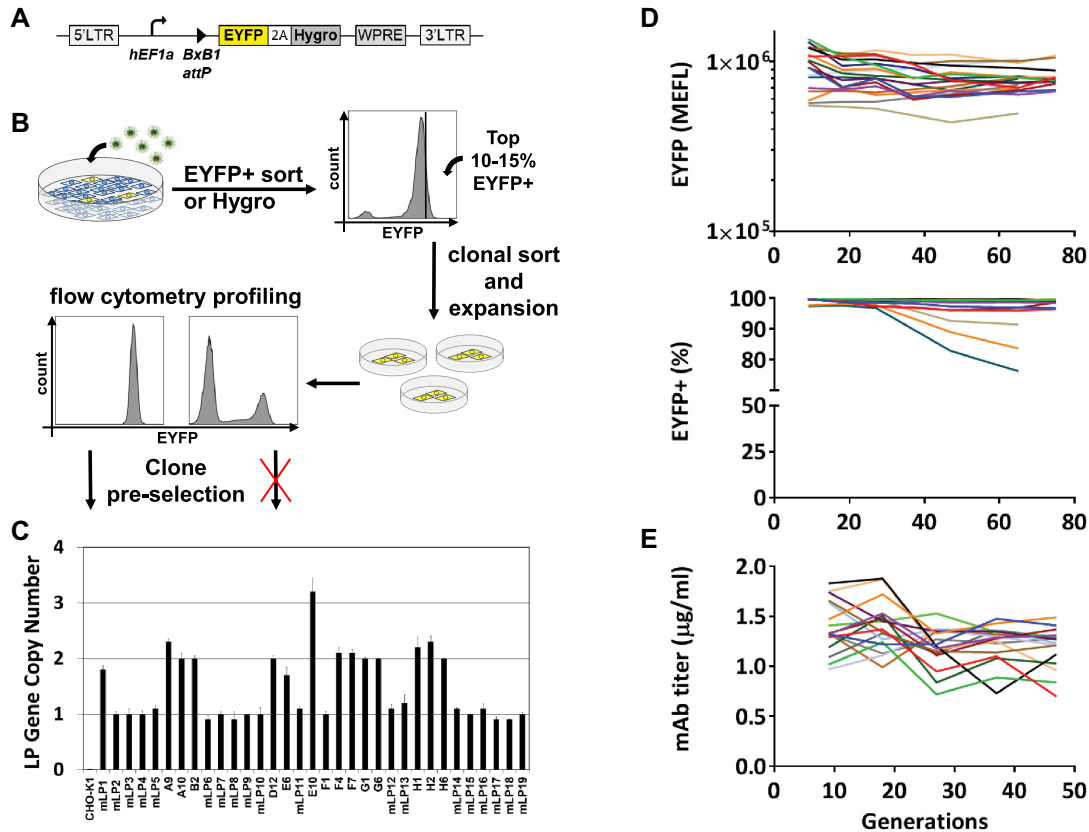
Figure 1. Twenty stable integration sites were discovered in CHO-K1 by lentiviral screen with an LP probe. (**A**) Schematic diagram of the lentiviral LP integration vector (pLV-LP) used for the lentiviral screen. Key components include 5′ and 3′LTRs necessary for lentiviral processing and integration, an attP attachment site for the BxB1 recombinase, a hEF1a constitutive promoter driving expression of an EYFP reporter and a hygromycin (Hygro) selection marker that are co-expressed using a 2A translation skip peptide, and a WPRE element (WHP Post-transcriptional Regulatory Element) for enhanced stability of the viral mRNA transcript. (**B**) Lentiviral infection and selection of candidate LP clones. Lentiviral particles were packaged with the LP vector and used to infect adherent CHO-K1 cells at a low MOI (0.05 and 0.005). EYFP expressing clones were picked with FACS (top ∼10 and ∼15% EYFP expressing cells), expanded and analyzed by flow cytometry. Clones with homogenous fluorescence profiles were isolated and subjected to further analysis. (**C**) ddPCR gene copy analysis of LP clones. Vertical axis denotes copy number for EYFP gene normalized to the housekeeping gene Cog1. Candidate clones isolated from the lentiviral screen were assessed by ddPCR analysis to select for single integration monoclonal LPs (named mLP2 to mLP19). Note that mLP1 is a double integration clone, and the F1 clone with a gene copy number of 1 was discarded due to genomic instability. (**D**) EYFP expression stability analysis of 18 lentiviral single integration monoclonal LPs. Cells were propagated for 2 months in the absence of antibiotic selection. Genetic stability was assessed weekly by monitoring EYFP expression with flow cytometry from which mean EYFP fluorescence intensity (top panel) and percentage of EYFP-positive cells (bottom panel) were derived. Fifteen clones exhibited >96% EYFP-positive cells after 2 months of stability assay, while three clones exhibited moderate stability (∼ 75–90% positive cells; see Supplementary Figure S3A-B and Table S1 for details). (**E**) Antibody expression stability analysis by pools of mLP clones integrated with the 1x-hEF1a-mAb circuit. The mAb circuit was integrated with a BxB1 recombinase, and pools of LP-mAb integrants were selected with puromycin. mAb titers were assayed weekly in conditioned media from polyclonal cell pools propagated without selection. Each data point represents one week of cell propagation and mean titer measurement with biological duplicates or triplicates (*n* = 2 or 3). A doubling time of 18 h was used to calculate the generation number of adherent CHO-K1 derived clones. The same symbols are used in panels D and E (see Supplementary Figure S4B and Table S1 for details).

pression levels during 2 months of stability study), while three clones exhibited stable mean EYFP expression levels, but a moderate drop in percentage of EYFP-expressing cells (∼75–90% positive cells after 2 months of stability assay; Figure 1D; see Supplementary Figure S3A-B and Table S1 for details). Southern blot analysis of selected clones from early and late passages (generation 9 and 65, respectively) also confirmed genomic stability of the LPs during at least 2 months of cell propagation without selection (Supplementary Figure S1).

Interestingly, despite the fact that all the LPs were located in different genomic contexts, there was little variation in EYFP levels between the LP clones, with most LPs show-

ing similar fluorescence levels and stability (Figure 1D; Supplementary Figure S3A-B and Table S1). Similar expression levels may result from the fact that transcription from our LP is driven by the same constitutive promoter, with the local genomic environment only having a minor effect. We also observed no noticeable difference in the growth rate or morphology of LP clones compared to wild-type CHO-K1 cells.

To assess the expression of recombinant proteins from the lentiviral LPs, we used BxB1-mediated recombination to integrate a DNA vector encoding a single copy of the heavy and light chain of human monoclonal antibody (mAb) JUG444 (1x-hEF1a-mAb vector; Supplementary Figure
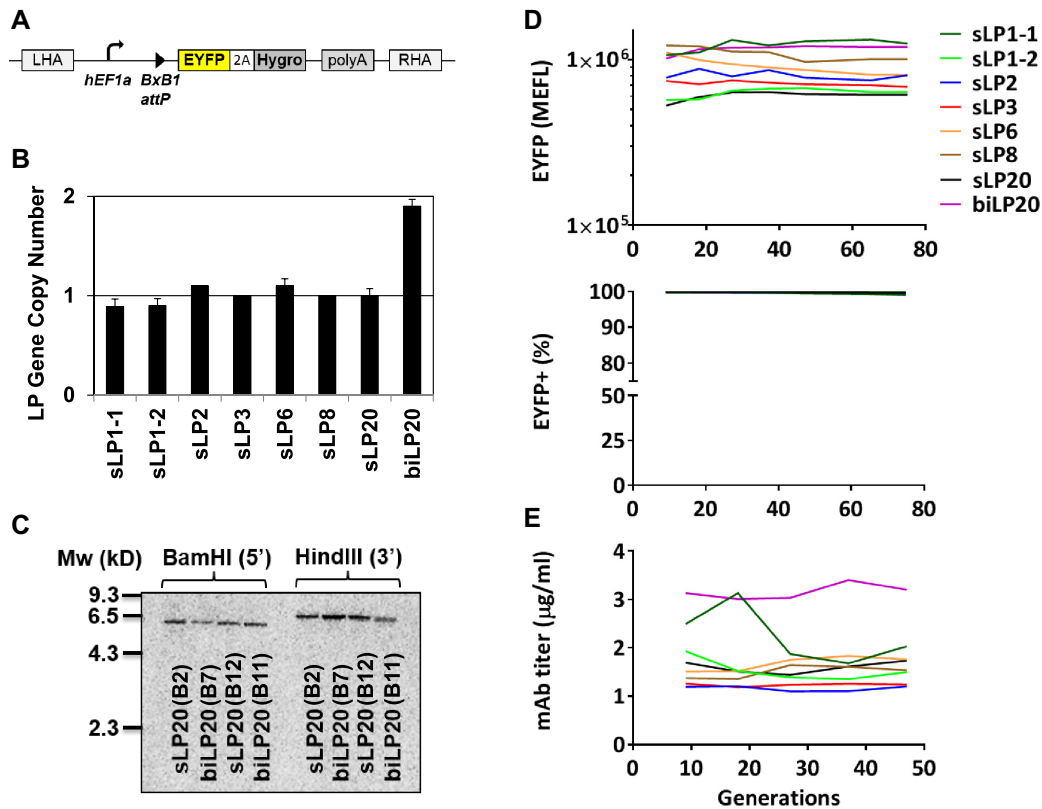
Figure 2. LP cell lines reconstructed with CRISPR/Cas9 maintain genetic stability and stable payload expression. (**A**) Schematic diagram of the LP donor vector used for CRISPR/Cas9 targeted insertion into the newly discovered stable integration sites of WT CHO-K1 cells. The LP cassette is similar to the lentiviral LP integration vector (pLV-LP, Figure 1A), but it includes locus-specific left and right homology arms (LHA and RHA) instead of lentiviral LTRs, and a polyA signal sequence instead of a WPRE. Single integration LP clones (sLPs) were constructed by targeted insertion of the LP probe into the genomic loci identified in the corresponding lentiviral mLP clones. (**B**) ddPCR gene copy analysis of selected LP clones constructed with CRISPR/Cas9. sLP1-1 and sLP1-2 are single integration lines bearing an LP cassette in the loci derived from the double integration clone mLP1. sLP2, sLP3, sLP6 and sLP8 are the reconstructed homologs of the lentiviral mLP clones. sLP20 and biLP20 are mono-allelic and bi-allelic integrants in the newly identified Rosa26 locus. Vertical axis denotes copy number for EYFP gene normalized to the housekeeping gene Cog1. (**C**) Southern blot analysis of sLP20 and biLP20 clones. gDNA was cleaved with BamHI (5′) or HindIII (3′) restriction enzymes and hybridized with an anti-hygromycin radiolabeled probe. Two sLP20 and two biLP20 clones were analyzed (B2, B12 and B7, B11, respectively). (**D**) EYFP expression stability analysis. Cells were propagated for 2 months in the absence of antibiotic selection. Genetic stability was assessed weekly by monitoring EYFP fluorescence intensity (top panel) and the percentage of EYFP-positive cells (bottom panel) by flow cytometry. All clones exhibited >99% EYFP-positive cells after 2 months of stability assay. (**E**) Antibody expression stability analysis. Cell pools of LP-mAb integrants stably expressing the 1x-hEF1a-mAb circuit were propagated in the absence of selection, and mAb levels were measured weekly in conditioned media. Each data point represents 1 week of cell propagation and mean mAb measurements with biological duplicates or triplicates ($n = 2$ or 3). A doubling time of 18 hr was used to calculate the generation number of adherent CHO-K1 derived clones. The same symbols are used in panels D and E.

S4A) into selected LP clones. Site-specific integration was indicated by gain of payload-encoded puromycin resistance and the disappearance of LP-encoded EYFP expression. Cell pools of DNA integrants were then propagated for 5 weeks in the absence of selection and assayed weekly for mAb expression. Most mLPs, including three clones with moderate EYFP stability, exhibited stable mAb expression with no drop in titers for over a month (Figure 1E; see Supplementary Figure S4B and Table S1 for details). In agreement with low variability in EYFP levels, similar levels of mAb titers were also observed among the various mLPs. We observed a weak correlation between EYFP levels and mAb titers of mLP clones (Supplementary Figure S5A), either due to the different expression mechanisms of EYFP and mAb (i.e. intracellular expression of EYFP and secreted expression of mAb) or due to the lack of variation in EYFP

and mAb expression between the LPs. Overall, we conclude that the newly discovered loci allow efficient site-specific integration of DNA payloads and support stable long-term expression of transgenes (including secreted recombinant proteins).

Identification of stable LP genomic integration loci

We used genome walking analysis to identify the lentiviral integration sites in the stable LP cell lines. Genomic sequences flanking the lentiviral integration sites were searched against the CHO-K1 genome database to identify the integration loci for 18 single-integration mLPs, a double-integration mLP1 clone, and the newly identified Rosa LP (LP20) (Table 1; Supplementary Tables S2 and 3). We identified 21 unique integration sites, with 9 sites located

in unannotated or non-coding regions and 12 sites located in annotated regions that could be mapped to genes (11/12 are in introns). The genes belong to various functional families and have not been previously reported as stable integration sites in mammalian cells.

We then sought to explore possible common functional and transcriptomic features shared by these annotated loci. To this end, we first examined the distribution of human and mouse orthologs of these genes across KEGG pathways (31). We found that 3 of the 12 genes—Bmp5, Dcn and Smad6—are in the transforming growth factor beta (TGF-β) pathway, a significant over-representation compared to random distribution ($P = 1.7 \times 10^{-5}$, FDR = 0.005 using human orthologs; $P = 1.2 \times 10^{-5}$, FDR = 0.003 using mouse orthologs). We also looked for over-representation of gene signatures from the literature and found enrichment in a gene set that corresponds to a signature that discriminated different categories of adult renal progenitor/stem cells (39). The observed enrichment is due to six genes of our LP set, Dcn, Clcc1, Ssbp2, Smad6, Fam114a1 and Lrba. We also found that the integration sites generally have significantly higher expression levels than other non-housekeeping genes in CHO-K1 cells, but lower than house-keeping genes (Supplementary Figure S6). Together, these results suggest that the patterns observed for lentiviral integration are not random, which is in agreement with previous observations (5–9). One possible preference for integration may be the vicinity of actively transcribed genes; for example, activation of the TGF-β pathway under cell culture conditions was observed for various types of mammalian cells (40). Presumably the integration of a large payload into these genes would disrupt their expression by either disrupting a coding sequence or interfering with splicing, but additional gene copies on other allele(s) could compensate for this. Further analysis is needed to investigate whether LP insertion deregulates the expression of the integration genes and neighboring genes.

Reconstruction of selected landing pads with CRISPR/Cas9

Some strategies for engineering cell lines require multiple large-scale integrations, which would be greatly facilitated by cell lines harboring multiple LPs. These cell lines can be produced using site-specific integration into the newly discovered stable integration loci. Thus, we sought to test whether the new integration sites can be utilized for the targeted (non-viral) integration and reconstruction of the LP cell lines and whether the resulting cell lines retain the expression stability of the lentiviral LPs. For this objective we engineered seven single LP cell lines (sLPs) using CRISPR/Cas9, including sLP1-1 and sLP1-2 (single integration LPs in loci 1 and 2 of the double integration clone mLP1), sLP2, sLP3, sLP6, sLP8 and sLP20 (Figure 2A–C; see Supplementary Table S4 for the complete list of all LP clones and their configurations). Integration sites were chosen based on the stability and expression levels of EYFP and mAb (Figure 1D and E), but given the low variation between the LPs our selection was somewhat arbitrary. We designed insertion sites to be within 1 kb of the integration loci identified in the lentiviral clones, with the main differences between an sLP and the corresponding mLP clone be-

ing the presence of viral LTRs that flank the integrated LP probe in the lentiviral mLPs and the precise location of integration due to Cas9/gRNA genome targeting constraints (Figures 1A and 2A; Supplementary Table S2).

Following Cas9-mediated insertion and selection for LP expression, we performed FACS to generate clonal cell lines and selected stable clones based on fluorescent reporter expression. Selected clones were verified with a number of genomic tests, including gene copy analysis and Southern blot (Figure 2B and C) and diagnostic PCR for on-target and off-target integration (Supplementary Figure S7A). Correct Cas9 mediated integration efficiency varied for different loci, ranging from ∼5% (sLP1-1) to as high as ∼50% (sLP20) in selected stably expressing clones. For the Rosa26 site, in addition to the single integration clones (sLP20, clones B2 and B12), we also generated bi-allelic integrants bearing a double copy of the LP in two alleles of the putative Rosa26 locus (biLP20, clones B7 and B11) (Figure 2B and C).

We used the same methods to characterize gene expression from the targeted LP lines that we developed to characterize the lentiviral insertion lines. Fluorescent protein expression was stable during 2 months of cell propagation in the absence of antibiotic selection and exhibited expression levels similar to their mLP homologs (Figure 2D and Supplementary Figure S7B). Interestingly, expression levels from Cas9-engineered sLPs were even more stable than from the lentiviral mLPs, showing a tight and homogeneous distribution of stability levels (>99% of EYFP+ cells after 2 months of cell culturing in the absence of selection; Figure 2D) and a weak correlation between EYFP levels and mAb titers (Supplementary Figure S5B). Notably, the bi-allelic Rosa26 LP (biLP20) was as stable as the mono-allelic Rosa26 LP (sLP20), but exhibited ∼2-fold higher EYFP fluorescence due to an extra LP copy, suggesting that a combination of stable single LPs can yield a stable multi-LP cell line with proportionally increased expression levels.

We also integrated the JUG444-expressing payload into the CRISPR/Cas9-generated LP lines and quantified antibody titers in pools of mAb integrants (Figure 2E and Supplementary Figure S7C). All clones exhibited stable mAb expression during one month of cell propagation in the absence of selection and titers were similar to those observed with the corresponding mLPs. As with EYFP expression, biLP20 exhibited stable mAb expression with ∼2-fold higher titers compared to LP20 due to an extra mAb copy integrated into the second LP. Overall, this data confirms that stable integration loci discovered with a lentiviral screen can be successfully reconstructed with CRISPR-Cas9, and the properties of the lentiviral clones (such as long-term stable expression of the LP and the integrated payload) are recapitulated in the engineered clones.

Construction and utility of a multi-LP platform for designer cell lines

After successfully reconstructing stable single LPs using a gene editing approach, we focused on generating cell lines bearing LPs integrated in multiple stable genomic sites. These designer cell lines may form a foundational platform for site specific chromosomal integration of various

Table 1. Landing pad integration loci in CHO-K1 genome

LP clone	Locus (scaffold, position)	Gene	Relative orientation (to genome scaffold, to gene landmark)	Intron/exon	Gene function
sLP1-1[a]	scaf 1924 pos 60148	unannotated region	reverse	non-coding	
sLP1-2[a]	scaf 1361 pos 1075377	unannotated region	reverse	non-coding	
mLP2	scaf 1558 pos 414422	unannotated region	reverse	non-coding	
mLP3	scaf 5035 pos 1864526	unannotated region	forward	non-coding	
mLP4 (A5)	scaf 934 pos 837184	Bmp5	forward, reverse	intron 1	TGF-β pathway
mLP5 (A8)	scaf 2625 pos 237547	Ssbp2	reverse, forward	intron 9	genome stability, tumor suppression, cell growth
mLP6 (B5)	scaf 156 pos 5618538	Trmt6	reverse, reverse	intron 8	tRNAs post-translational modification
mLP7 (C2)	scaf 424 pos 177811	unannotated region	reverse	non-coding	
mLP8 (C5)	scaf 2259 pos 86304	Clcc1	forward, reverse	exon 11	chloride channel in intracellular compartments
mLP9 (C10)	scaf 3405 pos 4675	Fam114a1 (Noxp20)	reverse, reverse	intron 1	apoptosis, regulation of cell proliferation
mLP10 (D9)	scaf 1727 pos 865981	Lrba	reverse, reverse	intron 40	intracellular vesicle transport, secretion of immune effector molecules
mLP11 (E8)	scaf 1385 pos 898266	unannotated region	reverse	non-coding	
mLP12 (G8)	scaf 1485 pos 438214	Dcn	reverse, forward	intron 2	cellular matrix proteoglycan, tumor suppression, TGF-β pathway
mLP13 (G11)	scaf 2165 pos 931631	Cep128	reverse, forward	intron 23	centrosomes-associated cell division
mLP14 (PL1-2)	scaf 832 pos 117635	Aacs	reverse, reverse	intron 5	fatty acid-synthesis during adipose tissue development
mLP15 (PL1-3)	scaf 262 pos 1571782	Aldh5a1	forward, reverse	intron 4	breakdown of the neurotransmitter GABA
mLP16 (PL1-10)	scaf 1796 pos 188123	Smad6	forward, forward	intron 2	transcriptional regulation of BMP and TGF-β signaling pathways
mLP17 (PL1-11)	scaf 624 pos 2206187	unannotated region	forward	non-coding	
mLP18 (PL1-17)	scaf 1753 pos 117475	unannotated region	reverse	non-coding	
mLP19 (PL1-20)	scaf 1149 pos 49532	Ptprq	reverse, forward	intron 8	cellular proliferation and differentiation
sLP20[b] (Rosa-B12)	scaf 2241 pos 132202	Rosa26 locus (downstream to Thumpd3)	reverse, reverse	non-coding	

[a]sLP1-1 and sLP1-2 are the loci identified in the double integration clone mLP1 and constructed as single integration landing pads with CRISPR-Cas9.
[b]sLP20 was identified by sequence homology search of CHO-K1 genome with mouse Rosa26 sequence and constructed with CRISPR-Cas9.

DNA payloads into multiple chromosomal locations, enabling quick and efficient engineering of stable cell lines for various biological and biotechnological applications. To build multi-LP clones we used CRISPR/Cas9 to sequentially integrate several LP cassettes encoding different reporter/selection payloads (EYFP-2A-Hygro or EBFP-2A-Bla) flanked by locus-specific homology arms (Supplementary Figure S8A and Table S4). Four integration sites (LP1-2, LP2, LP15 and LP20) were chosen for multi-

LP construction based on the stability data of sLPs and mLPs (Figures 1 and 2) and the efficiency of Cas9-mediated integration into various loci. A number of multi-LP cell lines were generated, including the double LP lines dLP1 (LP2/LP15) and dLP2 (LP1-2/LP2) and a triple LP tLP1 (biLP20/LP2; Supplementary Figure S8A). All clones were confirmed by diagnostic PCR, gene copy analysis and Southern blot (Supplementary Figure S8B and C). Similar to previous results, expression stability studies showed

strong and stable EYFP and EBFP expression for all dLP and tLP clones for at least 2 months of cell culturing in the absence of selection (Supplementary Figure S9). This confirmed that genetically stable cell lines comprising a multi-LP platform can be constructed from various combinations of our newly discovered stable single LPs.

To test the utility of our multi-LP platform for recombinant protein production using targeted multi-site payload integration, we used recombinase-mediated DNA integration to insert the 1x-hEF1a-mAb payload (Supplementary Figure S4A) into the single, dual and triple LP cell lines, and measured antibody titers over time to assess relative protein expression levels and long-term stability. One-shot integration of mAb payload into double-LP or triple-LP lines yields a mixture of mAb integrants that can be separated into distinct populations by a combined antibiotic selection and/or FACS sorting strategy as shown in Figure 3. For example, selecting dLP1 integrants with puromycin and blasticidin yields an EBFP+ pool with a single mAb copy integrated into the LP2 locus but leaves the LP15 locus expressing EBFP-2A-blasticidin intact; selection with puromycin and hygromycin yields an EYFP+ pool with a single mAb copy integrated into the LP15 locus, while selection with puromycin only and sorting for EYFP−/EBFP− double negative cells yields double mAb integrants (Figure 3). Alternatively, each of the three different populations can be isolated by puromycin selection followed by FACS. This approach allowed us to specifically target the payload into the different loci of the multi-LP clones and thus to generate all possible LP-payload combinations in one transfection.

Selected and sorted cell pools, as well as clonal integrants isolated from each population, were propagated for 1 month without selection and assayed weekly for mAb titers (Supplementary Figure S10). All clones exhibited stable mAb expression, and the titers increased linearly as a function of the number of LP sites integrated with a mAb payload. Thus, double and triple integrants exhibited ∼2- and 3-fold higher titers compared to a single mAb integrant (Figure 4 and Supplementary Figure S10).

Based on the above observations of increased mAb production in multi-LP clones, we pursued two strategies for further boosting mAb expression: increasing mAb copy number by integrating multi-copy mAb payloads into the multi-LP platform, and using a stronger promoter to drive mAb expression. We constructed multi-mAb vectors expressing two and three mAb copies under the hEF1a promoter (2x-hEF1a-mAb and 3x-hEF1a-mAb, respectively) and a payload expressing a single mAb copy under a stronger promoter, CMV (Supplementary Figure S4A). Four mAb constructs (1x-, 2x-, 3x-hEF1a-mAb and 1x-CMV-mAb) were integrated into sLP2, dLP1 and tLP1 cell lines. Integration efficiency decreased as the size of the DNA payload and the number of the integrated LP sites increased, but it was high enough to occupy all sites of the multi-LP platform using one-shot integration (Figure 3). For example, integration of the triple mAb payload (3x-hEF1a-mAb; 33 kb) into the triple LP (tLP1) yielded ∼ 5, 2 and 1% of single, double and triple LP-mAb integrants, respectively. Pools of single, double and triple integrants were then selected by a combination of antibiotic selection and sorting as described above.

For higher accuracy titer measurements, we isolated clonal cell lines from the pools of mAb integrants. Clonal integrants were verified by diagnostic PCR, gene copy analysis, and Southern blot of selected clones (Supplementary Figures S11 and 12). Gene copy analysis confirmed most candidate LP-mAb integration clones with up to 6 mAb copies (15/20 correct), while only 1/6 candidate clones with nine mAb copies had the correct gene copy number, and the other five had a lower gene copy number (Supplementary Figure S12). This could have resulted from multiple factors including genetic instability of the highly repetitive 3x-hEF1a-mab payload, growth disadvantage of the 9-copy mAb integrants, or a higher percentage of incorrect integrants in the relatively low population of triple integrants for this large payload. However, all the verified multi-LP integrants with one to nine total mAb copies, including a tLP1-derived cell line integrated with three copies of the 3x-hEF1a-mAb construct that bears a total of ∼100 kb of heterologous DNA (21 transcriptional units total), exhibited stable long-term mAb expression (Figure 4) and normal growth rates with no morphological abnormalities. For the hEF1a constructs, mAb titers increased proportionally to the number of the integrated mAb cassettes, with the triple LP integrated with 3x-hEF1a-mAb expressing ∼9-fold higher titers compared to a single 1x-hEF1a-mAb copy integrated in a single LP (sLP2) (Figure 5). The stronger promoter in the 1x-CMV-mAb construct yielded ∼3-fold higher titer compared to the 1x-hEF1a-mAb payload, and the 1x-CMV-mAb integration into the double and triple LP lines resulted in additional titer increase (up to 3-fold) in good correlation with the number of mAb cassettes in the payload (Figure 5). Overall, this data demonstrates that cell lines harboring multiple LPs have utility for stably integrating large amounts of heterologous DNA in a fast and reproducible manner with control over transgene copy number and integration specificity. The linear increase of mAb titer with the use of our multi-LP platform integrated with multi-copy DNA payloads suggests that expression from the three LPs is independent under these conditions. Additionally, the proportionality between transgene copy number and mAb titer may represent a novel strategy for boosting the expression level of recombinant proteins in a controlled and predictable manner when other factors, such as the secretory pathway's capacity, do not limit cellular productivity.

DISCUSSION

In this study we used a lentiviral screen to discover 20 genomic integration sites that support efficient and stable heterologous gene expression in CHO cells. We also identified a putative Rosa26 site in the CHO-K1 genome. Using CRISPR/Cas9 we integrated an LP at several of the newly identified loci and recapitulated the same robust and homogeneous long-term expression of the integrated DNA payloads that we observed in the lentiviral cell lines. Finally, we created cell lines with multiple LPs and demonstrated their utility for recombinant protein expression. Such multi-LP cell lines allow for targeted integration of multiple large DNA payloads into well-defined chromosomal contexts,
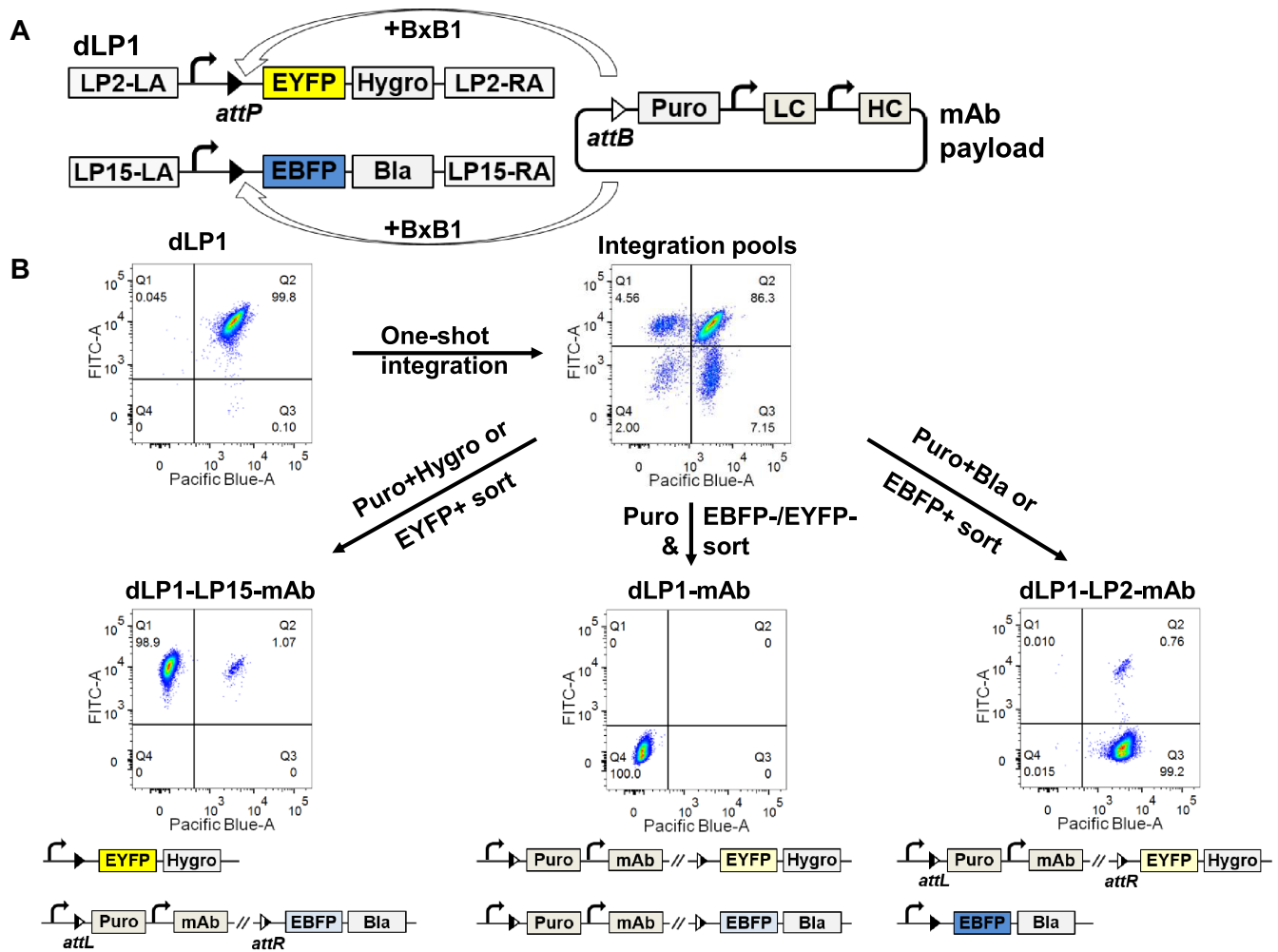
Figure 3. The multi-LP platform enables parallel construction of various payload expressing clones by one-shot payload integration. (**A**) Schematic representation of mAb payload integration into the dLP1 cell line. The dLP1 clone harbors LP2 and LP15 sites, each bearing the BxB1 attP site. BxB1-mediated recombination occurs between the LP's attP site and the payload's attB site, yielding a mixture of single integrants (LP2-mAb and LP15-mAb) and the double integrant (dLP1-mAb). (**B**) Flow cytometry diagrams and selection schemes of different pools of LP-mAb integrants following mAb payload integration into the dLP1 clone. Site-specific integration of a mAb payload into an LP site triggers expression of the payload's promoterless selection marker (puromycin), while expression of the LP cassette stops. Appropriate antibiotic selection schemes and sorting with FACS enable isolating three types of LP-mAb integrant pools where the mAb payload is integrated in only the LP2 site (LP2-mAb), only the LP15 site (LP15-mAb) or both (dLP1-mAb). Small residual populations of unintegrated LPs (e.g. an EYFP+ population in the LP2-mAb pool and an EBFP+ population in the LP15-mAb pool) can be removed by FACS sorting.

which may have broad applications in biomanufacturing and mammalian synthetic biology.

Genomic 'safe harbor' loci allow for predictable expression of integrated transgenes without adversely altering cellular function (41). We used lentiviral infection to screen for candidate integration loci because genome-wide studies have shown that lentiviruses preferentially insert in the vicinity of actively transcribed genes (6–9), and thus these sites may support stable expression of transgenes. Indeed, we found significantly higher expression of endogenous genes at lentiviral integration sites (Supplementary Figure S6) compared to other sites in the genome, which may indicate that these loci are relatively open and transcriptionally active. We also observed a significant over-representation of TGF-β signaling pathway among these genes, which may be the result of an activated epithelial–mesenchymal transition

response to the culture system (40,42). Interestingly, 20/21 integration sites are found in non-coding regions or gene introns. Similar enrichment of non-coding and intronic regions was observed in transcriptional hot spots in the human genome (43). Future analysis of a larger set of stable integration sites may help elucidate genomic features that make certain loci better LPs and could also lead to better criteria for identifying prospective stable integration sites in mammalian cells.

To validate the functionality of CHO cell lines harboring multiple LPs and their utility for biomanufacturing, we used them to engineer cell lines to produce monoclonal antibodies. Traditional development of mAb-producing cell lines via random integration often yields clones with heterogeneous, unstable expression of transgenes and necessitates labor intensive stability studies to
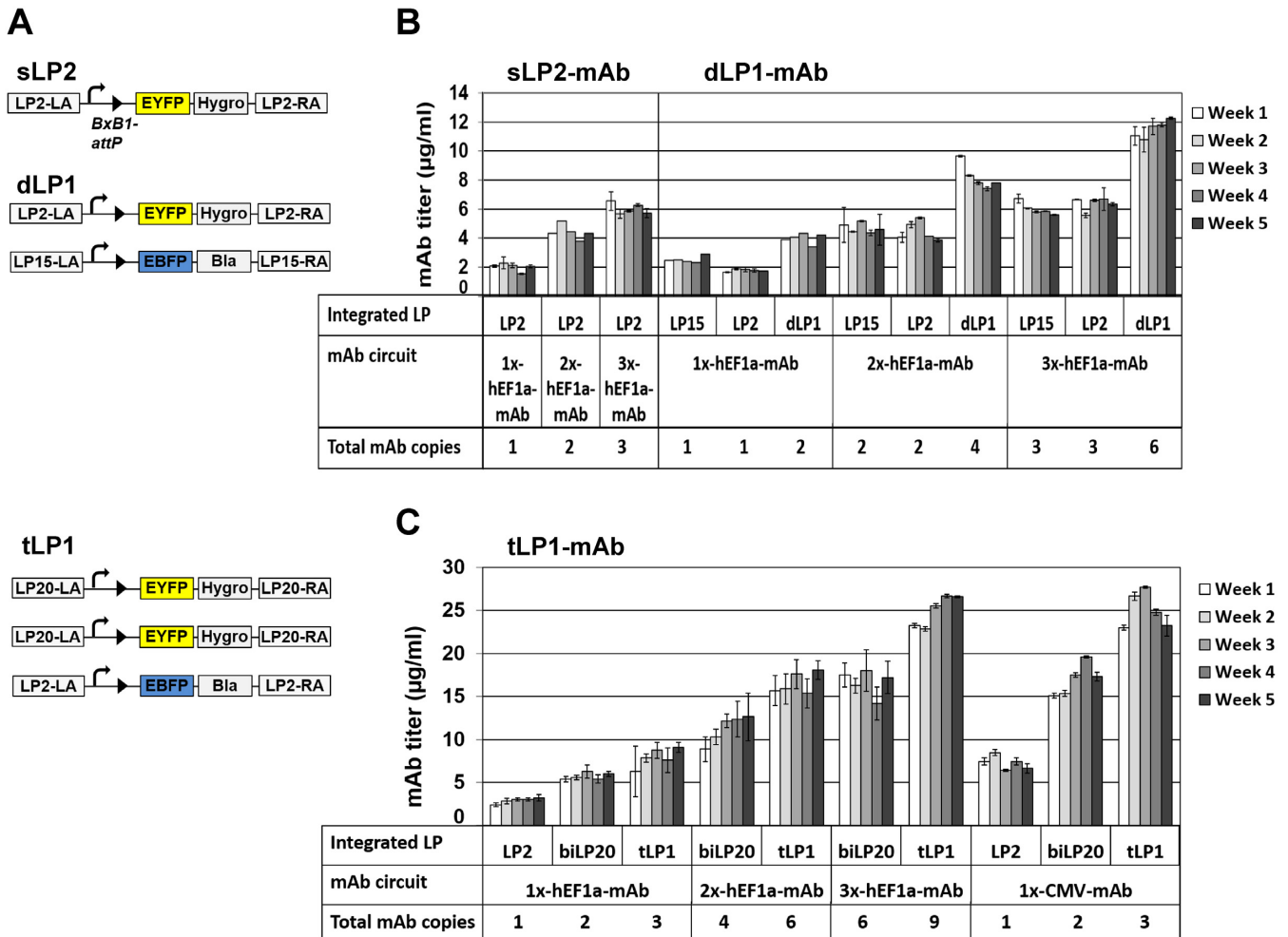
Figure 4. Stable and increased mAb expression with the use of multi-copy mAb circuits integrated in the multi-LP platform. (**A**) Schematic diagrams of LP configurations in single (sLP2), double (dLP1) and triple (tLP1) LP cell lines used for the multi-mAb payload integration. (**B**) Weekly mAb titers of clonal sLP2-mAb and dLP1-mAb integrants carrying multi-copy mAb constructs in different LP sites. mAb constructs were integrated into dLP1, and different pools of LP-mAb integrants carrying mAb payloads in single (LP2 or LP15) or double (dLP1) loci were selected and sorted clonally. For example, dLP2-hEF1a-mAb with 1 and 2 mAb cassettes were constructed by selecting and sorting clonal integrants of 1x-hEF1a-mAb payload in single (LP2 or LP15) or double (LP2/LP15) loci, respectively. Clonal mAb integrants were propagated in the absence of selection, and mAb levels were measured weekly in conditioned media. Each bar represents 1 week of cell propagation and mAb measurements (mean ± SD, *n* = 2 or 3). (**C**) Antibody titers of clonal tLP1-mAb integrants stably expressing multi-copy mAb payloads integrated into single (LP2), double (biLP20) or triple (LP2/biLP20) loci in the tLP1 clone. mAb constructs were integrated into tLP1, and different pools of LP-mAb integrants carrying the mAb payload in single (LP2), double (biLP20) or triple (biLP20/LP2) loci were selected and sorted clonally. For example, tLP1-hEF1a-mAb with 1, 2 and 3 mAb cassettes were constructed by selecting and sorting clonal integrants of 1x-hEF1a-mAb payload in single (LP2), double (biLP20) or triple (biLP20/LP2) loci, respectively. Clonal integrants were propagated in the absence of selection, and mAb levels were measured weekly in conditioned media. Each bar represents 1 week of cell propagation and mAb measurements (mean ± SD, *n* = 2 or 3). The table under the graph indicates the LP loci integrated with the payload, the integrated mAb payload and the total number of the integrated mAb cassettes. mAb gene copy number of all clonal integrants was verified by ddPCR.

produce a production-ready cell line (24,44–46). In contrast, recombinase-mediated approaches allow for insertion of DNA payloads into preselected and validated integration sites, providing full control over the chromosomal context and the copy number of the integrated transgenes (14–16). Here we demonstrate the creation and utility of a multi-LP platform that allows for the integration of large DNA payloads at loci that support predictable and stable expression of transgenes. One shot integration of a large mAb payload (∼33 kb) into a triple LP led to the occupancy of all three LP sites with reasonable efficiency (∼1% before selection), and the entire process of development and validation of monoclonal triple mAb integration cell lines was achieved in

∼3 weeks. We observed no drop in mAb production during >1 month of cell propagation without selection, and no decrease in growth rate of single and multi-LP clones bearing up to ∼100 kb of integrated DNA payload (the triple-LP bearing the 3x-mAb construct). Moreover, combining the multi-LP platform with a multi-copy mAb payload resulted in a linear increase in mAb titer up to nine mAb copies. For higher numbers of LPs, the one-shot integration efficiency into all sites may drop below the practical selection level, and in this case sequential integration should be used instead to allow for integration into additional sites. It should be noted that the linear relationship may not necessarily hold under high expression conditions where other factors,
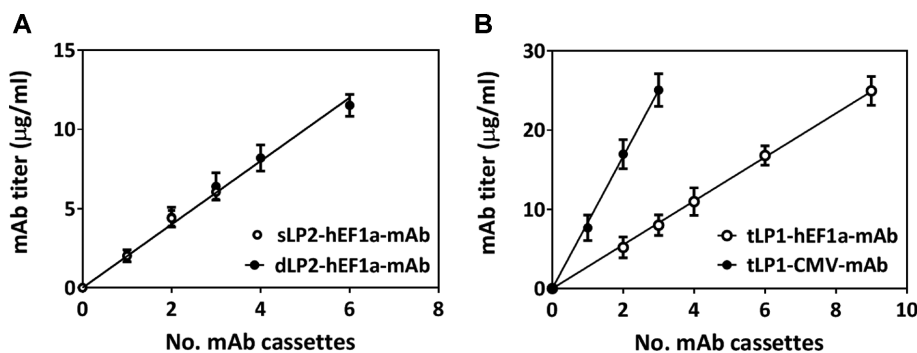
Figure 5. Linear increase in mAb titers with the use of multi-copy mAb circuits integrated in the multi-LP platform. Antibody titers of sLP2-mAb and dLP1-mAb (**A**) and tLP1-mAb (**B**) integrants of hEF1a-mAb and CMV-mAb circuits are shown against the total number of the integrated mAb cassettes. Total number of mAb cassettes is the number of integrated LP sites multiplied by the number of mAb copies in the mAb payload vector. mAb titers were calculated from the mAb stability assay (Figure 4) as the average titer values over 5 weeks of the stability assay (mean ± SD).

such as the secretory pathway's capacity, become the bottleneck for cellular productivity (47).

The mAb expression levels we observed are moderately low compared to industrial standards due to the use of a low-scale mAb production process in 24-well plate and adherent CHO-K1 cells that are not optimized for protein production. Our titers are not directly comparable to industrial titers obtained under heavily optimized biomanufacturing processes using optimized CHO strains and growth conditions, proprietary media and extended bioreactor processes (24). A recent study has in fact shown that an optimized biomanufacturing process can yield industrially relevant titers (∼1g/l) using a single LP for mAb expression (48). The application of our proof of concept multi-site integration platform to a production relevant CHO host may lead to further improvements of such systems.

Cell lines bearing multiple LPs that support stable long-term gene expression may also catalyze new advances in mammalian synthetic biology. Combined with inexpensive gene synthesis and modular cloning technologies, landing-pad cells could enable high-throughput construction and characterization of synthetic gene networks in mammalian cells, which may lead to new gene and engineered-cell therapies for human diseases. A multi-LP approach is particularly interesting in this context because it allows for rapid construction of combinatorial variants: one DNA payload could be kept constant while the others vary, or many possible variants could be combined in a single integration step to generate a diverse library of clones. Finally, multiple integration sites could help maintain independent expression from different parts of a synthetic gene circuit: situations where multiple transcriptional units are required but known to interfere with one-another could be separated into different genomic loci and thus isolated from each other.

In conclusion, our multi-LP platform enables reliable and predictable insertion of large payloads at well-characterized sites in the CHO-K1 genome. Future optimizations of the multi-LP platform may include using orthogonal recombination sites (e.g. BxB1 mutants) for recombinase mediated cassette exchange and one-shot integrations of libraries of DNA payloads (49). Together, these technologies promise improvements in cell line engineering for applications including expression of synthetic genes and gene circuits, gene

overexpression for functional genetics studies, ectopic expression of genes for reprogramming and lineage conversion, insertion of reporters for lineage tracing studies and construction of high yield therapeutic protein expressers for biomanufacturing.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Kinch,M.S. (2015) An overview of FDA-approved biologics medicines. *Drug Discov. Today*, **20**, 393–398.
2. Wurm,F.M. (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.*, **22**, 1393–1398.

3. Slusarczyk,A.L., Lin,A. and Weiss,R. (2012) Foundations for the design and implementation of synthetic genetic circuits. *Nat. Rev. Genet.*, **13**, 406–420.

4. Bacchus,W., Aubel,D. and Fussenegger,M. (2013) Biomedically relevant circuit-design strategies in mammalian synthetic biology. *Mol. Syst. Biol.*, **9**, 691–702.

5. Schott,J.W., Hoffmann,D. and Schambach,A. (2015) Retrovirus-based vectors for transient and permanent cell modification. *Curr. Opin. Pharmacol.*, **24**, 135–146.

6. Cattoglio,C., Facchini,G., Sartori,D., Antonelli,A., Miccio,A., Cassani,B., Schmidt,M., von Kalle,C., Howe,S., Thrasher,A.J. *et al.* (2007) Hot spots of retroviral integration in human CD34+ hematopoietic cells. *Blood*, **110**, 1770–1778.

7. Wu,X., Li,Y., Crise,B. and Burgees,S.M. (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science*, **300**, 1749–1751.

8. Schröder,A.R.W., Shinn,P., Chen,H., Berry,C., Ecker,J.R. and Bushman,F. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*, **110**, 521–529.

9. Desfarges,S. and Ciuffi,A. (2010) Retroviral integration site selection. *Viruses*, **2**, 111–130.

10. Smithies,O., Gregg,R.G., Boggs,S.S., Koralewski,M.A. and Kucherlapati,R.S. (1985) Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. *Nature*, **317**, 230–234.

11. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. and Church,G.M. (2013) RNA-Guided human genome engineering via Cas9. *Science*, **339**, 823–826.

12. Gaj,T., Gersbach,C.A. and Barbas,C.F. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.*, **31**, 397–405.

13. Cristea,S., Freyvert,Y., Santiago,Y., Holmes,M.C., Urnov,F.D., Gregory,P.D. and Cost,G.J. (2013) In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnol. Bioeng.*, **110**, 871–880.

14. O'Gorman,S., Fox,D. and Wahl,G. (1991) Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science*, **251**, 1351–1355.

15. Duportet,X., Wroblewska,L., Guye,P., Li,Y., Eyquem,J., Rieders,J., Rimchala,T., Batt,G. and Weiss,R. (2014) A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res.*, **42**, 13440–13451.

16. Turan,S., Zehe,C., Kuehle,J., Qiao,J. and Bode,J. (2013) Recombinase-mediated cassette exchange (RMCE)—a rapidly-expanding toolbox for targeted genomic modifications. *Gene*, **515**, 1–27.

17. Turan,S., Kuehle,J., Schambach,A., Baum,C. and Bode,J. (2010) Multiplexing RMCE: versatile extensions of the Flp-recombinase-mediated cassette-exchange technology. *J. Mol. Biol.*, **402**, 52–69.

18. Baser,B., Spehr,J., Büssow,K. and van den Heuvel,J. (2016) A method for specifically targeting two independent genomic integration sites for co-expression of genes in CHO cells. *Methods*, **95**, 3–12.

19. Zhang,L., Inniss,M.C., Han,S., Moffat,M., Jones,H., Zhang,B., Cox,W.L., Rance,J.R. and Young,R.J. (2015) Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line. *Biotechnol. Prog.*, **31**, 1645–1656.

20. Crawford,Y., Zhou,M., Hu,Z., Joly,J., Snedecor,B., Shen,A. and Gao,A. (2013) Fast identification of reliable hosts for targeted cell line development from a limited-genome screening using combined φC31 integrase and CRE-Lox technologies. *Biotechnol. Prog.*, **29**, 1307–1315.

21. Zhou,H., Liu,Z., Sun,Z., Huang,Y. and Yu,W. (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**, 122–129.

22. 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**, 100–111.

23. Kito,M., Itami,S., Fukano,Y., Yamana,K. and Shibui,T. (2003) Construction of engineered cho strains for high-level production of recombinant proteins. *Appl. Microbiol. Biotechnol.*, **60**, 442–448.

24. Wurm,F.M. (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat. Biotechnol.*, **22**, 1393–1398.

25. Lois,C., Hong,E.J., Pease,S., Brown,E.J. and Baltimore,D. (2002) Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science*, **295**, 868–872.

26. Engler,C., Gruetzner,R., Kandzia,R. and Marillonnet,S. (2009) Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS One*, **4**, e5553.

27. Guye,P., Li,Y., Wroblewska,L., Duportet,X. and Weiss,R. (2013) Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Res.*, **41**, e156.

28. Altschul,S.F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

29. Zambrowicz,B.P., Imamoto,A., Fiering,S., Herzenberg,L.A., Kerr,W.G. and Soriano,P. (1997) Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. *Proc. Natl. Acad. Sci. U.S.A.*, **94**, 3789–3794.

30. Irion,S., Luche,H., Gadue,P., Fehling,H.J., Kennedy,M. and Keller,G. (2007) Identification and targeting of the ROSA26 locus in human embryonic stem cells. *Nat. Biotechnol.*, **25**, 1477–1482.

31. Kanehisa,M. and Goto,S. (2000) Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **28**, 27–30.

32. Wang,J., Duncan,D., Shi,Z. and Zhang,B. (2013) WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.*, **41**, W77–W83.

33. 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**, 434–443.

34. Eisenberg,E. and Levanon,E.Y. (2013) Human housekeeping genes, revisited. *Trends Genet.*, **29**, 569–574.

35. Lewis,N.E., Liu,X., Li,Y., Nagarajan,H., Yerganian,G., O'Brien,E., Bordbar,A., Roth,A.M., Rosenbloom,J., Bian,C. *et al.* (2013) Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. *Nat. Biotechnol.*, **31**, 759–765.

36. Xu,X., Nagarajan,H., Lewis,N.E., Pan,S., Cai,Z., Chen,W., Xie,M., Wang,W., Hammond,S., Mikael,R. *et al.* (2012) The genomic sequence of the Chinese hamster ovary (CHO) K1 cell line. *Nat. Biotechnol.*, **29**, 735–741.

37. Haeussler,M., Schönig,K., Eckert,H., Eschstruth,A., Mianné,J., Renaud,J.-B., Schneider-Maunoury,S., Shkumatava,A., Teboul,L., Kent,J. *et al.* (2016) Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.*, **17**, 148–159.

38. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.

39. Sallustio,F., De Benedictis,L., Castellano,G., Zaza,G., Loverre,A., Costantino,V., Grandaliano,G. and Schena,F.P. (2010) TLR2 plays a role in the activation of human resident renal stem/progenitor cells. *FASEB J.*, **24**, 514–525.

40. Jäckel,C., Nogueira,M.S., Ehni,N., Kraus,C., Ranke,J., Dohmann,M., Noessner,E. and Nelson,P.J. (2016) A vector platform for the rapid and efficient engineering of stable complex transgenes. *Sci. Rep.*, **6**, 34365–34373.

41. Sadelain,M., Papapetrou,E.P. and Bushman,F.D. (2011) Safe harbours for the integration of new DNA in the human genome. *Nat. Rev. Cancer*, **12**, 51–58.

42. Moustakas,A. and Heldin,C.-H. (2016) Mechanisms of TGFβ-induced epithelial–mesenchymal transition. *J. Clin. Med.*, **5**, 63–96.

43. 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**, 1100–1109.

44. Pilbrough,W., Munro,T.P. and Gray,P. (2009) Intraclonal protein expression heterogeneity in recombinant CHO cells. *PLoS One*, **4**, e8432.

45. 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.

46. Du,Z., Mujacic,M., Le,K., Caspary,G., Nunn,H., Heath,C. and Reddy,P. (2013) Analysis of heterogeneity and instability of stable mAb-expressing CHO cells. *Biotechnol. Bioprocess Eng.*, **18**, 419–429.

47. Pieper,L.A., Strotbek,M., Wenger,T., Gamer,M., Olayioye,M.A. and Hausser,A. (2017) Secretory pathway optimization of CHO producer cells by co-engineering of the mitosRNA-1978 target genes CerS2 and Tbc1D20. *Metab. Eng.*, **40**, 69–79.

48. 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**, 1463–1467.

49. 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**, 1837–1846.