REVIEW

Novel approaches to *in vitro* transgenesis

A D Adamson, D Jackson and J R E Davis[1]

Faculty of Life Sciences, Manchester Interdisciplinary Biocentre, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK
[1]Developmental Biomedicine Research Group, University of Manchester, AV Hill Building, Manchester M13 9PT, UK
(Correspondence should be addressed to A D Adamson; Email: antony.adamson@manchester.ac.uk)

Abstract

The study of gene expression is a major focus in biological research and is recognised to be critical for our understanding of physiological and pathophysiological processes. Methods to study gene expression range from *in vitro* biochemical assays through cultured cells and tissue biopsies to whole organisms. In the early stages of project development, considerations about which model system to use should be addressed and may influence future experimental procedures. The aim of this review is to briefly describe advantages and disadvantages of the existing techniques available to study eukaryote gene expression *in vitro*, including the mechanism of transgene integration (transient or stable), the different transgenesis systems available, including plasmids, viruses and targeted integration and knockin approaches, and paying particular attention to expression systems such as bacterial artificial chromosomes and episomal vectors that offer a number of advantages and are increasing in popularity. We also discuss novel approaches that combine some of the above techniques, generating increasingly complex but physiologically accurate expression systems.

Journal of Endocrinology (2011) **208,** 193–206

Gene expression

Gene expression and function is regulated at a number of levels: 1) from an initial stimulus or signal (which can be either intracellular or extracellular, leading to activation of a signalling pathway); 2) alterations in the chromatin architecture of the DNA and the promoter sequences; 3) transcription and mRNA production (splicing, transport, stability and degradation); and 4) translation and the formation of the protein product (which may be subjected to post-translational modification). This review will focus on the ectopic expression systems that are currently available to utilise transgenesis and to study the regulation of gene transcription.

Endogenous versus ectopic gene expression

Specific RNA production by cells has long been held as a direct indicator of gene expression. Techniques such as northern blotting (Alwine *et al.* 1977) and RNAse protection assays (Sambrook *et al.* 1989) can be used to assay RNA levels within a population of cells, and nuclear run-on assays (Gariglio *et al.* 1981, Brown *et al.* 1984) have been used to measure transcription initiation. Later, reverse transcription (transcriptase)-PCR (RT-PCR), and more quantitative variants (real-time PCR and quantitative PCR) have been developed to analyse RNA expression. Each of these techniques relies on endogenous gene expression within the cell, and has their own advantages and disadvantages. However, commonly no suitable cell line exists that expresses the gene-of-interest. Furthermore, these assays measure gene expression by extracting RNA from a population of cells, thus averaging gene expression across an entire population. Single-cell analysis using sophisticated microscopy has revealed startling heterogeneity of gene transcription between cells even within clonal populations (Takasuka *et al.* 1998, Ashall *et al.* 2009, Eldar & Elowitz 2010). Analysis of gene expression at the single-cell level using microscopy requires the addition of an appropriate marker protein or dye to gene expression vectors in order to visualise these effects. This is most commonly assessed using reporter systems that can be ectopically expressed in the cell.

At the beginning of any project, critical decisions should be made to define the methodological approaches to be pursued. Considerations should include

- The cell line(s) to be used.
- The reporter gene(s) to be used (colourimetric enzymes, luminescent enzymes, fluorescent proteins, etc.).
- Whether an increase in the copy number of a gene may lead to aberrant effects (e.g. gene overexpression, perturbation of feedback mechanisms, etc.).

- What kind of delivery system would be most appropriate (plasmid, viral, episomal, stably integrated or transient expression).

These above considerations are not mutually exclusive and can often overlap. In this review, we shall identify the individual problems and advantages inherent to a range of experimental approaches used to analyse transgenesis and ectopic gene expression.

Choosing a suitable cell model system

The choice of cell line(s) should be determined early in any project, initially by analysing expression of the gene-of-interest in a number of relevant cell lines. For many genes, particularly endocrine genes that tend to be tissue- and cell-specific in their expression pattern, it is common to fail to identify a cell type suitable for gene expression analyses. For example, many endocrine genes are specifically restricted in expression to the pituitary, but their study is hampered by the absence of human pituitary-derived cell lines, and primary tissues rarely become available due to a lack of pituitary surgical intervention.

In the absence of human-derived cell lines for a specific investigation several alternative approaches to study gene regulation have been used. These include transient expression of specific transcription factors in HeLa cells (Bodner & Karin 1987), use of human cell lines isolated from different tissue origin that unusually expresses gene-of-interest (Gellersen *et al.* 1995) or the use of cells derived from the appropriate tissue, but from an alternative species, such as rodents (Ben-Jonathan *et al.* 2008). However, these approaches are not without potential criticism. It is difficult to account for cell-specific effects either lost or gained by using cells of different origins. Regulation of a specific gene may differ from species to species. For example, prolactin, which is expressed chiefly in pituitary lactotroph cells, has extensive differences in gene organisation and hormone function between humans and rodents (Gerlo *et al.* 2006, Ben-Jonathan *et al.* 2008, Bernichtein *et al.* 2010). In these circumstances, ectopic expression of human gene–promoter constructs in rat pituitary cells is a common approach for assessing the regulation of the human *prolactin* gene (e.g. Adamson *et al.* 2008, Semprini *et al.* 2009).

Additionally, cell lines have inconsistencies and limitations when compared with primary tissue. For example, many rodent pituitary cell lines lack functional dopamine receptors, and as a consequence are unable to respond to one of the most important *in vivo* pituitary regulatory signals. Although a lengthy and time-consuming process, it is possible to remedy this issue by generating stable cell lines expressing physiological levels of the two dopamine receptor variants resulting in a more physiologically relevant model system (An *et al.* 2003).

Often, compromises need to be made when selecting an appropriate cell line. Cell lines differ according to a number of characteristics including their growth and proliferation rates,

whether they are adherent, easily transfected or transduced and whether aberrant gene expression or chromosomal aneuploidy occurs. Single-cell time-lapse microscopy experiments are far easier to analyse if cells remain relatively static. Thus, on occasions the cell line selected may not necessarily be the most physiologically relevant and the compromises that are made need to be acknowledged.

Reporter genes

Studies of mammalian gene expression have relied heavily on the use of reporter genes that encode a measurable exogenous protein product that can give a quantitative index of the level of gene activation gene when linked to a given promoter. The library of reporter genes available has significantly expanded over the years, and includes enzymes such as β-galactosidase, which generate a colourimetric product or chloramphenicol acetyl transferase, which catalyses the synthesis of radio-labelled (and latterly fluorescent) product, to enzymes such as luciferase that catalyses the production of light from the substrate luciferin to give a measure of gene activation (Frawley *et al.* 1994, White *et al.* 1995). Luciferase can be used to analyse the induction of gene expression across a population of cells by assaying the whole cell lysate, and also to dynamically analyse gene expression in a single cell through microscopy, measuring the photon output from a single cell (Harper *et al.* 2010).

Also available are fluorescent proteins such as green fluorescent protein, originally derived from bioluminescent marine species (Tsien 1998, Shaner *et al.* 2007). Fluorescent proteins have revolutionised not only gene expression studies but, through fusion to the coding sequence of the gene-of-interest, have enabled visualisation of the behaviour of the protein within the cell over time (Rafalska-Metcalf & Janicki 2007, Spiller *et al.* 2010).

Currently, a multitude of vectors that express differently coloured fluorescent proteins are available; most are altered version of the natural DNA sequence (e.g. cherry, tomato; Chudakov *et al.* 2010). Fluorescent proteins display different properties, including variable brightness, ability to form monomers or multimers, photo-switching, photo-activation, photo-conversion, and pH and temperature sensitivity (Giepmans *et al.* 2006, Chudakov *et al.* 2010). Further modifications such as addition of PEST or CL-1 sequences (Rogers *et al.* 1986) shorten intracellular half-lives of fluorescent proteins to allow gene expression to be monitored more accurately. Increasing our understanding of dynamic and stochastic gene expression requires the facility to monitor rapid changes in the transcriptional rate of the gene-of-interest (Raser & O'Shea 2004). Accurate quantitative data for reporter gene synthesis and degradation is essential in order to incorporate mathematical modelling and systems biology approaches into studies (Finkenstädt *et al.* 2008). Thus, reporter gene development is an ongoing process and is important to ensure that reporter molecules can be selected and tailored to meet the requirements of specific gene expression studies.

Driving gene expression

Many studies designed to investigate protein behaviour within a cell utilise expression vectors with gene transcription driven by constitutively active promoters with a broad range of activity in most cells, such as the viral cytomegalovirus, Rous sarcoma virus and Simian virus 40 (SV40) promoters. The constitutive transcription from these promoters is by way of a concatenation of numerous different response elements and is suitable wherever the constant production of protein within a cell is necessary. However, these promoters do not express the protein appropriately with regard to the onset of transcription, induction by hormones or drugs and regulation by other factors that may constitute feedback loops. Furthermore, when some genes are expressed at incorrect levels it can lead to aberrant activation or repression of other genes. The use of the relevant promoter of the gene–of–interest greatly enhances the suitability of data generated. Promoter–mediated regulation is highly specific to each individual gene and can often involve multiple transcription factor binding sites found in a multitude of different locations relative to transcription start site. Transcription factor binding sites tend to be primarily within the proximal region of the gene for most $5'$ sequences, but enhancers and locus control regions can be many kilobases away, in both upstream and downstream and even intronic sequences (Carroll *et al.* 2005, West & Fraser 2005). An in–depth knowledge of promoter structure can greatly influence the choice of expression system; this will be discussed in further detail later in this review.

Gene delivery: mechanism of transgenesis

Before analysis of protein or gene expression can be analysed, the molecular ectopic expression systems must first be introduced to the cell. This can be achieved through a variety of biological, chemical or physical techniques including, for example, viral transduction, cationic lipid delivery or electroporation (Kim & Eberwine 2010) but is ultimately determined by the cellular model system(s) chosen. Many different expression systems are available, but all can be loosely categorised as either purified circular bacterial DNA structures (i.e. plasmids, episomal vectors and larger structures such as cosmids and bacterial artificial chromosomes (BACs); see later) or viral vectors.

Plasmids, small circular bacterial DNA structures, are perhaps the most commonly used ectopic expression system. They are easy to purify from bacterial cultures by alkaline lysis, and can be readily modified through well–characterised molecular cloning techniques. Plasmids can be introduced into cells using chemical transfection (including a wide variety of specialist compounds that differ markedly in efficiency, cell–type specificity and cost) and/or physical transfection through direct manipulation of the cells (e.g. electroporation and microinjection among other techniques). BACs are similar in many ways to other bacterial vectors in their propagation and purification and have several advantages when compared with plasmids, but modification and transfection are more challenging than using plasmids (see later).

Viral transduction approaches exploit the natural ability of viruses to deliver genetic material into infected cells. A number of viral-based vectors are commonly used to ectopically express reporter constructs, including adenovirus (AV), adeno-associated virus (AAV), herpes simplex virus (HSV) and retroviruses (principally lentiviruses (LV)). Each viral vector has a number of advantages and disadvantages and this has been reviewed extensively elsewhere (Verma & Weitzman 2005, Osten *et al.* 2007, Howarth *et al.* 2010). In this review, we present a brief summary of the general properties of viral transduction and summarise the characteristics and properties of commonly used viral vectors (Table 1).

Standard molecular cloning techniques are first used to generate a gene-of-interest-containing vector *in vitro* in suitable viral construct(s) containing genetic information critical for the formation of the virus. Next, after the infection of a host cell, facilitating replication of the viral genome within the host cell environment, and formation and propagation of new virions, the cells are lysed or the virus-containing supernatant is removed, centrifuged and added to the target cell line to infect and transduce the target cells. The viral titre or yield obtained from the host cell differs according to the virus used and some viruses, such as AAV, often require helper virus expression to increase the titre. Thus, preparation for performing transgenesis is lengthier and more labour intensive compared with using a simple bacterial vector construct.

Furthermore, viruses vary in their ability to infect different cell types, but usually target both dividing and non-dividing cells (Pfeifer 2004). Viruses often have limited cloning capacity (4–7 kb), with the exception of HSV, which can hold very large ($>$100 kb) DNA inserts (Cuchet *et al.* 2007). One key advantage of using viruses (with the exception of AVs; see Table 1) is the relative ease of generating cells with an integrated transgene without the need for selection markers and clonal selection strategies.

Gene delivery for gene transcription analysis: transient or stable expression

Transgenes can be either transiently expressed for a period of hours to days or stably integrated into the target cell genome through appropriate transformation techniques and/or the use of selection markers. Stable integration of transgenes can be a time consuming and laborious process, requiring weeks to months to generate suitable clonal cell populations depending on the transgenesis method used, whereas transient expression can yield data in a relatively short time frame (for a comparison of the different transgenesis methods and their transient and stable effects see Fig. 1).

Many authors have reported discrepancies between these expression systems. For example, the MMTV promoter, a sequence extensively used in transcriptional studies, varies

Table 1 Properties of viral vectors

Virus	Nucleic acid	Maximum insert or packaging size	Typical titre	Target cell line choice	Expression duration	Interaction with cell genome
Lentivirus (retrovirus)	ssRNA	Low ($\sim$7 kb)	High (10^7–10^9 cfu/ml)	Broad range	Lengthy	Integrates (some non-integrating; Yanez-Munoz et al. 2006)
Adenovirus	dsDNA	Low ($\sim$7 kb)	Very high (10^{10}–10^{12} cfu/ml)	Broad range (although some serotype specificity)	Transient	Does not integrate
Adeno-associated virus	ssDNA	Very low ($\sim$4·5 kb)	Medium (10^7 cfu/ml), and requires helper viruses	Broad range	Potentially lengthy	Does not integrate when containing ITRs (WT target integration into chr 19; McCarty et al. 2004)
Herpes simplex virus	dsDNA	Very high ($\sim$200 kb)	High (10^8–10^9 cfu/ml)	Limited range	Potentially lengthy	Does not integrate

greatly in activation by different transcription factors depending on transient or stable transfection (Archer *et al.* 1992, 1994). This was shown to be due to the incorrect nucleosome positioning of the DNA transiently transfected into cells (Hebbar & Archer 2008). It has been demonstrated that transiently transfected DNA often fails to efficiently form the appropriate chromatin structure and, in some cases, regulation of these transgenes may not accurately represent the *in vivo* behaviour of the promoter-of-interest. Some strategies exist to circumvent this issue by use of cell lines expressing SV40 T antigen and integrating the SV40 origin of replication into the vector. Using this approach, plasmid DNA is assembled into nucleosomes that mimic the endogenous gene, and the resulting mini-chromosomes are more accurately replicated and transcribed by the cellular machinery (Xu & Cook 2008). However, there are likely to be multiple copies of the transgene, which may not faithfully reproduce expression profiles.

Additionally, transient transfection efficiency also can differ extensively between experiments. A common practice is to co-transfect control reporter plasmids, to help determine the level of transfection between samples, although some authors report that regulation of these vectors themselves may lead to systematic error in the experimental data (Huszar *et al.* 2001). Furthermore, Ishikawa *et al.* (2004) have reported that co-transfected plasmids have the potential to form concatamers, through non-homologous end-joining ligation, leading to further problems in interpretation of the results.

Despite the problems associated with transient transfection, the technique remains a quick and simple method for analysis of gene transcription, particularly when screening a number of promoter–reporter constructs (such as in promoter deletion and mutagenesis assays). While stable cell lines are more likely to reflect the true nature of molecular interactions in gene regulation, especially with regard to chromatin and transcription dynamics, the generation of stable cell lines is time consuming and laborious and clones need to be well characterised. Furthermore, clones may suffer from 'site of integration effects' and copy number abnormalities (see below), and it may be advisable to experiment with a number of different clones, or a mixed population of clones, to ensure the activation of the target promoter is consistent throughout, but overall stable clones do provide a more consistent and well-characterised cell model system.

Increasing gene copy number

The introduction of reporter transgene(s) usually results (with the exception of gene knockins; see later) in altered gene copy number within a cell, the effects of which are understudied and, therefore, poorly understood. The most thorough analysis of how increasing the gene copy number affects expression *in vitro* has come from studies in yeast (Guan *et al.* 2007, Presser *et al.* 2008) and *Drosophila* (Sabl & Henikoff 1996). With regard to mammalian gene expression, transcriptomic data suggest that the number of transcripts from a gene roughly increases
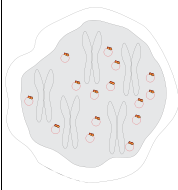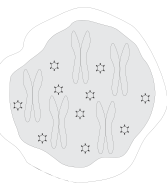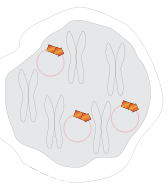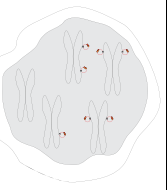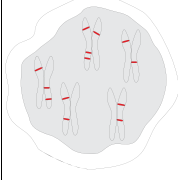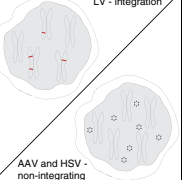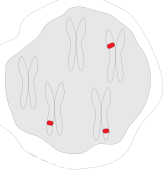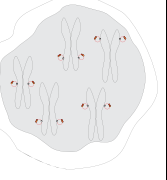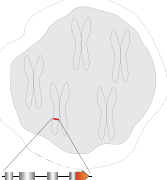
Figure 1 Transient versus stable transfection. Multiple methods for achieving transgenesis are shown. Column 1 – plasmids. Multiple copies found transiently, and multiple integration sites upon stable transfection. Column 2 – lentivirus (LV), adeno-associated virus (AAV) and herpes simplex virus (HSV). Transduction is long term and not of a transient nature. LV integrates with several copies; AAV and HSV are both non-integrative. Column 3 – adenovirus. Virus usually lacks replication ability, thus only transient transduction, at multiple copies, is achieved. Column 4 – bacterial artificial chromosomes (BACs). Transient expression results in low copy number within the nucleus, stable expression is integrative at a low copy number also. Column 5 – episomal vectors. Transient expression of several copies, vector associates with chromosome and nuclear matrix. For stable expression, the vector continues to associate with the chromosomal regions and nuclear matrix but does not integrate. Column 6 – gene knockin or targeted integration. These are stable only approaches and involve targeting either the endogenous gene locus or a predetermined locus.

in proportion to gene copy number (Zhang & Oliver 2007). It is therefore possible that ectopic transgenesis approaches that result in increased gene copy number within the cell may perturb the biological system. It is relevant to mention that some studies adopting a mathematical approach have suggested that increased gene dose can be compensated for through proportional effects on feedback loops (Mileyko & Weitz 2010). Based on the above observations it is likely that the effect of copy number is gene and system specific. Thus, the effect of increasing transgene copy number remains an unknown quantity in most systems and needs to be acknowledged as different ectopic expression systems vary in the number of transgene copies they introduce to the cell. It should also be noted that due to epigenetic silencing of non-mammalian DNA (see 'Integration effects' and Fig. 2B) that the effective copy number of a clone may reduce over time with increasing generation number (Jenke *et al.* 2004, Mutskov & Felsenfeld 2004; Table 2).

Integration effects

The site(s) within the genome at which an integrated reporter gene resides can have a profound influence on gene expression, either of the transgene itself or surrounding genomic regions. Integration of a transgene into dense, inactive heterochromatin regions can result in unstable position effect variation, resulting in little or no transgene expression (Fig. 2A; Dillon & Festenstein 2002). It is also possible that gradual silencing of the transgene can occur over time as heterochromatin can spread into the majority of integrated genes, especially when the

exogenous DNA contains bacterial DNA sequences (Fig. 2B; Jenke *et al.* 2004, Mutskov & Felsenfeld 2004). Integration of exogenous DNA may also disrupt important endogenous sequences including genes and regulatory elements, leading to unanticipated changes in the cell phenotype (insertional mutagenesis; Fig. 2C).

Integration within the 'range' of endogenous enhancers/promoters may lead to overexpression of the transgene (Fig. 2D). Given that enhancers and locus control regions can influence genes found several kilobases away this 'range' is potentially quite large. The site of integration can be critical for correct control of gene expression and there are a number of different methods to limit these effects, including the addition of insulator elements to flank the transgene and its promoter (Fig. 3A, upper panel; Bell *et al.* 2001). Insulator elements are natural genetic sequences that establish independent domains of transcriptional activity within the eukaryotic genome and have distinct roles in establishing gene expression; acting as barrier elements to prevent the spread of heterochromatin and blocking enhancer–promoter interactions (Fig. 3A, lower panel; Kuhn & Geyer 2003, Recillas-Targa *et al.* 2004). Thus, while insulators would not protect against insertional mutagenesis that may lead to aberrant gene expression of non-transgenes (Fig. 2C), flanking transgene with insulators confers position independence, allowing self-regulation, appropriate chromatin organisation and prevents transgene silencing by the spread of heterochromatin. Insulators have proved highly effective both *in vitro* (Anastassiadis *et al.* 2002, Qu *et al.* 2004) and also when used *in vivo* in the development of gene therapy approaches (Gallagher *et al.* 2009).
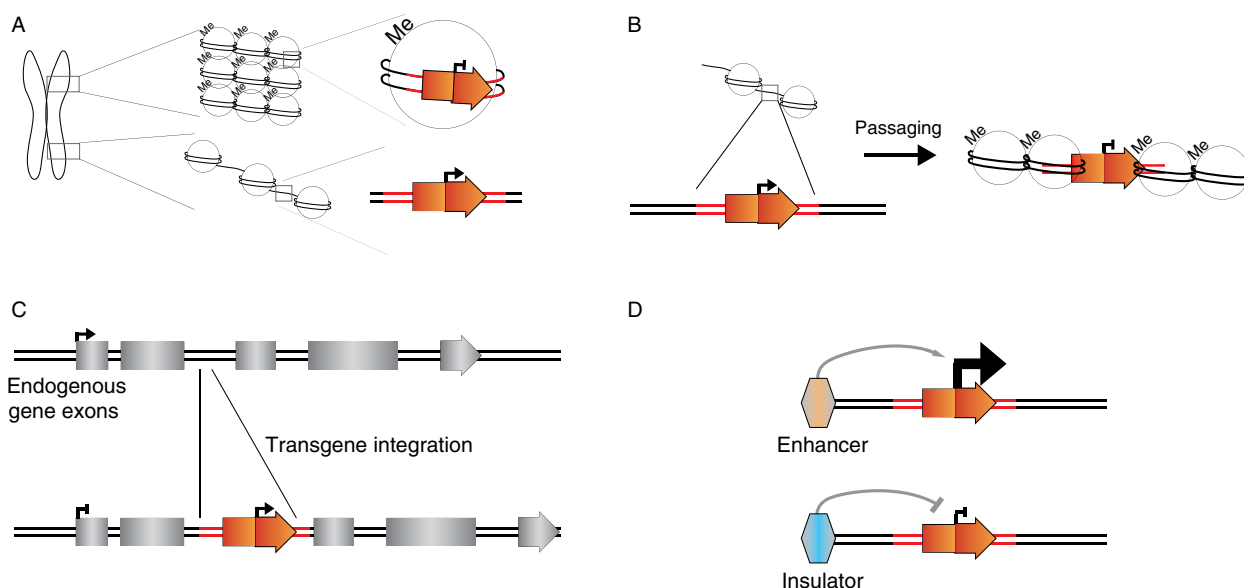
Figure 2 Site of integration effects. Black lines/boxes/arrows represent endogenous DNA, red lines/boxes/arrows represent transgene DNA. (A) Integration of transgene into different chromatin states. Upper panel shows inhibited transcription of transgene upon integration into dense heterochromatin state ('Me' indicates hypermethylation), lower panel shows transgene able to be transcribed when integrated into open, euchromatin state. (B) Insertional mutagenesis; disruption of an endogenous gene. Integration of transgene disrupts a random endogenous gene resulting in loss or altered expression. (C) Integration within the range of endogenous transcriptional control elements, left panel shows overexpression of transgene due to the effect of endogenous enhancer, right panel shows inhibited transgene expression due to the effect of endogenous insulator. (D) Transgene silencing over multiple generations. Initial integration results in transgene expression but passaging results in the spread of heterochromatin, particularly if the exogenous DNA contains non-mammalian DNA.

Targeted integration

Integration of transgenes into pre-determined endogenous genomic loci, and subsequent selection by an appropriate marker, known as recombinase-mediated cassette exchange (RMCE; Schlake & Bode 1994), is a method for generating stable cell lines and limiting site of integration effects (Fig. 3B). In this approach, a pool of stable cell lines is created that have recombinase target sites (including Flp recombination target (FRT), locus of X-over P1 (LoxP), and more recently phiC31 integrase sites; Oumard *et al.* 2006), integrated into the genome along with a selection marker. Screening of these founder clones can then identify the genomic position of these recombinase sites, and suitable clones can be selected and expanded. Criteria for clones include integration of a single copy, in a genomic locus of high transcriptional potential and a lack of insertional mutagenesis (Qiao *et al.* 2009). These clones can then be subjected to transfection with an expression vector carrying the gene-of-interest, and a complementary recombinase site. Transient expression of the relevant recombinase facilitates integration of the gene-of-interest into the pre-determined genomic loci (Thomas *et al.* 2004). This process can be engineered so as not to leave a selection marker or extensive prokaryotic DNA, reducing the likelihood of epigenetic silencing (Oumard *et al.* 2006).

A drawback to this approach is the random integration of the initial recombinase target site and clone screening, which may be time consuming and subject to chance. Once a suitable line is identified, however, it allows great flexibility and rapid and efficient generation of copy number-controlled stable cell lines in the future, especially useful if a number of constructs are to be screened (e.g. mutagenesis and promoter length studies). There are many easy-to-use commercially available kits for this process, as well as pre-generated recombinase target site cell lines, negating the first, lengthy step.

RMCE has numerous forms, each with their own advantages and disadvantages and varying levels of complexity (e.g. FRT site variants allowing for RMCE multiplexing; Bode *et al.* 1992, Turan *et al.* 2010) and has been expertly reviewed elsewhere (Oumard *et al.* 2006, Wirth *et al.* 2007). These methods provide high flexibility in experimental approach, but may require extensive molecular engineering contributing to generation time of the cellular model system, and concessions on the background cell line may have to be made.

Gene 'knockin'

A gene 'knockin' can be specifically targeted to the endogenous gene locus to homologously recombine in

Table 2 Comparison of various ectopic expression systems

Ectopic expression system	Construct generation time	Construct generation difficulty	Transfection efficiency	Promoter for gene expression	Transcript	Copy number	Stable cell generation	Genomic insertion site	Integration effects	Target cell line choice
Plasmids	Short (weeks)	Low	High	Short (max ~5 kb)	Direct from cDNA[a]	High	Weeks–months	Not predictable	Highest likelihood	Practically unlimited
Viral approach[b]	Short–medium (weeks–months)	Medium	Usually high	Short (max ~5 kb), except HSV (>30 kb)	Direct from cDNA[a] except HSV	Variable	Weeks–months	See Table 1	See Table 1	See Table 1
BACs	Medium–long (weeks–months)	Medium–high	Low	Large genomic region (100–300 kb)	Full	Low	Months[c]	Not predictable	Low risk	Practically unlimited[d]
Episomal constructs	Short (weeks)	Low–medium	High	Short (max ~5 kb)[e]	Direct from cDNA[a,e]	High[e]	Weeks–months but high efficiency	No insertion in genome	None	Practically unlimited
Targeted integration	Variable[f]	Variable[f]	High	Short (max ~5 kb)[g]	Direct from cDNA[a,g]	Controlled and low	Rapid after suitable founder clone identified	Pre-determined in founder clone[e]	Minimised	Limited
Gene knockin	Medium (>1 construct may be required)	High (complex targeting systems required)	N/A, but homologous recombination highly dependent on cell line	Full endogenous regulatory sequences	Full	Equivalent to endogenous	Single allele: months; both alleles: month–years	Pre-determined	None	Extremely limited

[a]Except for naturally short genes, small enough for the entire sequence with introns, exons and UTRs to fit into smaller constructs.
[b]General properties of viruses described, more detail found in Table 1.
[c]Further modifications of the construct, such as retrofitting selection markers, may be necessary, increasing time required.
[d]BAC transfection is less efficient than transfection of smaller constructs.
[e]See 'Combining approaches'.
[f]Can be very simple (e.g. using commercial system) or very complex depending on project.
[g]To our knowledge targeted insertion of large constructs has not been attempted.
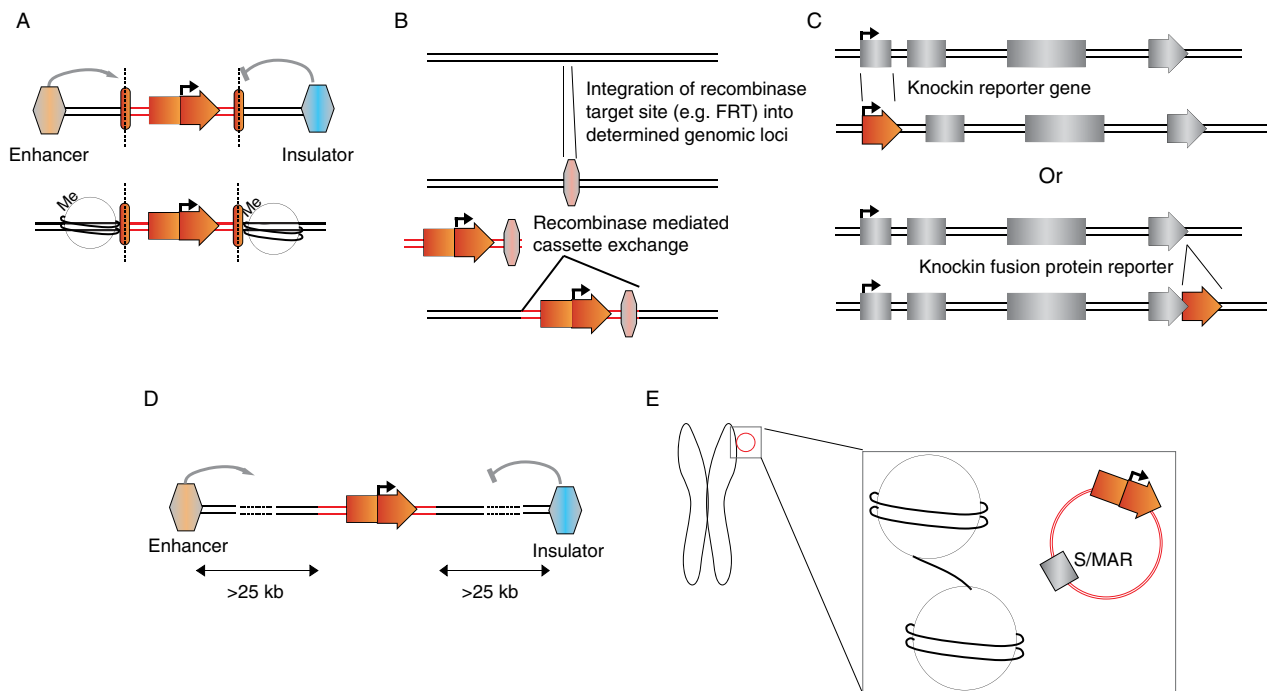
Figure 3 Methods to minimise site of integration effects. Black lines/boxes/arrows represent endogenous DNA, red lines/boxes/arrows represent transgene DNA. (A) Flanking transgene DNA with insulator sequences. If the transgene integrates within the range of endogenous transcriptional regulatory elements, insulators can block their effects (upper panel). Insulators can also halt the spread of heterochromatic and gene silencing (lower panel). (B) Targeted integration. Following determination of recombinase target integration site, the gene-of-interest can be specifically targeted to a favourable locus. (C) Knockin. Owing to the targeting of the endogenous gene locus no site of the integration effects will be observed, either when a reporter is knocked in to replace the endogenous gene (upper panel), or when a fusion protein knockin is created (lower panel). (D) Use of large constructs. The extended flanking sequences found in BACs leave the transgene itself out of range of enhancers and insulators. (E) Use of episomal vectors. Episomal vectors associate with the nuclear matrix close to the chromatin through the scaffold/matrix attachment region (S/MAR) but do not integrate.

the reporter construct of choice, either as direct readout of promoter activity (by replacing the gene itself with a reporter gene; e.g. Ishikawa *et al.* 2006) or combined with the endogenous gene as a fusion reporter (e.g. De Lorenzi *et al.* 2009). The advantage of this approach is the targeting of endogenous genes, resulting in no variation in gene copy number or site of integration effects, as well as the expression of the full gene (introns, exons, untranslated region), which contribute to gene regulation. This approach is commonly used in embryonic stem (ES) cells for the generation of knockin/reporter animals and in *Saccharomyces cerevisiae* in which the high homologous recombination rate in yeast has been exploited to generate epitope tags on a number of yeast genes and proteins (Gavin *et al.* 2006). Gene knockin approaches have also been applied to somatic human cell lines for *in vitro* gene analysis (Hendrickson 2008, Kim *et al.* 2008). However, homologous recombination in mammalian cells is a highly inefficient process (Yáñez & Porter 1999) and, in particular, human somatic cell lines have been shown to display an intrinsically low rate of homologous recombination (Sedivy *et al.* 1999, Hendrickson 2008). Lengthy homology arms (at least 2–3 kb) are necessary to increase recombination efficiency resulting in complex molecular construction.

Developments increasing the efficacy of this approach in somatic cells have proved invaluable. It has been shown that the delivery method of the recombination cassette DNA is critical. Hirata & Russell (2000) used a recombinant AAV approach that greatly increased recombination efficiency when compared with transfected plasmids (Topaloglu *et al.* 2005, Rago *et al.* 2007). However, again the cell model system was essential, as there was a huge variation in the efficiency of transfection and recombination between the cell types (Rago *et al.* 2007).

Furthermore, in addition to modifying the endogenous locus, the exogenous DNA can be randomly integrated at multiple, non-targeted sites, leading to both site of integration effects and an increase in transgene copy number. Sophisticated molecular approaches that increase targeting efficiency have been developed, including promoter traps where the transgene recombination cassette contains not only the desired modification and a selection marker flanked with loxP sites (or other suitable recombinase target) but also a splice acceptor immediately upstream of an internal ribosome entry site antibiotic-resistance gene (Topaloglu *et al.* 2005, Rago *et al.* 2007). The splice acceptor initiates transcription of the selection marker driven by the endogenous promoter of

the gene, thus selection marker expression only occurs when correct DNA integration is achieved. The marker can then be removed through transient expression of the necessary recombinase (the proximity of a constitutive promoter-driven marker gene expression could conceivably have indirect effects on the regulation of the endogenous gene; Hasegawa & Nakatsuji 2002). However, it is important to note that the endogenous target gene must be basally expressed at a sufficient level in order to drive selection marker expression, potentially limiting the usefulness of this approach for some genes and cell types.

Creating knockin somatic cell lines is a lengthy procedure, requiring generation of complex targeting cassettes, followed by delivery of the cassette via a suitable method to a limited number of cell lines at low efficiency. Following this, stable clones need to be selected and carefully characterised. However, the lack of copy number alterations and the site of integration effects (Figs 1 and 3C) result in an excellent model system for *in vitro* molecular tagging gene expression analysis (Kim *et al.* 2008).

Large constructs

Another approach to minimise site of integration effects is to use large constructs such as yeast, bacterial or P1 artificial chromosomes (YACs, BACs and PACs respectively). YACs were originally developed for cloning large genomic fragments into yeast cells and BACs and PACs were developed for cloning large genomic fragments into *Escherichia coli*. YACs can contain huge, megabase-sized DNA inserts (compared to 100–350 kb inserts of BACs and PACs) but YAC DNA is notoriously difficult to purify. YACs exist in multiple copies in bacteria that cause chimaerism, and random recombination events occur frequently (Copeland *et al.* 2001). YAC manipulation by recombination is an even lengthier process than that with BACs/PACs; thus, YACs are not as popular a model system as their bacterial counterparts for transgene vector construction.

BACs can be considered large plasmids, propagated in bacteria and consisting of a small amount of bacterial DNA derived from the single copy F-plasmid (Shizuya *et al.* 1992) containing prokaryotic genes essential for replication, partition and selection, and a large genomic fragment of mammalian DNA (100–350 kb). These vectors were a useful tool in the genome-sequencing projects (Lander *et al.* 2001, Venter *et al.* 2001) and as a result most regions of the human genome (as well as genomes of several other species) are available as BACs. With the development of homologous recombination systems in *E. coli* (Copeland *et al.* 2001), BACs can be genetically engineered to express reporter genes, or to place a transgene under the control of inducible or conditional promoters making them an increasingly popular method of gene expression in the mammalian cells. The sheer size of BAC DNA vectors helps to minimise the site of integration effects (Fig. 3D).

In addition to minimising the integration effects, BACs have a number of other advantages. Often, proximal upstream promoter elements are used to drive gene expression and small constructs such as plasmids and viruses may have up to 5 kb of DNA immediately flanking the gene-of-interest to drive reporter gene expression. However, transcriptional regulatory mechanisms are highly gene specific and proximal-flanking DNA is often insufficient to drive transgene expression (e.g. human GH; Palmiter & Brinster 1986), or at least does not reflect the full promoter behaviour for gene activation (Semprini *et al.* 2009). In such cases, the use of larger constructs has shown the necessity of surrounding genomic sequences, containing enhancers and locus control regions, resulting in a more accurate representation of *in vivo* gene expression (e.g. Townes *et al.* 1985, Jones *et al.* 1995). Using a larger construct that contains a sizeable fragment of the human genome surrounding the gene-of-interest will therefore increase the probability that any such regulatory elements are included. Moreover, bioinformatic programmes are increasing in sophistication and regulatory regions such as enhancers and locus control regions can be predicted from the genomic sequence. This process can either be based on the identification of predicted or probable binding sites for specific transcription factors, or alternatively through the analysis of conserved non-coding DNA sequences (Sakabe & Nobrega 2010). Combining bioinformatic approaches with the selection of BAC clones allows the inclusion or exclusion of potential regulatory sequences.

In recent years, there has been increasing interest of the role of RNA in gene regulation, including RNA splicing and gene regulation by microRNAs (Jackson & Standart 2007). As BACs are essentially a fragment of the human genome they will contain the full gene structure, including untranslated regions, exons and introns, alternative promoters and splice sites and microRNA coding sequences. This should result in full mRNA processing and splicing when transcribed, and will also include any microRNA targets sites within the mRNA sequence and produce the full complement of protein isoforms once translated. Thus, genes expressed from BACs mirror endogenous gene expression far more accurately.

Despite their size, it is still possible to transiently express BACs in mammalian cell lines, though the efficiency of transfection tends to be lower than for smaller plasmids (Magin-Lachmann *et al.* 2004) but they are introduced into cells at a lower copy number (Sparwasser & Eberl 2007).

BAC recombineering technology includes different selection strategies that can minimise any disruption to genomic sequences within the BAC. Some methods can cause minimal alterations to a BAC sequence and leave a post-recombination scar, usually comprising recombinase target site such as FRT or loxP. Other methods use selection/counter selection, which results in a seamless alteration of the BAC, allowing for accurate manipulation without altering neighbouring sequences (e.g. the GalK system described by Warming *et al.* (2005)). This is important in the generation of fusion protein, expressing BACs where the reading frame of

the insertion is critical. A fusion protein might not behave naturally; therefore, transient transfection of the BAC construct should be performed initially to confirm expression of the gene-of-interest and the correct behaviour of the protein before undertaking the time-consuming step of stable clone generation. It is also advisable to perform initial studies with plasmid vectors that express the same (or very similar) fusion proteins before embarking on BAC construction.

There are a number of disadvantages associated with using BACs: i) a construct consisting of a large genomic fragment is likely to contain non-related genes that may lead to indirect, non-specific gene expression and potentially undesirable phenotypes; ii) the generation and screening of recombinant BAC constructs can take considerably more time than that of plasmids or other gene expression vectors; iii) large DNA constructs require special handling as they are more susceptible to degradation and shearing; and iv) recombination systems may suffer random recombination events, for instance the problems associated with cryptic LoxP sites resulting in random Cre-mediated recombination (Semprini et al. 2007). Moreover, BACs that contain repeating homologous sequences are prone to intramolecular rearrangements, not only reducing the efficiency of recombination but also, in some selection/counter-selection approaches there are a high number of false-positive clones, increasing the time spent screening (Narayanan 2008).

In summary, BACs have numerous advantages over conventional plasmids in that they insulate the gene from the site of integration effects, allow effective chromatinisation of the promoter DNA and include distal regulatory regions within the construct. Expression driven by the genomic region in which the gene is located results in accurate transcription regulation and promoter activity and promoter feedback. However, given their size, the technical difficulties in handling them properly and the potential for non-related gene expression, careful consideration of the investment in both time and resources for using BACs as a gene expression model system must be addressed.

Episomal gene expression systems

Integration effects can be avoided altogether through use of extra-chromosomal/episomal vectors designed to replicate independently of the genome (Fig. 3E). These systems fall into three categories: mammalian/human artificial chromosomes (MACs/HACs), viral episomal expression systems and vector-based episomal expression systems (Lufino et al. 2008).

MACs/HACs are chromosomal-derived vector systems, which replicate autonomously, segregate into daughter cells and are maintained at a low copy number (Conese et al. 2004, 2007). However, construction of these vectors is very difficult due to their large size and the time-consuming recombination techniques required, thus limiting the use of MACs and HACs as ectopic expression systems.

Many episomal systems are derived from viral sequences, a variety of which have been manipulated. The most successful

among these has been the HSV-derived constructs, which are well retained through generations and have a relatively large insert capacity (Lufino et al. 2008).

The original EPI vector (pEPI-1) was constructed by Piechaczek et al. (1999) and contains chromosomal scaffold/matrix attachment region isolated from the human β-interferon gene cluster (Bode et al. 1992). This region interacts with major nuclear matrix proteins (Jenke et al. 2002) and this is thought to enable co-replication with chromosomal DNA. The vectors are mitotically stable and exist at a copy number of around 5 to 15 copies per cell (depending on the stage of the cell cycle; Stehle et al. 2003), which is probably relatively low when compared with transiently transfected plasmids but compared to stably integrated plasmid vectors will not have aberrant site of integration effects on the host. Stable transfection with EPI vectors is often highly efficient, compared to the integration of conventional plasmid vectors. EPI vectors have also been shown to remain in cells in the absence of selection for several hundred generations (Piechaczek et al. 1999, Baiker et al. 2000) probably because there is no chance of heterochromatin spread.

Combining approaches for transgenesis

In recent years, attempts have been made to combine different techniques to minimise and negate some of the disadvantages, while exploiting the various advantages.

Targeted integration utilising AAV

As discussed, one of the major drawbacks of targeted integration is the random integration of recombinase target sites when generating 'founder' cell lines prior to integration of the gene-of-interest. DeKelver et al. (2010) have utilised the natural ability of AAV to target and integrate in a specific genomic locus on chromosome 19 (commonly referred to as AAVS1 locus; McCarty et al. 2004). Although integration at this site disrupts the PPP1R12C gene, this does not alter cellular phenotype (Smith et al. 2008, Hockemeyer et al. 2009) nor has any pathophysiological effect from AAV infection been observed. DeKelver et al. (2010) thus report this particularly well-characterised genomic region to be a transgene 'safe harbour', as no phenotypic effect has been observed by its disruption, yet the region is transcriptionally active. The authors exploited this natural AAV integration site to generate somatic cell lines with easily targetable recombinase sites at this locus. Founder line integration site is therefore consistent across cell types, thus reducing the time and effort necessary for cell line screening and characterisation prior to targeted integration of the gene-of-interest.

Using BACs for targeted 'knockin'

Creating knockin and knockout cell lines is a lengthy and challenging process. One of the key components of the development is the generation of the homologous

recombination cassette, usually several kilobases in length. Recently, Song *et al*. (2010) exploited the coverage of the human genome by BACs to generate a recombination cassette with extended homology arms for the creation of p53 knockout human ES cells. This BAC-based targeting approach had several advantages, such as the introduction of large DNA sequences, leading to a high efficiency of homologous recombination in various genetic backgrounds. This technique avoided the more complex and technically difficult cloning strategies by using BAC recombineering technology to generate the recombination cassette (Song *et al*. 2010). One reported drawback of this BAC-based targeting approach was the difficulty in confirming that the homologous recombination event had occurred. It should also be noted that ES cells are more amenable to such an approach compared to somatic cell lines. Thus, generating a recombinant BAC *in vitro* can improve targeted integration and recombination in *in vitro* cell cultures.

Episomal BAC expression

When producing cell lines for study, the uncontrolled integration of BACs into genomic DNA could lead to (albeit minimised) the site of integration effects, the BAC might be truncated unpredictably and the low transfection efficiency of BACs makes generating stable cell lines challenging.

Utilising BAC technology in combination with episomal sequences might be an effective way to stably introduce BAC constructs into the mammalian cells. Episomal constructs replicate autonomously within the cell and do not integrate in the genome, in addition to being relatively simple to use for generating stable cell lines. Lufino *et al*. (2007) were able to fuse the pEPI vector (Stehle *et al*. 2003) with both HSV viral sequences and a BAC to express the low-density lipoprotein (LDL) receptor in a cell line that was deficient in this gene. Subsequent viral infection of this construct resulted in LDL receptor-expressing cells, with the construct maintained at a low copy number ($\sim$2) in most cells for many generations. Thus, the combination of these two strategies led to exploitation of the major advantages of both systems and reduced the impact of the respective disadvantages for both techniques. Furthermore, the presence of HSV sequences within the EPI-BAC allows viral transduction of the construct, circumventing the problematic issues with BAC transfection. This also negates the need for special BAC DNA handling in the laboratory, and results in increased efficiency for delivery of DNA to target cells.

Conclusions

The study of gene expression is an ever expanding field of research, with a variety of technologies and approaches available. Important decisions are required when selecting a model system for analysis in any given project. All approaches discussed have their advantages and limitations, but methods are constantly evolving and adapting, expanding the options available to researchers.

References

Adamson AD, Friedrichsen S, Semprini S, Harper CV, Mullins JJ, White MRH & Davis JRE 2008 Human prolactin gene promoter regulation by estrogen: convergence with tumor necrosis factor-{alpha} signaling. *Endocrinology* **149** 687–694. (doi:10.1210/en.2007-1066)

Alwine JC, Kemp DJ & Stark GR 1977 Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *PNAS* **74** 5350–5354. (doi:10.1073/pnas.74.12.5350)

An JJ, Cho S-R, Jeong DW, Park KW, Ahn YS & Baik J-H 2003 Anti-proliferative effects and cell death mediated by two isoforms of dopamine D2 receptors in pituitary tumor cells. *Molecular and Cellular Endocrinology* **206** 49–62. (doi:10.1016/S0303-7207(03)00236-3)

Anastassiadis K, Kim J, Daigle N, Sprengel R, Schöler HR & Stewart AF 2002 A predictable ligand regulated expression strategy for stably integrated transgenes in mammalian cells in culture. *Gene* **298** 159–172. (doi:10.1016/S0378-1119(02)00979-4)

Archer TK, Lefebvre P, Wolford RG & Hager GL 1992 Transcription factor loading on the MMTV promoter: a bimodal mechanism for promoter activation. *Science* **255** 1573–1576. (doi:10.1126/science.1347958)

Archer TK, Lee HL, Cordingley MG, Mymryk JS, Fragoso G, Berard DS & Hager GL 1994 Differential steroid hormone induction of transcription from the mouse mammary tumor virus promoter. *Molecular Endocrinology* **8** 568–576. (doi:10.1210/me.8.5.568)

Ashall L, Horton CA, Nelson DE, Paszek P, Harper CV, Sillitoe K, Ryan S, Spiller DG, Unitt JF, Broomhead DS *et al*. 2009 Pulsatile stimulation determines timing and specificity of NF-{kappa}B-dependent transcription. *Science* **324** 242–246. (doi:10.1126/science.1164860)

Baiker A, Maercker C, Piechaczek C, Schmidt SB, Bode J, Benham C & Lipps HJ 2000 Mitotic stability of an episomal vector containing a human scaffold/matrix-attached region is provided by association with nuclear matrix. *Nature Cell Biology* **2** 182–184. (doi:10.1038/35004061)

Bell AC, West AG & Felsenfeld G 2001 Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* **291** 447–450. (doi:10.1126/science.291.5503.447)

Ben-Jonathan N, LaPensee CR & LaPensee EW 2008 What can we learn from rodents about prolactin in humans? *Endocrine Reviews* **29** 1–41. (doi:10.1210/er.2007-0017)

Bernichtein S, Touraine P & Goffin V 2010 New concepts in prolactin biology. *Journal of Endocrinology* **206** 1–11. (doi:10.1677/JOE-10-0069)

Bode J, Kohwi Y, Dickinson L, Joh T, Klehr D, Mielke C & Kohwi–Shigematsu T 1992 Biological significance of unwinding capability of nuclear matrix-associating DNAs. *Science* **255** 195–197. (doi:10.1126/science.1553545)

Bodner M & Karin M 1987 A pituitary-specific trans-acting factor can stimulate transcription from the growth hormone promoter in extracts of nonexpressing cells. *Cell* **50** 267–275. (doi:10.1016/0092-8674(87)90222-4)

Brown AM, Jeltsch JM, Roberts M & Chambon P 1984 Activation of pS2 gene transcription is a primary response to estrogen in the human breast cancer cell line MCF-7. *PNAS* **81** 6344–6348. (doi:10.1073/pnas.81.20.6344)

Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoute J, Shao W, Hestermann EV, Geistlinger TR *et al.* 2005 Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122** 33–43. (doi:10.1016/j.cell.2005.05.008)

Chudakov DM, Matz MV, Lukyanov S & Lukyanov KA 2010 Fluorescent proteins and their applications in imaging living cells and tissues. *Physiological Reviews* **90** 1103–1163. (doi:10.1152/physrev.00038.2009)

Conese M, Auriche C & Ascenzioni F 2004 Gene therapy progress and prospects: episomally maintained self-replicating systems. *Gene Therapy* **11** 1735–1741. (doi:10.1038/sj.gt.3302362)

Conese M, Boyd AC, Di Gioia S, Auriche C & Ascenzioni F 2007 Genomic context vectors and artificial chromosomes for cystic fibrosis gene therapy. *Current Gene Therapy* **7** 175–187. (doi:10.2174/156652307780859026)

Copeland NG, Jenkins NA & Court DL 2001 Recombineering: a powerful new tool for mouse functional genomics. *Nature Reviews. Genetics* **2** 769–779. (doi:10.1038/35093556)

Cuchet D, Potel C, Thomas J & Epstein AL 2007 HSV-1 amplicon vectors: a promising and versatile tool for gene delivery. *Expert Opinion on Biological Therapy* **7** 975–995. (doi:10.1517/14712598.7.7.975)

DeKelver RC, Choi VM, Moehle EA, Paschon DE, Hockemeyer D, Meijsing SH, Sancak Y, Cui X, Steine EJ, Miller JC *et al.* 2010 Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Research* **20** 1133–1142. (doi:10.1101/gr.106773.110)

De Lorenzi R, Gareus R, Fengler S & Pasparakis M 2009 GFP-p65 knock-in mice as a tool to study NF-kappaB dynamics *in vivo*. *Genesis* **47** 323–329. (doi:10.1002/dvg.20468)

Dillon N & Festenstein R 2002 Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends in Genetics* **18** 252–258. (doi:10.1016/S0168-9525(02)02648-3)

Eldar A & Elowitz MB 2010 Functional roles for noise in genetic circuits. *Nature* **467** 167–173. (doi:10.1038/nature09326)

Finkenstädt B, Heron EA, Komorowski M, Edwards K, Tang S, Harper CV, Davis JRE, White MRH, Millar AJ & Rand DA 2008 Reconstruction of transcriptional dynamics from gene reporter data using differential equations. *Bioinformatics* **24** 2901–2907. (doi:10.1093/bioinformatics/btn562)

Frawley LS, Faught WJ, Nicholson J & Moomaw B 1994 Real time measurement of gene expression in living endocrine cells. *Endocrinology* **135** 468–471. (doi:10.1210/en.135.1.468)

Gallagher PG, Nilson DG, Steiner LA, Maksimova YD, Lin JY & Bodine DM 2009 An insulator with barrier-element activity promotes {alpha}-spectrin gene expression in erythroid cells. *Blood* **113** 1547–1554. (doi:10.1182/blood-2008-06-164954)

Gariglio P, Bellard M & Chambon P 1981 Clustering of RNA polymerase B molecules in the 5′ moiety of the adult beta-globin gene of hen erythrocytes. *Nucleic Acids Research* **9** 2589–2598. (doi:10.1093/nar/9.11.2589)

Gavin A-C, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dumpelfeld B *et al.* 2006 Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440** 631–636. (doi:10.1038/nature04532)

Gellersen B, Kempf R, Telgmann R & DiMattia GE 1995 Pituitary-type transcription of the human prolactin gene in the absence of Pit-1. *Molecular Endocrinology* **9** 887–901. (doi:10.1210/me.9.7.887)

Gerlo S, Davis JR, Mager DL & Kooijman R 2006 Prolactin in man: a tale of two promoters. *BioEssays* **28** 1051–1055. (doi:10.1002/bies.20468)

Giepmans BNG, Adams SR, Ellisman MH & Tsien RY 2006 The fluorescent toolbox for assessing protein location and function. *Science* **312** 217–224. (doi:10.1126/science.1124618)

Guan Y, Dunham MJ & Troyanskaya OG 2007 Functional analysis of gene duplications in *Saccharomyces cerevisiae*. *Genetics* **175** 933–943. (doi:10.1534/genetics.106.064329)

Harper CV, Featherstone K, Semprini S, Friedrichsen S, McNeilly J, Paszek P, Spiller DG, McNeilly AS, Mullins JJ, Davis JRE *et al.* 2010 Dynamic organisation of prolactin gene expression in living pituitary tissue. *Journal of Cell Science* **123** 424–430. (doi:10.1242/jcs.060434)

Hasegawa K & Nakatsuji N 2002 Insulators prevent transcriptional interference between two promoters in a double gene construct for transgenesis. *FEBS Letters* **520** 47–52. (doi:10.1016/S0014-5793(02)02761-8)

Hebbar PB & Archer TK 2008 Altered histone H1 stoichiometry and an absence of nucleosome positioning on transfected DNA. *Journal of Biological Chemistry* **283** 4595–4601. (doi:10.1074/jbc.M709121200)

Hendrickson EA 2008 Gene targeting in human somatic cells. In *Source Book of Models for Biomedical Research*, pp 509–525. Ed. PM Conn. Totowa, NJ: Humana Press, Inc.

Hirata RK & Russell DW 2000 Design and packaging of adeno-associated virus gene targeting vectors. *Journal of Virology* **74** 4612–4620. (doi:10.1128/JVI.74.10.4612-4620.2000)

Hockemeyer D, Soldner F, Beard C, Gao Q, Mitalipova M, DeKelver RC, Katibah GE, Amora R, Boydston EA, Zeitler B *et al.* 2009 Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nature Biotechnology* **27** 851–857. (doi:10.1038/nbt.1562)

Howarth JL, Lee YB & Uney JB 2010 Using viral vectors as gene transfer tools (Cell Biology and Toxicology Special Issue: ETCS-UK 1 day meeting on genetic manipulation of cells). *Cell Biology and Toxicology* **26** 1–20. (doi:10.1007/s10565-009-9139-5)

Huszar T, Mucsi I, Terebessy T, Masszi A, Adamko S, Jeney C & Rosivall L 2001 The use of a second reporter plasmid as an internal standard to normalize luciferase activity in transient transfection experiments may lead to a systematic error. *Journal of Biotechnology* **88** 251–258. (doi:10.1016/S0168-1656(01)00277-2)

Ishikawa T, Lee EJ & Jameson JL 2004 Nonhomologous end-joining ligation transfers DNA regulatory elements between cointroduced plasmids. *Molecular and Cellular Biology* **24** 8323–8331. (doi:10.1128/MCB.24.19.8323-8331.2004)

Ishikawa TO, Jain NK, Taketo MM & Herschman HR 2006 Imaging cyclooxygenase-2 (Cox-2) gene expression in living animals with a luciferase knock-in reporter gene. *Molecular Imaging and Biology* **8** 171–187. (doi:10.1007/s11307-006-0034-7)

Jackson RJ & Standart N 2007 How do microRNAs regulate gene expression? *Science's STKE: Signal Transduction Knowledge Environment* **2007** p.re1.

Jenke BH, Fetzer CP, Stehle IM, Jonsson F, Fackelmayer FO, Conradt H, Bode J & Lipps HJ 2002 An episomally replicating vector binds to the nuclear matrix protein SAF-A *in vivo*. *EMBO Reports* **3** 349–354. (doi:10.1093/embo-reports/kvf070)

Jenke ACW, Stehle IM, Herrmann F, Eisenberger T, Baiker A, Bode J, Fackelmayer FO & Lipps HJ 2004 Nuclear scaffold/matrix attached region modules linked to a transcription unit are sufficient for replication and maintenance of a mammalian episome. *PNAS* **101** 11322–11327. (doi:10.1073/pnas.0401355101)

Jones BK, Monks BR, Liebhaber SA & Cooke NE 1995 The human growth hormone gene is regulated by a multicomponent locus control region. *Molecular and Cellular Biology* **15** 7010–7021.

Kim TK & Eberwine JH 2010 Mammalian cell transfection: the present and the future. *Analytical and Bioanalytical Chemistry* **397** 3173–3178. (doi:10.1007/s00216-010-3821-6)

Kim J-S, Bonifant C, Bunz F, Lane WS & Waldman T 2008 Epitope tagging of endogenous genes in diverse human cell lines. *Nucleic Acids Research* **36** e127. (doi:10.1093/nar/gkn566)

Kuhn EJ & Geyer PK 2003 Genomic insulators: connecting properties to mechanism. *Current Opinion in Cell Biology* **15** 259–265. (doi:10.1016/S0955-0674(03)00039-5)

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al.* 2001 Initial sequencing and analysis of the human genome. *Nature* **409** 860–921. (doi:10.1038/35057062)

Lufino MMP, Manservigi R & Wade-Martins R 2007 An S/MAR-based infectious episomal genomic DNA expression vector provides long-term regulated functional complementation of LDLR deficiency. *Nucleic Acids Research* **35** e98. (doi:10.1093/nar/gkm570)

Lufino MM, Edser PA & Wade-Martins R 2008 Advances in high-capacity extrachromosomal vector technology: episomal maintenance, vector delivery, and transgene expression. *Molecular Therapy* **16** 1525–1538. (doi:10.1038/mt.2008.156)

Magin-Lachmann C, Kotzamanis G, D'Aiuto L, Cooke H, Huxley C & Wagner E 2004 *In vitro* and *in vivo* delivery of intact BAC DNA – comparison of different methods. *Journal of Gene Medicine* **6** 195–209. (doi:10.1002/jgm.481)

McCarty DM, Young SM & Samulski RJ 2004 Integration of adeno-associated virus (AAV) and recombinant AAV vectors. *Annual Review of Genetics* **38** 819–845. (doi:10.1146/annurev.genet.37.110801.143717)

Mileyko Y & Weitz JS 2010 Bifurcation analysis of gene regulatory circuits subject to copy number variation. *SIAM Journal on Applied Dynamical Systems* **9** 799–826. (doi:10.1137/090771247)

Mutskov V & Felsenfeld G 2004 Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO Journal* **23** 138–149. (doi:10.1038/sj.emboj.7600013)

Narayanan K 2008 Intact recombineering of highly repetitive DNA requires reduced induction of recombination enzymes and improved host viability. *Analytical Biochemistry* **375** 394–396. (doi:10.1016/j.ab.2008.01.022)

Osten P, Grinevich V & Cetin A 2007 Viral vectors: a wide range of choices and high levels of service. *Handbook of Experimental Pharmacology* **178** 177–202. (doi:10.1007/978-3-540-35109-2_8)

Oumard A, Qiao J, Jostock T, Li J & Bode J 2006 Recommended method for chromosome exploitation: RMCE-based cassette-exchange systems in animal cell biotechnology. *Cytotechnology* **50** 93–108. (doi:10.1007/s10616-006-6550-0)

Palmiter RD & Brinster RL 1986 Germ-line transformation of mice. *Annual Review of Genetics* **20** 465–499. (doi:10.1146/annurev.ge.20.120186.002341)

Pfeifer A 2004 Lentiviral transgenesis. *Transgenic Research* **13** 513–522. (doi:10.1007/s11248-004-2735-5)

Piechaczek C, Fetzer C, Baiker A, Bode J & Lipps HJ 1999 A vector based on the SV40 origin of replication and chromosomal S/MARs replicates episomally in CHO cells. *Nucleic Acids Research* **27** 426–428. (doi:10.1093/nar/27.2.426)

Presser A, Elowitz MB, Kellis M & Kishony R 2008 The evolutionary dynamics of the *Saccharomyces cerevisiae* protein interaction network after duplication. *PNAS* **105** 950–954. (doi:10.1073/pnas.0707293105)

Qiao J, Oumard A, Wegloehner W & Bode J 2009 Novel tag-and-exchange (RMCE) strategies generate master cell clones with predictable and stable transgene expression properties. *Journal of Molecular Biology* **390** 579–594. (doi:10.1016/j.jmb.2009.05.012)

Qu Z, Thottassery JV, Van Ginkel S, Manuvakhova M, Westbrook L, Roland-Lazenby C, Hays S & Kern FG 2004 Homogeneity and long-term stability of tetracycline-regulated gene expression with low basal activity by using the rtTA2S-M2 transactivator and insulator-flanked reporter vectors. *Gene* **327** 61–73. (doi:10.1016/j.gene.2003.10.029)

Rafalska-Metcalf IU & Janicki SM 2007 Show and tell: visualizing gene expression in living cells. *Journal of Cell Science* **120** 2301–2307. (doi:10.1242/jcs.008664)

Rago C, Vogelstein B & Bunz F 2007 Genetic knockouts and knockins in human somatic cells. *Nature Protocols* **2** 2734–2746. (doi:10.1038/nprot.2007.408)

Raser JM & O'Shea EK 2004 Control of stochasticity in eukaryotic gene expression. *Science* **304** 1811–1814. (doi:10.1126/science.1098641)

Recillas-Targa F, Valadez-Graham V & Farrell CM 2004 Prospects and implications of using chromatin insulators in gene therapy and transgenesis. *BioEssays* **26** 796–807. (doi:10.1002/bies.20059)

Rogers S, Wells R & Rechsteiner M 1986 Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* **234** 364–368. (doi:10.1126/science.2876518)

Sabl JF & Henikoff S 1996 Copy number and orientation determine the susceptibility of a gene to silencing by nearby heterochromatin in *Drosophila*. *Genetics* **142** 447–458.

Sakabe NJ & Nobrega MA 2010 Genome-wide maps of transcription regulatory elements. *Wiley Interdisciplinary Reviews. Systems Biology and Medicine* **2** 422–437. (doi:10.1002/wsbm.70)

Sambrook J, Fritsch EF & Maniatis T 1989 In: *Molecular Cloning: a laboratory manual*, 2nd edn. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, NY.

Schlake T & Bode J 1994 Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry* **33** 12746–12751. (doi:10.1021/bi00209a003)

Sedivy JM, Vogelstein B, Liber HL, Hendrickson EA & Rosmarin A 1999 Gene targeting in human cells without isogenic DNA. *Science* **283** 9. (doi:10.1126/science.283.5398.9a)

Semprini S, Troup TJ, Kotelevtseva N, King K, Davis JRE, Mullins LJ, Chapman KE, Dunbar DR & Mullins JJ 2007 Cryptic loxP sites in mammalian genomes: genome-wide distribution and relevance for the efficiency of BAC/PAC recombineering techniques. *Nucleic Acids Research* **35** 1402–1410. (doi:10.1093/nar/gkl1108)

Semprini S, Friedrichsen S, Harper CV, McNeilly JR, Adamson AD, Spiller DG, Kotelevtseva N, Brooker G, Brownstein DG, McNeilly AS *et al.* 2009 Real-time visualization of human prolactin alternate promoter usage *in vivo* using a double-transgenic rat model. *Molecular Endocrinology* **23** 529–538. (doi:10.1210/me.2008-0399)

Shaner NC, Patterson GH & Davidson MW 2007 Advances in fluorescent protein technology. *Journal of Cell Science* **120** 4247–4260. (doi:10.1242/jcs.005801)

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y & Simon M 1992 Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *PNAS* **89** 8794–8797. (doi:10.1073/pnas.89.18.8794)

Smith JR, Maguire S, Davis LA, Alexander M, Yang F, Chandran S, ffrench-Constant C & Pedersen RA 2008 Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration. *Stem Cells* **26** 496–504. (doi:10.1634/stemcells.2007-0039)

Song H, Chung S-K & Xu Y 2010 Modeling disease in human ESCs using an efficient bac-based homologous recombination system. *Cell Stem Cell* **6** 80–89. (doi:10.1016/j.stem.2009.11.016)

Sparwasser T & Eberl G 2007 BAC to immunology – bacterial artificial chromosome-mediated transgenesis for targeting of immune cells. *Immunology* **121** 308–313. (doi:10.1111/j.1365-2567.2007.02605.x)

Spiller DG, Wood CD, Rand DA & White MRH 2010 Measurement of single-cell dynamics. *Nature* **465** 736–745. (doi:10.1038/nature09232)

Stehle IM, Scinteie MF, Baiker A, Jenke AC & Lipps HJ 2003 Exploiting a minimal system to study the epigenetic control of DNA replication: the interplay between transcription and replication. *Chromosome Research* **11** 413–421. (doi:10.1023/A:1024962308071)

Takasuka N, White MRH, Wood CD, Robertson WR & Davis JRE 1998 Dynamic changes in prolactin promoter activation in individual living lactotrophic cells. *Endocrinology* **139** 1361–1368. (doi:10.1210/en.139.3.1361)

Thomas H, Senkel S, Erdmann S, Arndt T, Turan Glz, Klein-Hitpass L & Ryffel GU 2004 Pattern of genes influenced by conditional expression of the transcription factors HNF6, HNF4α and HNF1β in a pancreatic β-cell line. *Nucleic Acids Research* **32** e150. (doi:10.1093/nar/gnh144)

Topaloglu O, Hurley PJ, Yildirim O, Civin CI & Bunz F 2005 Improved methods for the generation of human gene knockout and knockin cell lines. *Nucleic Acids Research* **33** e158. (doi:10.1093/nar/gni160)

Townes TM, Lingrel JB, Chen HY, Brinster RL & Palmiter RD 1985 Erythroid-specific expression of human beta-globin genes in transgenic mice. *EMBO Journal* **4** 1715–1723.

Tsien RY 1998 The green fluorescent protein. *Annual Review of Biochemistry* **67** 509–544. (doi:10.1146/annurev.biochem.67.1.509)

Turan S, Kuehle J, Schambach A, Baum C & Bode J 2010 Multiplexing RMCE: versatile extensions of the Flp-recombinase-mediated cassette-exchange technology. *Journal of Molecular Biology* **402** 52–69. (doi:10.1016/j.jmb.2010.07.015)

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA *et al.* 2001 The sequence of the human genome. *Science* **291** 1304–1351. (doi:10.1126/science.1058040)

Verma IM & Weitzman MD 2005 Gene therapy: twenty-first century medicine. *Annual Review of Biochemistry* **74** 711–738. (doi:10.1146/annurev.biochem.74.050304.091637)

Warming S, Costantino N, Court DL, Jenkins NA & Copeland NG 2005 Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Research* **33** e36. (doi:10.1093/nar/gni035)

West AG & Fraser P 2005 Remote control of gene transcription. *Human Molecular Genetics* **14** R101–R111. (doi:10.1093/hmg/ddi104)

White MR, Masuko M, Amet L, Elliott G, Braddock M, Kingsman AJ & Kingsman SM 1995 Real-time analysis of the transcriptional regulation of HIV and hCMV promoters in single mammalian cells. *J Cell Sci* **108** 441–445.

Wirth D, Gama-Norton L, Riemer P, Sandhu U, Schucht R & Hauser H 2007 Road to precision: recombinase-based targeting technologies for genome engineering. *Current Opinion in Biotechnology* **18** 411–419. (doi:10.1016/j.copbio.2007.07.013)

Xu M & Cook PR 2008 Similar active genes cluster in specialized transcription factories. *Journal of Cell Biology* **181** 615–623. (doi:10.1083/jcb.200710053)

Yáñez RJ & Porter ACG 1999 Influence of DNA delivery method on gene targeting frequencies in human cells. *Somatic Cell and Molecular Genetics* **25** 27–31. (doi:10.1023/B:SCAM.0000007137.28557.73)

Yanez-Munoz RJ, Balaggan KS, MacNeil A, Howe SJ, Schmidt M, Smith AJ, Buch P, MacLaren RE, Anderson PN, Barker SE *et al.* 2006 Effective gene therapy with nonintegrating lentiviral vectors. *Nature Medicine* **12** 348–353. (doi:10.1038/nm1365)

Zhang Y & Oliver B 2007 Dosage compensation goes global. *Current Opinion in Genetics and Development* **17** 113–120. (doi:10.1016/j.gde.2007.02.002)