Quantitative Analysis of DNA Transposon-mediated Gene Delivery: the *Sleeping Beauty* System as an Example

Orsolya Kolacsek
Institute of Molecular Pharmacology, RCNS
Hungarian Academy of Sciences, Hungary

Zsuzsanna Izsvák
Mobile DNA Group
Max-Delbrück Center for Molecular Medicine, Germany

Zoltán Ivics
Division of Medical Biotechnology
Paul Ehrlich Institute, Germany

Balázs Sarkadi
Membrane Research Group of HAS
Semmelweis University and National Blood Center, Hungary

Tamás I. Orbán
Institute of Molecular Pharmacology, RCNS
Hungarian Academy of Sciences, Hungary

1 Introduction – DNA Transposons as Genetic Tools

Transposons are special genetic elements that are capable of moving from one DNA locus to another. They were discovered in maize by Barbara McClintock in the 1940s, and these ground-breaking experiments later earned her the Nobel Prize in 1983. Transposons can be classified based on their replication intermediate (Class I retrotransposons and Class II DNA transposons), on their replication manner (replicative versus non-replicative transposons), or on their ability to disperse independently (autonomous versus non-autonomous elements) (Burns and Boeke, 2012). Once considered to be selfish units of replication, they are now believed to also represent inevitable driving forces for evolution, proven by their presence in all genomes examined so far (Kazazian, 2004; Hedges & Batzer, 2005; Feschotte & Pritham, 2007). The human genome is not an exception, as approximately 45% of our genetic material is made up of transposons (Biemont & Vieira, 2006; Mills *et al.*, 2006; Wicker *et al.*, 2007; Goodier & Kazazian, 2008; Solyom & Kazazian, 2012). The majority of the human transposons belong to the Class I retrotransposons (or RNA transposons) which move around by the replicative "copy and paste" mechanism. These contain currently active mobile elements which by all means played a significant role in human evolution (Mills *et al.*, 2007; Shen *et al.*, 2011). Moreover, LINE-1 retrotransposons were recently proven to be responsible for certain types of somatic mozaicisms present in vertebrate neurons (Singer *et al.*, 2010). Class II DNA transposons, however, make up a significantly smaller proportion (~3%) of the human genome. This moderate fraction might be attributable to the fact that the majority of them spread by the non-replicative "cut and paste" mechanism. In addition, none of the DNA transposons have been shown to be active in the human genome (Feschotte & Pritham, 2007; Collier & Largaespada, 2007; Izsvak *et al.*, 2010; Solyom & Kazazian, 2012).

Transposons as genetic tools have been widely used in invertebrate model organisms (*Drosophila* species or *Caenorhabditis elegans*), mainly exploiting active DNA transposons of the particular species. The most prominent examples are the P-elements which were used as proof of principle for the two component transposon gene delivery system in *Drosophila melanogaster* (Rubin & Spradling, 1982; Spradling & Rubin, 1982). For vertebrates, however, applications were limited to retrotransposons for a long time, with the obvious disadvantages of higher mutational rate (due to the reverse transcription process) and the long term genetic instability of the modified cells because of the potential remobilization of the integrated transgene (Uren *et al.*, 2005; Ostertag *et al.*, 2007). A significant breakthrough in vertebrate genetics was the creation of an artificial Tc1/Mariner-type transposon, the *Sleeping Beauty* (SB) system, which was the first DNA transposon proven to be active in vertebrates, including human cells (Ivics *et al.*, 1997). Its simple structure (Figure 1) made it easy to modify and to establish a controllable system by separating the transposase from its targets (the originally flanking terminal repeat sequences), thereby allowing the controlled delivery of any gene of interest into the genome (Izsvak *et al.*, 2000). However, the efficiency of the originally resurrected SB variant was still significantly lower than the widely used viral vectors so its potential seemed to be behind those other genetic vehicles, especially in human applications.

The success of "awakening" a new active DNA transposon initiated a wave of research aiming at establishing efficient novel transposon systems applicable in human cells. Apart from other reconstructed species (such as the *Frog Prince* from *Rana pipiens*, Miskey *et al.,* 2003), an active DNA transposon (*Tol2*) was discovered in medaka fish and successfully applied in various vertebrate species (Balciunas *et al.*, 2006). Moreover, a transposon from another insect species (*Trichoplusia ni*) called *piggyBac* (PB) was shown to be highly active in human cells (Ding *et al.*, 2005). In the meantime, hyperactive versions

of the previously used DNA transposons were also established (Zayed *et al.*, 2004; Baus *et al.*, 2005; Pledger & Coates, 2005), opening the possibility of efficient non-viral gene delivery applications. The most promising of all was the 100 times more active form of SB (SB100x), providing a highly effective alternative to the existing viral gene delivery methods (Mates *et al.*, 2009).

In a recent study, the most hyperactive versions of three transposon systems (SB, PB and *Tol2*) were systematically tested and compared in terms of delivery efficiency, copy number and integration profile of the transgene (Grabundzija *et al.*, 2010). It was revealed that SB and PB are the most efficient gene delivery vehicles and, although transposition efficiency is known to decrease with the cargo size (Izsvak *et al.*, 2010), they are able to carry and integrate inserts of ≥10 kb, outweighing the packaging capacity of the most efficient viral vectors. It was also shown that in conditions where the amount of DNA transposon is limiting (modeling most gene therapy applications), SB is superior even to PB. Moreover, the integration profile of SB seems to be the most favorable one: in fact, it is the closest to random on the genomic level among all tested viral- and transposon-based systems so far (Vigdal *et al.*, 2002; Liu *et al.*, 2005; Yant *et al.*, 2005; Grabundzija *et al.*, 2010), providing the lowest risk for insertional mutagenesis and making the SB system particularly suitable for gene therapy applications. In addition, as opposed to PB, no potential endogenous elements resembling SB are present in the human genome (Ivics *et al.*, 1997; Ivics *et al.*, 2004) which is an important safety issue that further supported the initiation of a clinical trial experiment using this transposon system (Williams, 2008).

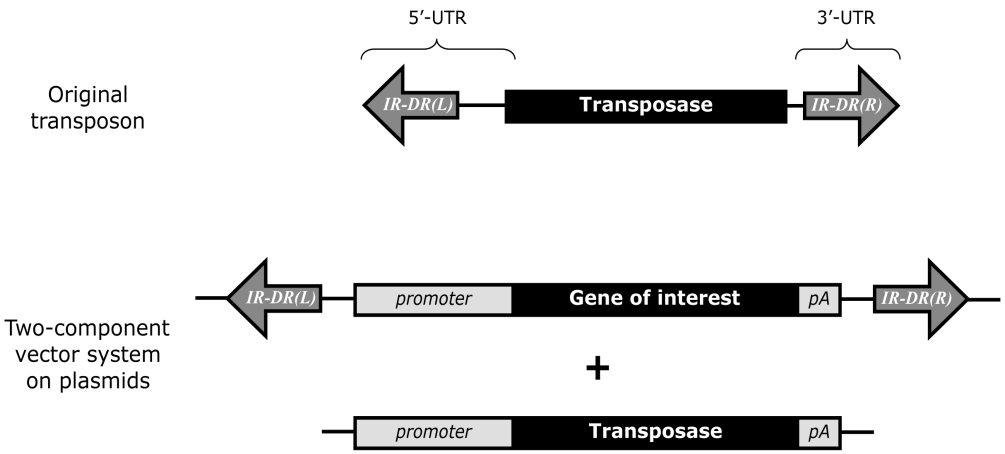

Figure 1: The structure of the *Sleeping Beauty* transposon system. In the natural transposon, the transposase gene is flanked by untranslated regions (UTRs) that include the terminal Inverted Repeat Direct Repeat regions (IRDR-L and IRDR-R), containing binding sites for the transposase. When used as a gene delivery vector system, the transposase coding region is replaced by a gene of interest within the transposable element that is maintained in a plasmid. This non-autonomous transposon can be mobilized if the transposase is supplied in trans by expression from a separate plasmid vector containing a suitable promoter. Co-transfection of the two components into candidate cells provides the platform for transposition from the donor plasmid to the cellular genome.

Considering all aspects, however, other efficient transposon systems with integration preferences into transcriptionally active regions (such as the PB system, Wilson *et al.*, 2007) might be more suitable

when performing "traditional" forward genetic screens (Collier & Largaespada, 2005; Chew *et al.*, 2011; Guo *et al.*, 2011). Also, as the PB transposase is able to "tracelessly" remove the integrated transgene from the genome, it might be a method of choice when such a feature is desirable, exemplified by the removal of the reprogramming cassette after generating induced pluripotent stem cells from fibroblasts (Kaji *et al.*, 2009; Woltjen *et al.*, 2009). Nevertheless, there are still various aspects of the transposon-based technology that should be rigorously tested. Such issues include the potential silencing of the transgene which is often the drawback of viral vectors (Ellis, 2005), especially in the case of embryonic stem cells which are particularly prone to silence viral promoters (Meilinger *et al.*, 2009; Rowe *et al.*, 2010). So far, the already applied DNA transposon sequences did not seem to face this problem as in the case of SB the effect of silencing was shown to depend on the cargo sequence, and not on the transposon vector (Garrison *et al.*, 2007; Zhu *et al.*, 2010).

According to gene therapy guidelines or mutagenesis protocols, one of the most important issues is the exact and fast determination of transgene copy numbers (Bian & Belmont, 2010; Sivalingam *et al.*, 2010; Huang *et al.*, 2010). Various methods are available to perform this, including "traditional" blotting techniques (Southern blotting/dot blotting), or several polymerase chain reaction (PCR)-based techniques (Wicks *et al.*, 2000; Devon *et al.*, 1995). These usually involve the application of radioactively or fluorescently labeled probes, or – depending on the nature of the transgene – utilize the inherent signal originating from the transgene itself (such as quantifying GFP fluorescence, Moeller *et al.*, 2003). Apart from often requiring hazardous chemicals or being laborious, the general problem of these methods is that they are usually set up for a specific transgene, and changing the gene of interest will require optimizing the applied parameters of the method once again. Using the SB system in our laboratory, we aimed to develop an accurate and fast method to quantify transposon copy numbers that is applicable to any SB-based gene delivery experiments without *a priori* optimization of the protocol. We worked out a real-time PCR technique which is independent of the transgene sequence, hence we named it a "transgene independent" quantitative PCR technology (Kolacsek *et al.*, 2011). Apart from being sensitive, accurate and fast, this approach also offers a powerful non-radioactive technique as an alternative against other canonical methodologies.

In this chapter, using the SB transposon system as a prominent example, we address quantitative issues regarding the transposon-based gene delivery methods. We focus in detail on a transgene-independent qPCR method recently developed in our laboratory, providing more information in depth on the theoretical background and the technical aspects of this methodology published earlier (Kolacsek *et al.*, 2011). Additionally, we also cover a technique of measuring transposase activity at the excision phase of the reaction which allows comparative analysis of different transposase variants, different transposon systems or different conditions of application. We believe that these quantitative aspects of transposase activity are of great importance especially in light of the applicability of the DNA transposons for gene therapy purposes.

2 Selecting and Separating Transgenic Clones after Transposition

2.1 Selection Methodology

The basis of generating stable transgenic clones is to apply the most efficient but the least harmful gene delivery into the chosen cell types. The SB transposon is the method of choice in this respect as it offers a stable transgene integration technology with the least mutagenic potential among all available gene de-

livery techniques (see Section 1 above). Nevertheless, the bottleneck of this technology is that it involves the transfection of DNA into the host cells which might have low efficiency rates for certain cell types, such as embryonic stem cells. It is therefore necessary to apply an efficient and preferably non-invasive selection protocol to establish homogenous transgene expressing cells following transfection and transposition. Selection methods may sometimes be carried out utilizing the expression of the transgene but very often it is inescapable to use an additional marker gene even at the expense of having a larger genetic cargo, thereby lower delivery efficiency.

In applications where the cell source is not limiting, various chemical selection methods such as antibiotic selection can be applied to enrich for transgene expressing cells (Figure 2). However, depending on the cell type, this method might significantly disturb cell physiology, therefore other approaches are necessary to be applied. For example, the advantage of using fluorescent markers is that the transfected cells can be separated by Fluorescent Activated Cell Sorting (FACS) analysis (Figure 3), although some cell types may not tolerate such physical stress and this method could also decrease cell viability. The aim is to optimize the transfection/selection procedure for the particular cell type reaching the highest possible gene delivery efficiency with the lowest possible cell mortality rate.

2.2 Separating uniform clones

Some transgenic applications (e.g. transgenic animals) require genetically uniform cell populations, which can be precisely characterized from various aspects, including copy number and integration sites. These features can contribute to the transgenic phenotype to a large extent. To develop a "reliable" designated method for copy number determination, our goal was to detect the lowest (1) stable copy per cell, and to clearly differentiate cells differing in copy numbers by one (e.g. cells carrying 2 copies from those with 1 or 3 copies). In other words, with this method we wanted to detect the copy number as corpuscular units in the cells. This was an important reason why we aimed at working with genetically uniform clones that carry the same transgenic cassette in different copy numbers.

The simplest way for cloning is the threshold limit dilution of heterogeneous transgenic cell population, previously selected by the transgene or the marker gene expression. In this method, a serial dilution of the cells is spread in a 96-well cell culture plate and those wells are considered to represent one clone where only one colony can be seen by microscopy; those cell clones are then further maintained and utilized. However, this method cannot be applied to all cell types, such as human embryonic stem cells which naturally grow only in clumps. In such cases, other manual methods may be applied, including FACS selection or using cloning rings.

2.3 Isolation of a Single Copy Insertion Serving as a Calibration Unit

The major difficulty to start a particular copy number measurement project is the lack of reference samples with known copy numbers, also known as calibrators. If such samples determined by other techniques are not available, the first step is to isolate clones carrying 1 copy of the transgene.

Applying the hyperactive SB transposases often results in high copy numbers and working with such high efficiency transposon delivery may provide only small number of cells carrying 1 copy of the transgene. An obvious 1 copy clone source is the random integration of plasmids that most of the cases results in 1 transgenic copy. These integrations will contain the majority or the whole plasmid sequence, due to a random breakage of the transfected transposon vector. This can be achieved by transfecting the transposon donor plasmid either with the inactive mutant transposase variant or without the transposase expressing helper plasmid. Random integration is very ineffective, occurs usually in less than 1% of the

transfected cells, but they can be selected out and can be cloned as well. This transfection serves as transposition control (Figure 2) resulting in traces of random integrations contrary to the active transposition, which has much higher integration efficiency, and it could provide us an excellent source of one copy clones.

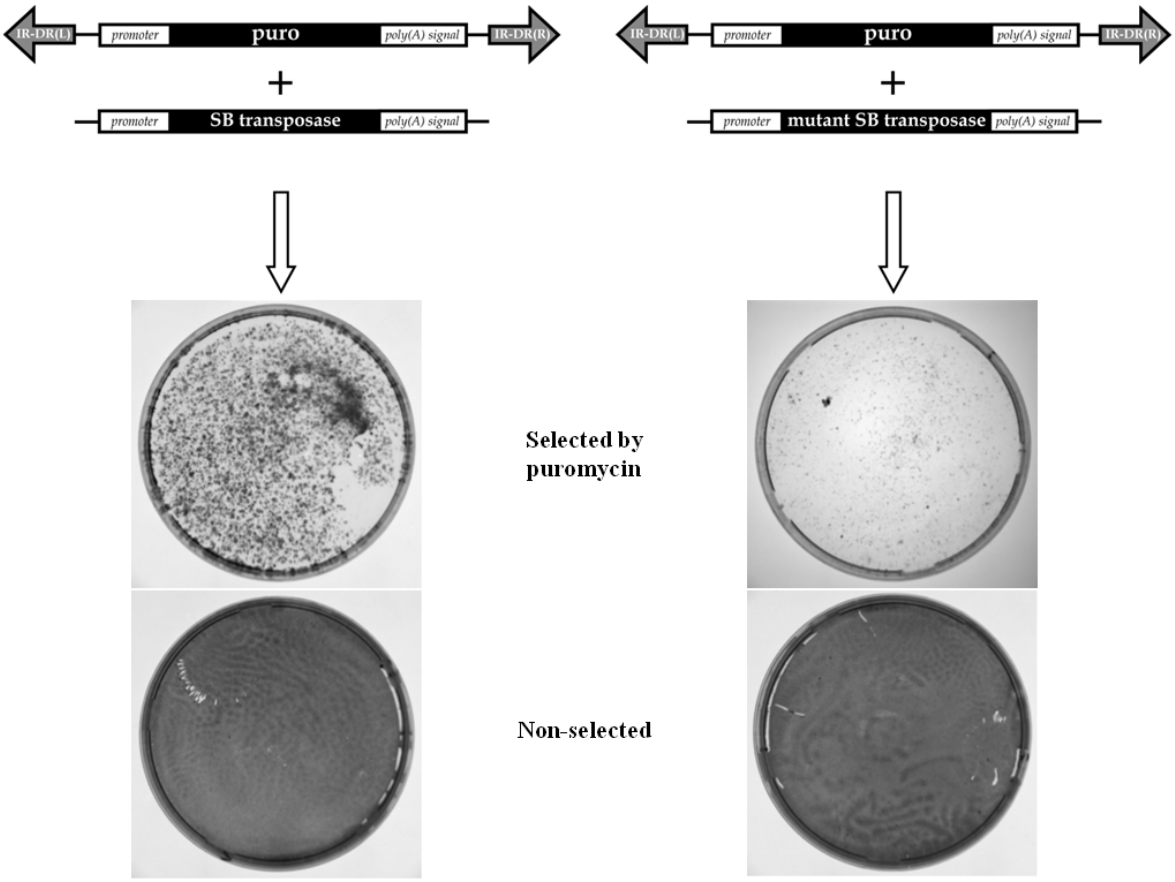

Figure 2: A typical experiments using SB tranposons: establishing transgenic HEK-293 cells expressing a puromycin resistance gene (puro). 2 days posttransfection, cells were passed into puromycin containing medium and selected for 10 days; living cells were visualized by Giemsa-staining following selection. The efficiency of transposition is obvious when comparing the selected cells after co-transfection with the transposase expressing helper plasmid (left) to the control experiment with the mutant transposase (right), the latter one indicating random integration events. As a negative control, non-selected cells are also shown in both experimental setups.

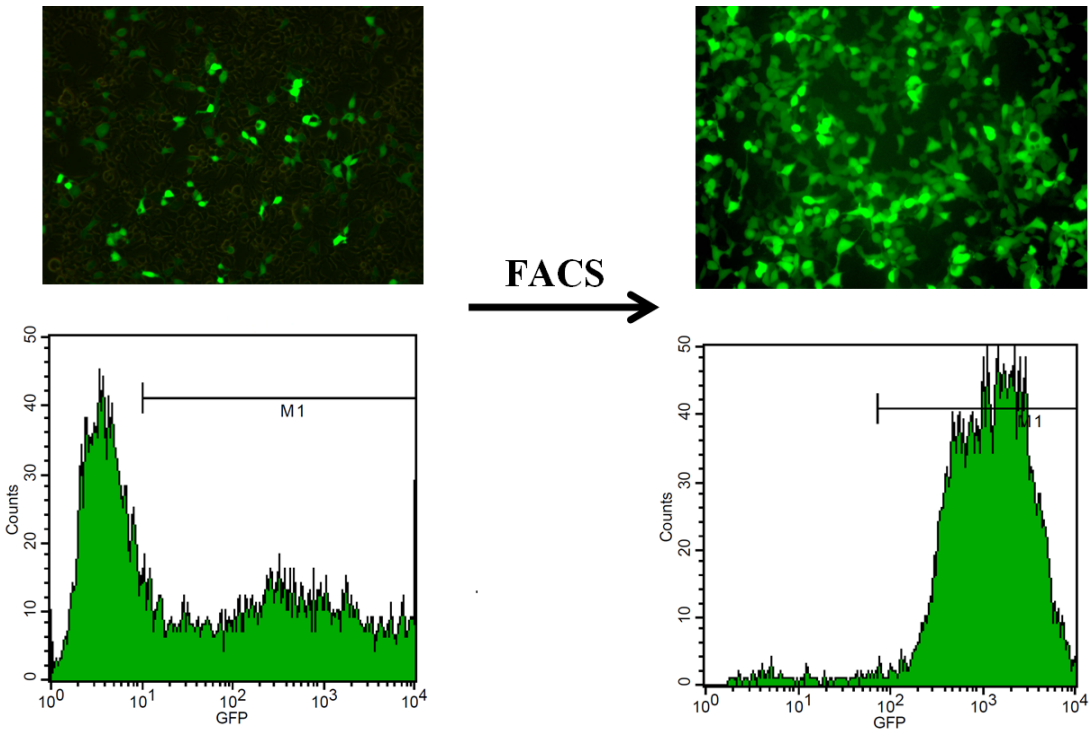

Figure 3: Transgene selection after transposition using GFP expression. The expression cassette is delivered into HEK-293 cells by the SB transposon system resulting in a heterogeneous cell population (left panels). Transgenic cells were selected by FACS experiment (right panels). Fluorescence microscopy images of x200 magnification and FACS histograms of GFP intensities can be seen. Cell numbers (Counts) are shown as a function of fluorescence intensity plotted in a logarithmic scale. M1: marker indicating GFP expressing cells.

3 Real-time PCR for Sequence Quantification

The polymerase chain reaction (PCR) is a technique for the in vitro amplification of specific DNA sequences by the simultaneous primer extension of complementary strands of DNA (Kleppe *et al.*, 1971; Mullis & Falona, 1987). It was a major development in molecular biology because it has simplified existing technologies and enabled a rapid development of new techniques which otherwise would not have been possible. PCR theoretically amplifies DNA exponentially, doubling the number of double stranded sequences present in each amplification cycle. After the logarithmic (log) phase of the reaction, the amount of the PCR product reaches the plateau phase (Figure 4). The amount of the product is proportional to the starting sequence copy number during the log phase, providing the basis for reliable quantitative comparisons.

In traditional (endpoint) PCR, detection and quantification of the amplified sequence are performed at the end of the reaction after the last PCR cycle, and involve post-PCR analysis such as gel electrophoresis, signal detection or image analysis. However, this allows only semi-quantitative analysis

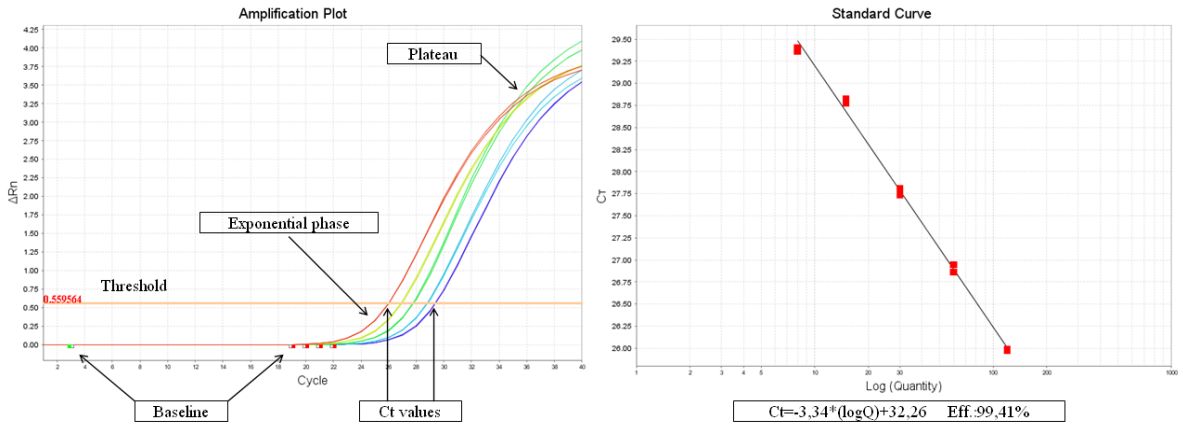

Figure 4: Example of a real-time PCR experiment determining the efficiency of the GFP TaqMan® assay using a standard curve of dilution points. Twofold dilutions were prepared from gDNA of pooled transgenic clones with the same copy number. Measurements were performed on a StepOnePlus™ Real-Time PCR platform (Applied Biosystems, Foster City, CA). Data were analyzed with the StepOne™ Software v2.1. Amplification plot (left panel) shows the increase of fluorescence signal as a function of cycle numbers (baseline fluorescence is subtracted, hence ΔRn; different colors represent different dilution reactions). Typical parameters of the reaction are indicated. The right panel shows the standard curve derived from the experiment, showing Ct values as a function of gDNA input (expressed in logarithm of DNA quantity in nanograms). The line of best fit is calculated by linear regression using the standard points; the equation is also shown from which efficiency of the reaction is calculated.

carried out in samples collected at multiple points throughout the amplification process, thus ensuring the analysis before the plateau is reached. This approach is usually combined with analysis of dilution series of the samples, it also requires known standards and provides a detection range of usually only tenfold difference (Chelly *et al.*, 1988; Wang *et al.*, 1989). In real-time quantitative PCR (qPCR), the amount of PCR product is measured at each cycle by the use of fluorescent dyes (Higuchi *et al.*, 1992; Livak *et al.*, 1995). This ability to monitor the reaction during its exponential phase enables the user to determine the initial amount of target with great precision. Apart from being simple and fast, the powerful benefit of qPCR is the increased dynamic range of comparisons.

The most popular fluorescent detection technologies are double-stranded DNA (dsDNA) binding agents, e.g. SYBR® Green, and fluorescent probes (Livak *et al.*, 1995; Wittwer *et al.*, 1997; Morrison *et al.*, 1998). SYBR® Green signal is measured at the end of each extension step and the intensity depends on the amount of dsDNA that is present. This technology is simple because the dye can be added to any kind of sequence amplification, but lacks specificity because it will also bind to PCR artifacts, e.g. primer-dimers. Good primer design and quality of starting materials are critical to avoid nonspecific products. Specificity of the reaction could be assessed using a melting curve measured at the end of the reaction (Figure 5). Melting curve determines the melting point (Tm) characteristic to the specific PCR product which assures differentiation of valid qPCR reactions from PCR artifacts (Ririe *et al.*, 1997). Those melting curves showing multiple peaks or one peak with rather different Tm than of the main product result from nonspecific PCR products indicating invalid reactions with false Ct values.

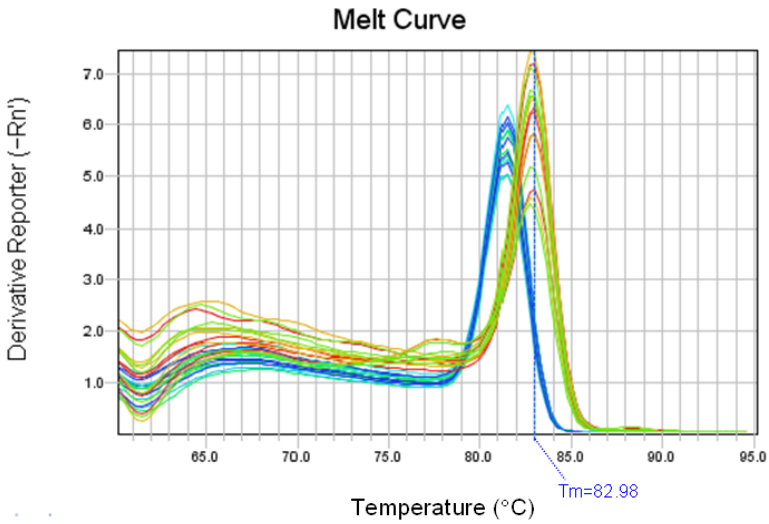

Figure 5: Representative melting curve analysis of a SYBR® Green qPCR experiment. Derivative of the fluorescence signal of the reporter (Rn') is shown as a function of the temperature. Two sequences (a target and an endogenous control) were amplified in various wells of this plate, therefore two melting peaks are visible; the melting temperature of the target is indicated. Different colors represent different reaction wells of the plate.

Fluorescent oligonucleotide probes (e.g. TaqMan® probes) are designed to hybridize to the sequence amplified by the primers. At the annealing step of each cycle, the probe will bind to the target sequence, and will be subsequently cleaved by the 5' nuclease activity of the polymerase during the extension phase. A dual labeled probe with a reporter dye at the 5' end and a quencher dye at the 3' end of the oligonucleotide will generate a fluorescent signal when the probe is degraded which is detected at the end of the extension phase (Livak *et al.*, 1995; Heid *et al.*, 1996). Probe-based systems provide highly specific detection of DNA, however, dual-labeling and complex design make them more expensive.

In qPCR, the cycle number in which the signal appears at the beginning of the log phase is considered to be inversely correlated to the amount of starting template, as a higher amount will result in sooner amplification. Threshold is the level of signal that reflects a statistically significant increase over the baseline fluorescent signal (Figure 4). In most cases, the real-time PCR software automatically sets the threshold at least 10 times the standard deviation of the fluorescence value of the baseline. However, the positioning of the threshold can be set manually at any point in the exponential phase of PCR. Threshold cycle (Ct) is the cycle number determined by the software at which the amplification plot crosses the threshold. Passive reference dyes (usually added to the qPCR master mixes) are frequently used in qPCR to normalize the fluorescent signal of reporter dyes (Rn). This allows the correction of fluctuations in fluorescence that is non-PCR-based, e.g. changes from well to well in reagent concentration or volume, or in instrument scanning (Figure 4).

Validation of the qPCR assay is generally carried out by the analysis of the slopes from standard curves. A standard curve is generated by plotting the results of a dilution series of the template against the Ct for each dilution (Figure 4). In theory, if the reaction is 100% efficient, the PCR duplicates the template in each cycle, and in the log scale of template amount, the slope will be –3.32 (1/[lg(x)-lg(2x)] =

–3.32). The reaction efficiency is related to assay sensitivity, which can be calculated from the slope (Real Time PCR Handbook, http://tools.invitrogen.com/content.cfm?pageid=12257):

$$\text{Efficiency} = 10^{(-1/\text{slope})} - 1.$$

The template used to generate the standard curve should match – as closely as possible – that is being used for the experiment (e.g. the same total RNA or DNA sample). The dilution range or dynamic range should span the concentration range expected for the unknown samples. The simplest way to ensure this is pooling the unknown samples (such as gDNAs from transgenic clones), and using it as a standard. The acceptable range of the efficiency, which most scientists agree on is between 90% – 110%. If efficiency is higher than 100%, it can reflect an inhibitory effect. In this case, scaling down the starting material usually helps by lowering the concentration of the suspected inhibitor. The desirable window of 90 to 110% defines the range of input template quantities that may be measured in a particular qPCR.

In the following three sections, we describe three different approaches which could be applied to determine transgene copy numbers of transgenic clones using a real-time PCR-based strategy.

4 Determining Transgene Copy Numbers of Transgenic Clones

4.1 First Approach: Verifying Presumably One Copy Clones with Absolute Quantification using the Marker Sequence

We generated SB transgenic clones carrying a GFP expressing cassette separating them by FACS analysis. Few clones were prepared also from random integration experiments, being obvious sources of 1 copy integrations as mentioned formerly. Since random integrations can be derived from breakage of the plasmid at any point, the presence of fluorescence in these clones assures the presence of the GFP transcription unit, so the GFP sequence can be reliably used for the copy number analysis. Therefore, specific TaqMan® assays were designed for the two terminal Inverted Repeat Direct Repeat (IRDR) motifs of the SB transposon (left and right, IRDR-L and IRDR-R), as well as for the GFP sequence. Sequences of primers and probes can be found in our previous publication (Kolacsek *et al*., 2011).

In the absence of a reference clone with known copy number, we have to compare the absolute Ct values to known plasmid dilutions containing the transposon sequence. The recommended amount of gDNA input is in the range of 10 to 40 ng (we used 30 ng) but other input size of the starting material can also be accepted if in previous pilot experiments, the efficiency at that point was shown to be in the desired range. For calculations of the required plasmid dilutions, first we need to know how many genome copies are present in the 30ng gDNA input. The average molecular weight of a DNA base pair is 618g/mol, so using the Avogadro's number of 6.02×10^{23} entities/mol, the molecular weight of a single haploid genome is (3×10^9 bp/genome × 618 g/mol) / 6.02×10^{23} = 3.08 pg. Therefore, 30ng gDNA contains 30000 pg / 2 × 3.08 = 4870 copy of diploid cell genome and so a one copy clone must contain 4870 copy of the transgene. (Genome weight of other species can be found in the genome size database at www.genomesize.com.) Since we used a plasmid of 5800 bp in length, for the signal equivalent to the single copy clone, we needed approximately 4870 × (5800 bp × 618 g/mol) / 6.02×10^{23} = 0.029 pg of plasmid input. Considering a plasmid with concentration of 100 ng/µl, at least a 3.45×10^6-fold dilution is required to be in the similar range. The equivalent plasmid amount was put in the middle of our standard curve and two neighboring points of twofold dilutions were taken for the standard curve (Figure 6). With this setup applying the GFP TaqMan® assay, 5317 ± 195 copy was calculated for a randomly inte-

grated GFP expressing clone, slightly differing from the desired 4870, estimating the copy number as (5317 ± 195) / 4870 = 1.09 ± 0.04.

In absolute quantification, each template must have sufficient purity and the input amount needs to be precisely quantified. The accuracy of the assay is directly related to the quality of the standard curve. Several dilution steps preceding each assay have to be performed with special attention, however, no matter how much care is taken, real-time PCR sensitivity amplifies minute human errors. In addition, the plasmid template used to generate the standard curve might not be an ideal specimen as it might not really represent the complex properties of the unknown samples. Due to such difficulties, absolute quantification seemed inconvenient for our routine transposon applications. Nevertheless, once we have successfully selected one copy clones, we could use them as reference samples for comparative analysis.

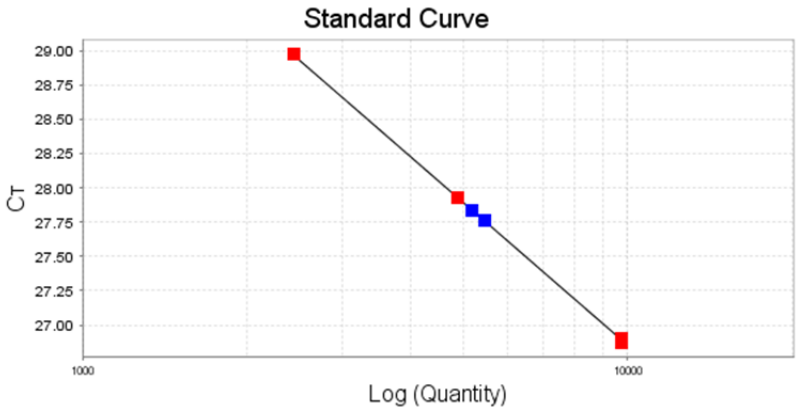

Figure 6: Absolute quantification of a presumably one copy clone using the GFP TaqMan® assay: comparison of the Ct values to known plasmid dilutions. Ct values are shown as a function of the quantity. Standard curve points of plasmid dilutions around the expected quantity are shown in red, whereas Ct values measured for the clone were put on this curve and are shown in blue. Quantity is illustrated as plasmid copy number input. The interpolated copy number for this clone was around 5300. See more informations and calculations in the text.

4.2 Second Approach: Comparative Quantification using One Copy Calibrator Samples

The alternative to identify a potential calibrator clone is to screen for the lowest transgene or marker gene expression. We have analyzed GFP expressing HEK-293 clones by FACS, and measured different green fluorescent intensities (Figure 7). The fluorescence intensities of the selected clones were compared to that of a one copy clone previously identified by absolute quantification. The GFP intensity of this latter one was similar to most of the lowest level expressing clones confirming that these clones are suitable for reference samples with one copy integration for subsequent analysis. Although expression level can be affected by its insertion site, we may say that one copy clones have significantly and uniformly lower GFP intensity than the few copy ones, but other transgene expressions may show wider differences.

Comparative qPCR quantification, while still technically challenging, does not require the same level of stringency. In this approach, the assay for target sequence is compared to a reference sample (a calibrator), and instead of precise copy number determination, it focuses on relative fold changes. In our application the copy number can be calculated based on the relative quantity of a single copy insertion. The method is based on the assumption that the threshold number of the sample and reference molecules

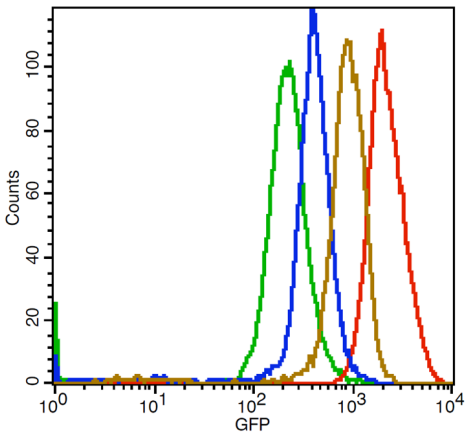

Figure 7: FACS histograms of representative HEK-293 clones expressing GFP as a transgene. Differences in GFP intensities correlate with GFP copy numbers.

is equal (Livak, 1997):

$$R_{Ct\ Sample} = R_{Ct\ Reference}$$

where RCt stands for the number of molecules at the threshold cycle. With the theoretical 100% efficiency, PCR duplicates the target in each cycle:

$$R_{Ct\ Sample} = R_{0\ Sample} \times 2^{Ct\ Sample}$$

$$R_{Ct\ Reference} = R_{0\ Reference} \times 2^{Ct\ Reference}$$

where R0 Sample and R0 Reference are the initial number of molecules of the sample and reference, respectively. The fold difference can therefore be calculated as:

$$R_{0\ Sample} / R_{0\ Reference} = 2^{-\Delta Ct},$$

where:

$$\Delta Ct = Ct_{Sample} - Ct_{Reference}$$

However, when the efficiency is not 100% but it is reproducibly identical, efficiency correction should be incorporated into the ΔCt method.

$$\text{Fold difference: } (1 + E)^{-\Delta Ct}.$$

For each novel assay, it is advisable to determine the efficiency values using the standard curve methodology discussed previously. As formerly mentioned, the most suitable specimen for such standard curve analysis is pooled gDNA of transgenic clones, because it has similar complexity to the unknown samples being analyzed. For comparison, the interpolated values from the standard curves can be used. A concrete example for the copy number calculation is as follows: the derived amount of the unknown sample is 24,99±0,15ng, whereas that of the single copy reference sample is 6.37 ± 0.18ng, so:

Fold difference – unknown/reference: (24.99 ± 0.15ng)/(6.37 ± 0.18ng)

Relative quantity (RQ) = 3.93 RQ Min = 3.80 RQ Max = 4.06

The upper example is still based on comparison of the absolute values of target Ct, however, a single Ct does not always reflect the expected number because of specimen discrepancies. Abandoning absolute quantification did not eliminate the deviations arising from differences in sample quality of the compared clones. Although the input of the template is always standardized, Ct deviations can still be attributable to certain errors, such as DNA concentration measurements. Normalization to an endogenous control sequence can overcome this problem. The control sequence should be similar in abundance to the target sequence and it must be present at a consistent level among all samples being compared. As an endogenous control for the human genome, the RPPH1 gene (the H1 RNA subunit of the RNaseP enzyme complex) was chosen which is a widely-accepted one copy gene of the haploid human genome. As mentioned earlier, without knowing the efficiency values, standard curves of both sequences have to be applied (Figure 8).

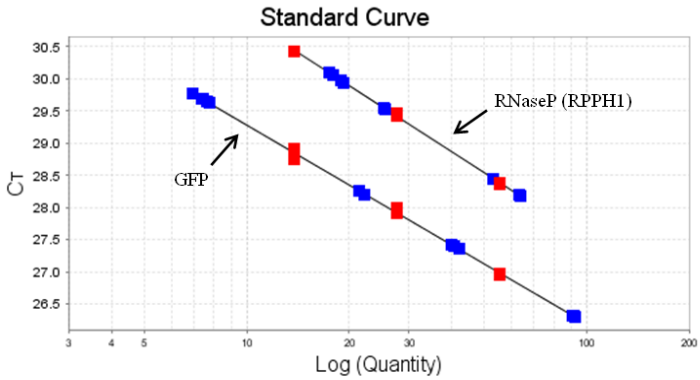

Figure 8: Example of a relative standard curve experiment. Standard curves for GFP and RNaseP (RPPH1) TaqMan® assay can be seen. Dilutions were made from pooled gDNA of unknown clones analyzed in the measurement; quantity is expressed in nanograms of gDNA input. Regression lines calculated from the measurement points of the two assays are parallel to each other, indicating very similar efficiency values. Red points represent points of the standard curves, whereas blue ones represent measurements of examined clones. See more informations and calculations in the text.

A relative standard curve experiment is based on sequential comparisons. First, both the target and the endogenous control are interpolated separately from the respective standard curves. Specimen comparisons showed formerly can be carried out only after normalizations to the endogenous control, for example:

Normalizations – Target/EndCont:

Sample → (54.23 ± 2.60ng) / (42.34 ± 0.82ng)

Reference → (6.13 ± 0.70ng)/(19.43 ± 0.25ng)

Fold difference – Sample/Reference:

Relative quantity (RQ) = 4.07 RQ Min = 3.76 RQ Max = 4.35

Alternatively, comparative analysis could be based on the real efficiency values:

Fold difference – Difference in the target/Difference in the endogenous control:

$$(1 + E_{Target})^{-\Delta Ct\ Target} / (1 + E_{EndCont})^{-\Delta Ct\ EndCont}$$

where:

$$\Delta Ct_{Target} = Ct_{Target\ Sample} - Ct_{Target\ Reference}$$

$$\Delta Ct_{EndCont} = Ct_{EndCont\ Sample} - Ct_{EndCont\ Reference}$$

However, if efficiencies of both the target and the endogenous control are proven to be close to identical, the ΔΔCt method could be chosen:

$$\text{Fold difference: } (1 + E)^{-\Delta\Delta Ct}$$

where:

$$\Delta\Delta Ct = \Delta Ct_{Target} - \Delta Ct_{EndCont}$$

Ideally, efficiencies of both the target and the endogenous control are close to 100%.

$$\text{Fold difference: } 2^{-\Delta\Delta Ct}$$

The main requirement for the ΔΔCt method is that the efficiencies of the assays are identical. The deviation of the two efficiencies can be determined by plotting the ΔCt-s from the standard curves points, and if the slope of this so called relative efficiency plot (Figure 9) is in the range of < 0.1, then it is acceptable to employ the ΔΔCt method. Another way to test the applicability is analyzing the relative standard curve experiment data with the ΔΔCt method. Similar outcome of the two methods will show the acceptability of the comparative Ct method, and in such cases, standard curves can be leaved behind. As the GFP sequence worked well in combination with the RPPH1 endogenous control, we continued to examine if assays designed for SB transposon sequences can also be utilized for copy number determination with similar methodology.

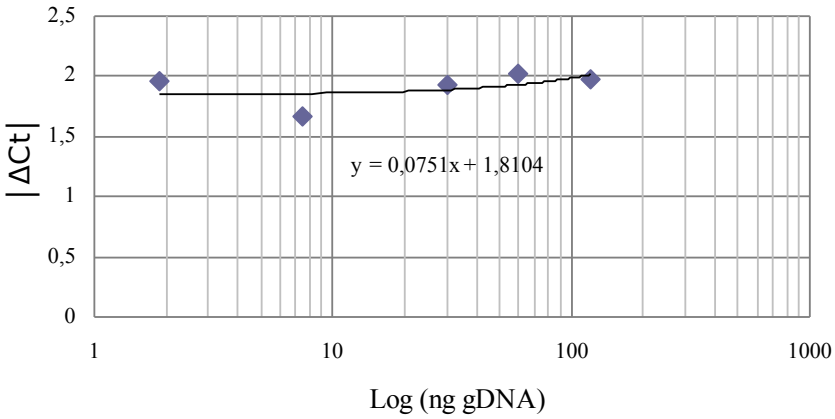

Figure 9: Relative efficiency plot comparing TaqMan® assays of IRDR-L and RNaseP (RPPH1). Dilution series were prepared from pools of gDNA containing one copy of SB transposon. Absolute values of ΔCt (target minus endogenous control) from the standard curve points were plotted against gDNA quantities and the equation of the regression line calculated from the measurement points is shown. A line with a slope of very small value (close to 0) indicates identical efficiencies of the two assays across all input concentrations.

4.3 Third Approach: Comparative Quantification using the SB IRDR Sequence Independently of the Transgene

As the SB transposon system is generally applied in our laboratory, we aimed at developing a real-time PCR-based technique that is transgene-independent, specific for the transposon regions, and therefore widely applicable. As for most transposon flanking sequences, the two IRDR regions are repeat-rich DNA sequences which make PCR primer design relatively difficult. Moreover, the left and the right IRDRs are very similar to each other which further narrows the possibility to design specific assays for them. Nevertheless, we could still develop specific TaqMan® assays for each: neither of the IRDR-L nor the IRDR-R probe set gives signals in the exclusive presence of the other template.

Next we tested both assays designed for the transposon sequences, whether these fit to the GFP copy numbers. As formerly mentioned, random integrations can be the result from the breakage of the plasmid at any points, therefore randomly integrated one copy clones are not reliable candidates for reference samples when utilizing transposon sequences. Only after validating the one copy candidates cloned from the active transposition experiments with the GFP assay, were these clones used as reference samples with the transposon specific assays. The results based on GFP and the IRDR-L were in agreement with each other (Figure 10) and most standard curve analysis showed similar result with the comparative Ct method indicating that we could directly use the ΔΔCt method. In addition, technical errors could be further decreased using a pool of gDNA samples with known copy number as a reference. Here we show the calculation with the row Ct values by the example of G2C2 clone in Figure 10 with Cts resulted in IRDR-L assay. Ct mean values are calculated as the mean of Cts of 3 simultaneous PCR reactions:

$Ct_{\text{Mean IRDR-L}}$ of G2C2: 31.065

$Ct_{\text{Mean RNaseP}}$ of G2C2: 30.951

$Ct_{\text{Mean IRDR-L}}$ of one copy clone pool: 31.457

$Ct_{\text{Mean RNaseP}}$ of one copy clone pool: 29.773

Calculations of ΔCt – s:

$\Delta Ct_{\text{IRDR-L}} = 31.065\text{-}31.457 = -0.392$

$\Delta Ct_{\text{RNaseP}} = 30.951\text{-}29.773 = 1.178$

Calculation of ΔΔCt: $\Delta\Delta Ct = -0.392 - 1.178 = -1.57$

Relative quantity: $RQ = 2^{-\Delta\Delta Ct} = 2^{1.57} = 2.969$ (3 copies)

These experiments therefore supported the use of the IRDR-L repeat specific assay for transposon copy number determination as it gave the same results as the assay specific for the carried internal transgene (GFP). However similar application of IRDR-R TaqMan® assay for the previously analyzed clones was unreliable to determine the exact copy number and the assay usually showed lower efficiency than the GFP and the IRDR-L. So initially we concluded that we have to leave out the specific but less efficient assay for the IRDR-R region (Kolacsek *et al*., 2011).

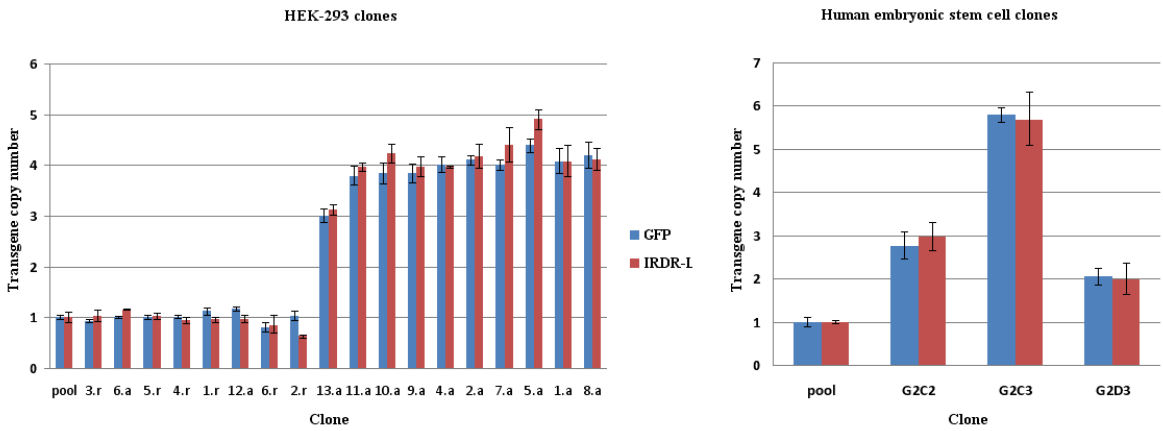

Figure 10: Copy numbers determined by the GFP or the IRDR-L TaqMan® assays in different clones of HEK-293 (left panel) and human embryonic stem cells (right panel). Results of the transgene independent, SB transposon specific IRDR-L assay correlated well with the GFP-based copy numbers. Pool: a mixture of equal amounts of one copy clones. Figure partially adopted from Kolacsek et al., Mobile DNA (2011), 2(1):5, published by BioMed Central.

Concerning SB transposon mutants and variants, "symmetrical" SB transposons with two IRDR-L (but not with two IRDR-R) flanking sequences were shown to be functional (Izsvak *et al.*, 2002). The assay for the left transposon sequence is also applicable for such constructs with a correction factor of 0.5. Searching for such mutant transposon clones, we used the IRDR-R assay to determine the presence of the right terminal repeat sequence. During this presence-absence examination studies, we could achieve higher sensitivity of the IRDR-R TaqMan® with elevating the input range of the gDNA. However, the sensitivity and the reliability of IRDR-R was still below to that of the IRDR-L assay, and only lower copy number clones showed reliable measurements (Figure 11). In addition, we still have to emphasize here that sample quality has a deep impact on reproducibility of all assays.

As a major general recommendation, we routinely analyze more parts of the inserted transgene sequence (e.g. IRDR-L and GFP regions or IRDR-L and IRDR-R regions). It is also advisable using at least two separate single copy clones as reference samples and two few copy (3 – 4) clones as controls in all analyses. The application of more than one reference clone makes it possible to choose the most appropriate one with which low copy control clones give the best precise round copy numbers. In fact, the use of the few (3 – 4) copy clones as references is helpful to approximate extreme copies (> 15) more precisely.

5 Validation of Transgene-independent qPCR Copy Number Quantification

To compare our transgene-independent quantification approach with other techniques, we measured copy numbers of clones that were generated from different cell types by transposons containing various transgene sequences. Such clones were ideal for comparison due to the different transgene sequence and because copy numbers in those cases were also determined either by the Southern/dot blotting techniques,

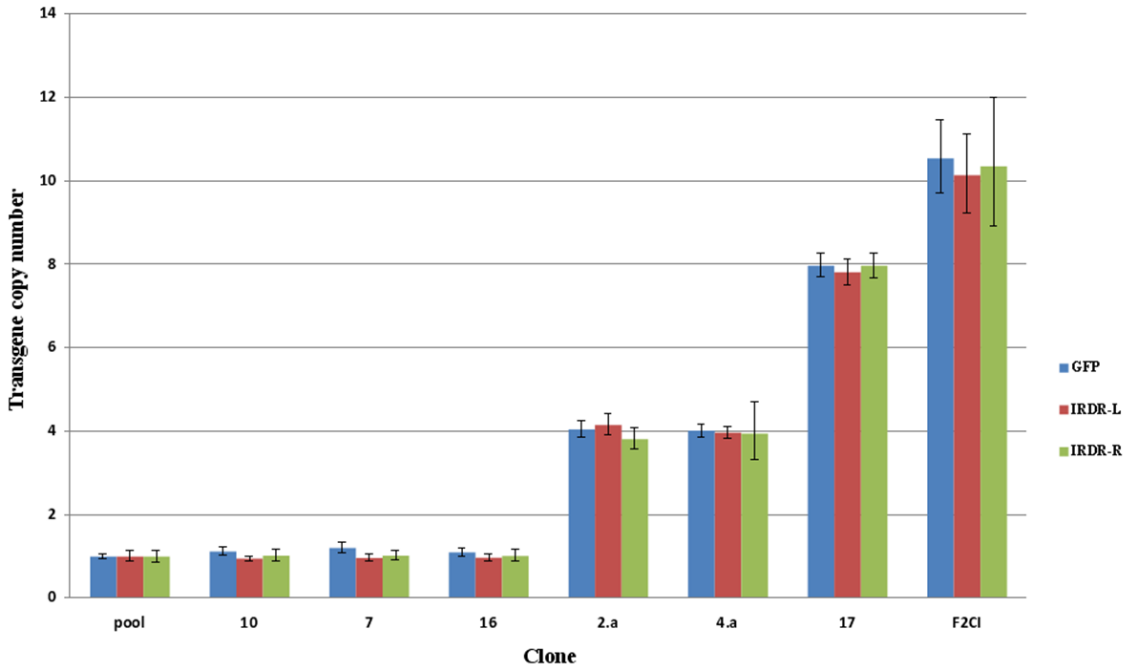

Figure 11: Comparison of transposon copy number determinations based on three different TaqMan® assays. Using higher amount of input gDNA (90ng) resulted in reliable correlation of the IRDR-R assay with the copy numbers determined by the GFP and IRDR-L specific qPCR measurements.

or by the transposon display method, or by estimations from transgene integration assays (Splinkerette PCR/Inverse PCR). Using the IRDR-L assay with the ΔΔCt methodology, copy numbers were estimated ranging from 1 to 50 copies in various clones (Table 1 and 2). The measured transposon copies were almost always the same by the qPCR as by the canonical methods. As canonical methods – with the exception of dot blot – utilize restriction enzyme sites flanking the integration point, which may be sensitive to sequence environment, therefore all integrated copies may not be reliably detected by the canonical methods. Perhaps this is why copy numbers were underestimated in some of the cases comparing to our qPCR method. For higher (> 5) copy-number clones, our method was also accurate, with occasional low relative-error margins (≤ 9%). The slight differences in these cases could be due to the inaccuracy of the standard methods for this range. In addition, it has been suggested that precise values of very high copy numbers are more reliably measured by dot blot rather than transposon display method.

By the above described experiments, the newly developed transgene independent method for determining SB transposon copy numbers could be validated since (i) it provided the same results as the assays specific for the carried transgene sequence and (ii) it could also reliably replace widely used canonical radioactive techniques. The TaqMan® assay designed for the IRDR-L region of the transposon provides the basis for transgene independence as it is present in all SB constructs (Kolacsek *et al.*, 2011).

Clone	Canonical methods	Copy numbers	
		by canonical methods	by qPCR IRDR-L
2/1	Transposon display/Southern blotting	8 – 10	8
2/2	Transposon display/Southern blotting	3	4
2/3	Transposon display/Southern blotting	10 – 12	10
2/9	Transposon display/Southern blotting	1	1
1	Transposon display/Southern blotting	12 – 13	13
4	Dot blot	52	50
5	Transposon display/Southern blotting	15	15
6	Transposon display/Southern blotting	12	11
7	Transposon display/Southern blotting	1	1
8	Transposon display/Southern blotting	2	2
9	Transposon display/Southern blotting	1	1
A3	Splinkerette PCR/Inverse PCR	2	2
A4	Splinkerette PCR/Inverse PCR	4	4
A5	Splinkerette PCR/Inverse PCR	4	4,5
A6	Splinkerette PCR/Inverse PCR	2	2
B1	Splinkerette PCR/Inverse PCR	1	2
B2	Splinkerette PCR/Inverse PCR	2	2
B3	Splinkerette PCR/Inverse PCR	3	3
B5	Splinkerette PCR/Inverse PCR	2	2
C3	Splinkerette PCR/Inverse PCR	1	2
C5	Splinkerette PCR/Inverse PCR	2	2
3	Splinkerette PCR/Inverse PCR	1	1
16	Splinkerette PCR/Inverse PCR	1	1
3.a	Splinkerette PCR/Inverse PCR	1	1
6.a	Splinkerette PCR/Inverse PCR	1	1
12.a	Splinkerette PCR/Inverse PCR	1	1

Table 1: Comparing the IRDR-L qPCR-based method with other canonical techniques for transposon copy number determination. Additional data is added to the table adopted from Kolacsek et al., Mobile DNA (2011), 2(1):5, published by BioMed Central.

Clone	Copy by qPCR	Sequence	Chrom. pos.	Ref.seq.	Nt.pos. of ref.seq.	Region	Seq. orient.	Method
C3	2	TTCCTTGTTACTCCTTCAAAATGCT TAcagttgaagt	**Hs 6p22**	NT_007592.15	51406968	intergenic	**L**eft	splinkerette pGEM
C5	2	CAGTTTTATTATTATTGTCATTGTA TAcagttgaagt	**Hs Xq28**	NT_167198.1	2288320	*GABRA3* intron3	**L**	splinkerette pGEM
		CAGATCCTGATAAATGTTAGTTAC TAcagttgaagt	**Hs 10q21**	NT_030059.13	4645679	*PRKG1* intron7	**L**	inverse PCR pGEM
3	1	TTCCTCCCTGGACTTTAGGACATA TAcagttgaagt	**Hs 5q22**	NT_034772.6	6866220	intergenic	**L**	splinkerette
		AGGAAGTATACACTCACCCTGTGA TAcagttgaagt	Hs 5q22	NT_034772.6	6866220		**R**ight	splinkerette
16	1	GACACAGCTGATTTTGAAGTCAGA TAcagttgaagt	**Hs 4q28**	NT_016354.19	65845752	*SCOC* intron1	**L**	splinkerette
		GAAGACTACCCTATCACTGTATTA TAcagttgaagt	Hs 4q28	NT_016354.19	65845752		**R**	splinkerette
3.a	1	TAGCACAATGAGTACTTTATCACA TAcagttgaagt	**Hs 1p31**	NT_032977.9	82218216	intergenic	**L**	inverse PCR pGEM
6.a	1	AATACATGGAAGGTAGAACAGATC TAcagttgaagt	**Hs 1p31**	NT_032977.9	88638434	intergenic	**L**	splinkerette
		TAAATTATACATGTCTGTTAAAGCA TAcagttgaagt	Hs 1p31	NT_032977.9	88638434		**R**	splinkerette
12.a	1	TCTGGAAAGCCACATTCGGGAACT TAcagttgaagt	**Hs 18q21**	NT_025028.14	21860434	intergenic	**L**	splinkerette

Table 2: Examples of *bona fide* SB transposon integration sites in clones of human cell lines where transposon copy numbers were determined previously. L or R indicates whether the particular transposon integration site was determined from the direction of the left or the right terminal repeat region of SB. The 'TA' sequence (shown in red) next to the transposon terminal repeats provides evidence that these integrations are the results of *bona fide* transposition, as this dinucleotide marks the SB target sequence that is duplicated during the reaction. Hs=Homo sapiens.

6 Quantitative Excision PCR

In our laboratory, we have successfully applied different DNA transposon systems for various in vitro cell culture applications, including studies on directed tissue differentiation from embryonic stem cells and modeling *ex vivo* introduction of therapeutic transgenes into patient cells. Controlling the delivery efficiency is a crucial task for these applications which is usually characterized by testing the transgenic rate of transfectants. However these test methods are time consuming and not applicable in all of the cases, so we have developed a qPCR method to characterize transposition activity. This quantification is extremely useful when optimizing for conditions of a given delivery system, or comparing the efficiency of different transposon systems or variants of a particular transposase enzyme.

There are two major steps involved in transposition, the excision of the transposon from the donor site and the integration of the transposon into the target site. These steps are proven to be coupled since excision frequency of different transposon mutants from a donor plasmid was correlated to overall transposition efficiency (Liu *et al.*, 2004). Excision events can be detected by a PCR reaction with primers flanking the transposon sequence at the donor site (Figure 12). Following excision, the donor plasmid will be circularized by the double-strand DNA break repair mechanism of the cell, and only these excised and repaired plasmids will serve as a template for exponential amplification, because the transposon content is usually large enough not to be amplified from the original uncut sequence. Excision PCR can be carried out on samples taken on the second/third day after transfection, and both isolated plasmids, as well as cell lysates can be used as input material (we generally use isolated plasmids as input.) A semi quantitative version of this excision PCR has been applied by Liu *et al.* (2004) to describe the excision step of SB transposition.

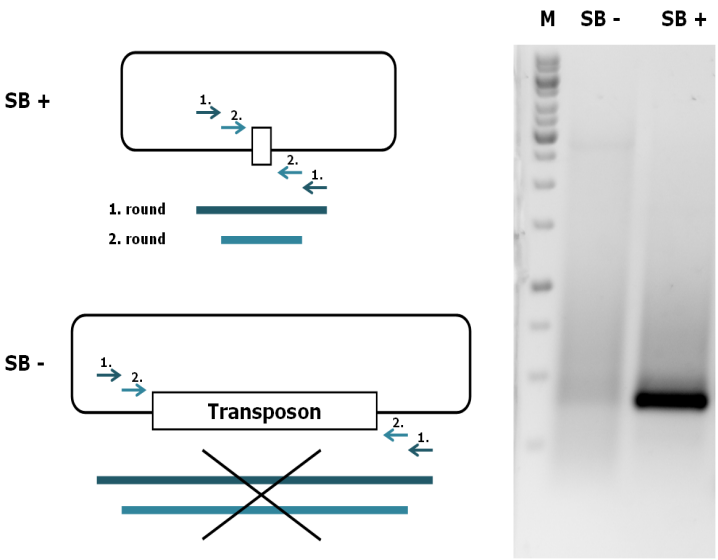

Figure 12: Principle of the excision PCR using transposon containing donor plasmids. This is a nested PCR technique using two sets of primers specific to the transposon flanking sequences, and amplifying the products in two consecutive rounds of PCR. In case of transposition (SB+), there is a distinct product that can be visualized by agarose gel electrophoresis (right panel). In the absence of transposition (SB-), the PCR cannot amplify the target sequence due to its large size. For more details, see the text.

Due to the error-prone double-stranded break repair mechanism, the joining of the donor ends is remarkably imprecise, therefore real-time quantification of excision based on a TaqMan® probe with stringent sequence requirement could not be considered, and hence we adopted the SYBR® Green technology for this application. In some cases, since very small portion of the amount of transposon donor plasmids undergo excision, to increase the traceability, more than one round of nested PCR is necessary to quantify the excision reaction. For this reason, we apply a 10 – 13 cycle pre-amplification (first round of PCR), and after a thousand-fold dilution we measure the real-time round (second round of PCR) with nested primers using SYBR® Green. Amplification from another segment of the plasmid backbone (in our case, the ampicillin resistance gene sequence) could serve as a normalization control. Due to its large excess, ampicillin sequence is not pre-amplified but it is permanently present in our samples during all processes until the second round of excision PCR, so it can be correctly measured in all samples. Although the primers we routinely use are specific to the backbone of our transposon plasmid constructs which can differ among SB users, our primers might be useful for those who apply constructs with the same origin (Table 3). PCR efficiencies for the target and the control sequence were measured by serial dilutions of pooled samples and the ΔΔCt method was proved to be reliable for comparison of excision efficiencies.

	name	sequence 5′ → 3′
1	exc.1 for	GCGAAAGGGGGATGTGCTGCAAGG
2	exc.1 rev	TCTTTCCTGCGTTATCCCCTGATTC
3	exc.2 for	CGATTAAGTTGGGTAACGCCAGGG
4	exc.2 rev	CAGCTGGCACGACAGGTTTCCCG
5	amp. for	TTTGCTCACCCAGAAACGC
6	amp. rev	AGTTGGCCGCAGTGTTATCAC

Table 3: Sequences of primers used for the quantitative excision PCR. Primers with .1 extension are used for the first round, whereas others with .2 extension are used for the second round of the nested PCR; amp primers anneal to the sequence of the ampicillin resistance gene.

An excellent validation of our quantification method developed for measuring the transposition activity was to compare different SB transposase versions resulted from gradual improvement of the activity (Mates *et al.*, 2009). Relative quantification of excision efficiency of SB11x, SB32x and SB100x is correlated to their expected activity in HEK-293 and in HeLa cells as well (Figure 13).

Normalization to the plasmid backbone sequence makes this relative quantification technique independent of the transfection efficiency. This provides the basis for comparing various transposon delivery experiments, including different transposon systems (e.g. SB or PB), different transposase or transposon variants, different host cell types, or different experimental settings. Applying this quantitative technique could therefore be a reliable and fast screening approach for different transposon systems and gene delivery conditions before any applications with the desired transgene.

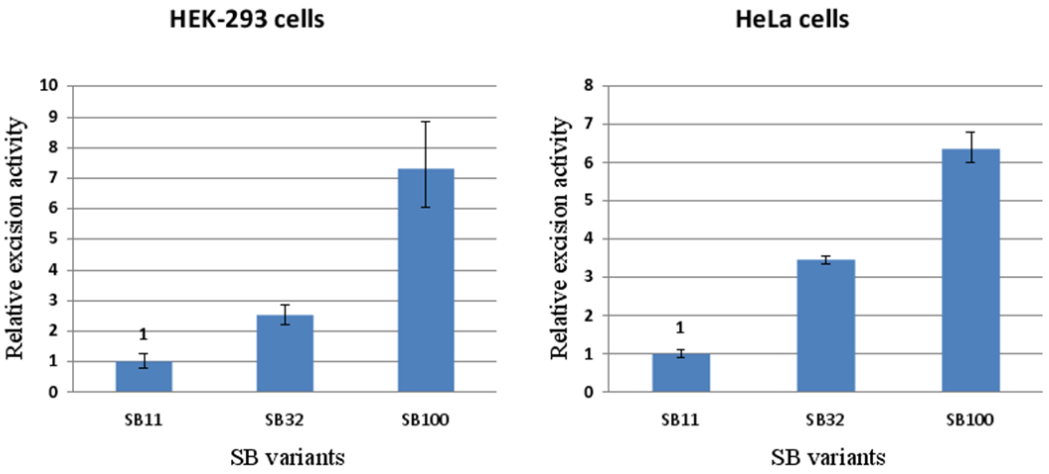

Figure 13: Relative excision efficiencies of SB transposase variants in different cell lines measured by real-time qPCR. Normalizing the excision PCR to the amount of the ampicillin sequence by specific PCR, here we quantify the excision events among all transfected transposon donor plasmids.

7 Related Works

As it was used for quantitative aspects of transposon-based gene delivery experiments, real-time PCR measurements are widely used and have become integral part of the methodology for gene delivery and gene therapy applications. On the other hand, careful design of the experiments or choosing suitable methodology are often missing from such studies, weakening the conclusion of the results. For example, Bian & Belmont (2010) have made absolute quantification of transgenic multi-copy insertions derived from linearized plasmid transfection, although they did not carry out any normalization of the input material. Huang *et al.* (2010) have detected random integration level with absolute quantification normalizing to the RNaseP one copy sequence. They used "empty" gDNA to mix with the target sequence containing plasmid for the standard curve samples; however, normalizing the absolute quantification with external template incorporates additional variability in the measurement. In the meantime, well-designed studies provide excellent new examples of combining existing technologies. Charrier *et al.* (2011) have applied an elegant solution for absolute quantification to determine lentiviral vector copy numbers with normalization to endogenous albumin. Standard curves were made from the same plasmid containing both the target and the endogenous control sequence, resulting in the smallest possible variability in their measurement. Ballester *et al.* (2004) determined copy number of transgenic mice carrying goat lactoglobulin gene using relative quantification with the comparative Ct method. The unique design of the measurement was the application of goat gDNA as a reference sample and the choice of glucagon sequence as an endogenous control, the latter one being strongly conserved between the two species.

Concerning SB transposon applications, a recent paper described a simultaneous analysis of excision activity of *Sleeping Beauty* and the resulting transgene copy number (Jin *et al.*, 2011). Excision was analyzed at days 1 to 3 after electroporation using the ΔΔCt method with a TaqMan® assay where the probe was specific to one of the transposon flanking regions. They also normalized to a plasmid backbone sequence, thus similarly to our qPCR studies, their quantification is independent of the electrotrans-

fer efficiency. However, as the repair of the plasmid after excision is an error-prone process (see Section 7), using a strict TaqMan® probe sequence may not detect all excision products. They also determined the average copy number in transgenic cell population after four weeks of selection. They used an absolute quantification approach using dilution series of a one copy clone gDNA as a standard which is in fact an alternative to relative standard curve method using one copy as a reference.

8 Conclusions

Transposon-based technology is an emerging new method of choice for gene delivery and for gene therapy applications. Compared to viral vectors, transposon systems offer several advantages. Apart from being less expensive in terms of the required safety facilities, for several DNA transposons, the integration profile of the delivered transgene is more close to random, showing no preferences for coding regions, therefore making its application less susceptible for insertional mutagenesis. SB seems to be the safest delivery technique for two reasons: 1) its transgene integration profile is the closest to random among all known gene delivery vehicles; 2) SB was resurrected from an ancient non-functional fish mobile element, therefore no potential transposons are present in vertebrate species, including the human genome, that can remobilize the integrated transgene. Based on these characteristics, the SB system is the most suitable for cell and gene therapy applications, even if compared to other transposon vehicles.

Our work contributed to the quantification of transposon delivery in two important aspects. We have developed sensitive and reliable real-time PCR-based methods to measure (i) the first step of the transposition and (ii) the resulting copy number of the delivery. Quantitative excision PCR is extremely useful to control and to optimize transposon mediated gene delivery, whereas copy number determination is essential to characterize transgenic cells. Comparing our copy number method with widely used canonical methods, it was proved to be just as accurate as those, also offering a faster and non-radioactive approach at the same time. However, the real advantage of this method is the transgene-independence which makes it applicable for any scientists working with SB transposon constructs.

In this chapter, we gave a detailed protocol for designing quantitative measurements of integrated DNA sequences such as transposons, and show examples for developing reliable quantitative assays specific to any sequence. In general, however, this description may also serve as a stepwise guide providing a strategy for similar quantification purposes.

Acknowledgements

Tamás I. Orbán is a recipient of the János Bolyai Scholarship of the Hungarian Academy of Sciences. Research in our laboratory was supported by grants from OTKA (NK83533), STEMKILL (OM00108/2008), and KMOP-1.1.2-07/1-2008-0003.

References

Balciunas, D., Wangensteen, K. J., Wilber, A., Bell, J., Geurts, A., Sivasubbu, S., Wang, X., Hackett, P. B., Largaespada, D. A., McIvor, R. S., & Ekker, S. C. (2006). Harnessing a high cargo-capacity transposon for genetic applications in vertebrates. PLoS Genetics, 2(11), e169.

Ballester, M., Castello, A., Ibanez, E., Sanchez, A., & Folch, J. M. (2004). Real-time quantitative PCR-based system for determining transgene copy number in transgenic animals. Biotechniques, 37(4), 610-613.

Baus, J., Liu, L., Heggestad, A. D., Sanz, S., & Fletcher, B. S. (2005). Hyperactive transposase mutants of the Sleeping Beauty transposon. Molecular Therapy, 12(6), 1148-1156.

Bian, Q. & Belmont, A. S. (2010). BAC TG-EMBED: one-step method for high-level, copynumber-dependent, position-independent transgene expression. Nucleic Acids Research, 38(11), e127.

Biemont, C. & Vieira, C. (2006). Genetics: junk DNA as an evolutionary force. Nature, 443(7111), 521-524.

Burns, K. H. & Boeke, J. D. (2012). Human transposon tectonics. Cell, 149(4), 740-752.

Charrier, S., Ferrand, M., Zerbato, M., Pre´cigout, G., Viornery, A., Bucher-Laurent, S., Benkhelifa-Ziyyat, S., Merten, O. W., Perea, J., & Galy, A. (2011). Quantification of lentiviral vector copy numbers in individual hematopoietic colony-forming cells shows vector dose-dependent effects on the frequency and level of transduction. Gene Therapy, 18(5), 479-487.

Chelly, J., Kaplan, J. C., Maire, P., Gautron, S., & Kahn, A. (1988). Transcription of the dystrophin gene in human muscle and non-muscle tissue. Nature, 333(6176), 858-860.

Chew, S. K., Rad, R., Futreal, P. A., Bradley, A., & Liu, P. (2011). Genetic screens using the piggyBac transposon. Methods, 53(4), 366-371.

Collier, L. S. & Largaespada, D. A. (2005). Hopping around the tumor genome: transposons for cancer gene discovery. Cancer Research, 65(21), 9607-9610.

Collier, L. S. & Largaespada, D. A. (2007). Transposable elements and the dynamic somatic genome. Genome Biology, 8 Suppl 1, S5.

Devon, R. S., Porteous, D. J., & Brookes, A. J. (1995). Splinkerettes—improved vectorettes for greater efficiency in PCR walking. Nucleic Acids Research, 23(9), 1644-1645.

Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., & Xu, T. (2005). Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. Cell, 122(3), 473-483.

Ellis, J. (2005). Silencing and variegation of gammaretrovirus and lentivirus vectors. Human Gene Therapy, 16(11), 1241-1246.

Feschotte, C. & Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. Annual Reviews of Genetics, 41, 331-368.

Garrison, B. S., Yant, S. R., Mikkelsen, J. G., & Kay, M. A. (2007). Postintegrative gene silencing within the Sleeping Beauty transposition system. Molecular and Cellular Biology, 27(24), 8824-8833.

Goodier, J. L. & Kazazian, H. H., Jr. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. Cell, 135(1), 23-35.

Grabundzija, I., Irgang, M., Mates, L., Belay, E., Matrai, J., Gogol-Doring, A., Kawakami, K., Chen, W., Ruiz, P., Chuah, M. K., VandenDriessche, T., Izsvak, Z., & Ivics, Z. (2010). Comparative analysis of transposable element vector systems in human cells. Molecular Therapy, 18(6), 1200-1209.

Guo, G., Huang, Y., Humphreys, P., Wang, X., & Smith, A. (2011). A PiggyBac-based recessive screening method to identify pluripotency regulators. PLoS One, 6(4), e18189.

Hedges, D. J. & Batzer, M. A. (2005). From the margins of the genome: mobile elements shape primate evolution. Bioessays, 27(8), 785-794.

Heid, C. A., Stevens, J., Livak, K. J., & Williams, P. M. (1996). Real time quantitative PCR. Genome Research, 6(10), 986-994.

Higuchi, R., Dollinger, G., Walsh, P. S., & Griffith, R. (1992). Simultaneous amplification and detection of specific DNA sequences. Biotechnology, 10(4), 413-417.

Huang, X., Haley, K., Wong, M,, Guo, H., Lu, C., Wilber, A., & Zhou, X. (2010). Unexpectedly high copy number of random integration but low frequency of persistent expression of the Sleeping Beauty transposase after trans delivery in primary human T cells. Human Gene Therapy, 21(11), 1577-1590.

Ivics, Z., Hackett, P. B., Plasterk, R. H., & Izsvak, Z. (1997). Molcular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell, 91(4), 501-510.

Ivics, Z. & Izsvak, Z. (2004). Transposable elements for transgenesis and insertional mutagenesis in vertebrates: a contemporary review of experimental strategies. Methods in Molecular Biology, 260, 255-276.

Izsvak, Z., Ivics, Z., & Plasterk, R. H. (2000). Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. Journal of Molecular Biology, 302(1), 93-102.

Izsvak, Z., Khare, D., Behlke, J., Heinemann, U., Plasterk, R. H., & Ivics, Z. (2002). Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in Sleeping Beauty transposition. The Journal of Biological Chemistry, 277(37), 34581-34588.

Izsvak, Z., Hackett, P. B., Cooper, L. J., & Ivics, Z. (2010). Translating Sleeping Beauty transposition into cellular therapies: victories and challenges. Bioessays, 32(9), 756-767.

Jin, Z., Maiti, S., Huls, H., Singh, H., Olivares, S., Mates, L., Izsvak, Z., Ivics, Z., Lee, D. A., Champlin, R. E., & Cooper, L. J. N. (2011). The hyperactive Sleeping Beauty transposase SB100X improves the genetic modification of T cells to express a chimeric antigen receptor. Gene Therapy, 18(9), 849-856.

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., & Woltjen K. (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature, 458(7239), 771-775.

Kazazian, H. H., Jr. (2004). Mobile elements: drivers of genome evolution. Science, 303(5664), 1626-1632.

Kleppe, K., Ohstuka, E., Kleppe, R., Molineux. L., & Khorana, H. G. (1971). Studies on polynucleotides. XCVI. Repair replications of short synthetic DNA's as catalyzed by DNA polymerases. Journal of Molecular Biology, 56(2), 341-361.

Kolacsek, O., Krízsik, V., Schamberger, A., Erdei, Z., Apáti, A., Várady, G., Mátés, L., Izsvák, Z., Ivics, Z., Sarkadi, B., & Orbán, T. I. (2011). Reliable transgene-independent method for determining Sleeping Beauty transposon copy numbers. Mobile DNA, 2(1), 5.

Liu, G., Aronovich, E. L., Cui, Z., Whitley, C. B., & Hackett, P. B. (2004). Excision of Sleeping Beauty transposons: parameters and applications to gene therapy. The Journal of Gene Medicine, 6(5), 574-583.

Liu, G., Geurts, A. M., Yae, K., Srinivasan, A. R., Fahrenkrug, S. C., Largaespada, D. A., Takeda, J., Horie, K., Olson, W. K., & Hackett, P. B. (2005). Target-site preferences of Sleeping Beauty transposons. Journal of Molecular Biology, 346(1), 161-173.

Livak, K. J, Flood, S. J., Marmaro, J., Giusti, W., & Deetz, K. (1995). Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. Genome Research, 4(6), 357-362.

Livak, K. J. (1997). ABI Prism 7700 Sequence Detection System, User Bulletin #2. Applied Biosystems.

Mates, L., Chuah, M. K., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D. P., Schmitt, A., Becker, K., Matrai, J., Ma, L., Samara-Kuko, E., Gysemans, C., Pryputniewicz, D., Miskey, C., Fletcher, B., Vandendriessche, T., Ivics, Z., & Izsvak, Z. (2009). Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nature Genetics, 41(6), 753-761.

Meilinger, D., Fellinger, K., Bultmann, S., Rothbauer, U., Bonapace, I. M., Klinkert, W. E., Spada, F. & Leonhardt, H. (2009). Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. EMBO Reports, 10(11), 1259-1264.

Mills, R. E., Bennett, E. A., Iskow, R. C., Luttig, C. T., Tsui, C., Pittard, W. S., & Devine, S. E. (2006). Recently mobilized transposons in the human and chimpanzee genomes. The American Journal of Human Genetics, 78(4), 671-679.

Mills, R. E., Bennett, E. A., Iskow, R. C., & Devine, S. E. (2007). Which transposable elements are active in the human genome? Trends in Genetics, 23(4), 183-191.

Miskey, C., Izsvak, Z., Plasterk, R. H., & Ivics, Z. (2003). The Frog Prince: a reconstructed transposon from Rana pipiens with high transpositional activity in vertebrate cells. Nucleic Acids Research, 31(23), 6873-6881.

Moeller, F., Nielsen, F. C., & Nielsen, L. B. (2003). New tools for quantifying and visualizing adoptively transferred cells in recipient mice. Journal of Immunological Methods, 282(1-2), 73-82.

Morrison, T. B., Weis, J. J., & Wittwer, C. T. (1998) Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. Biotechniques, 24(6), 954-962.

Mullis, K. & Faloona, F. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods in Enzymology, 155, 335-350.

Ostertag, E. M., Madison, B. B., & Kano, H. (2007). Mutagenesis in rodents using the L1 retrotransposon. Genome Biology, 8 Suppl 1, S16.

Pledger, D. W. & Coates, C. J. (2005). Mutant Mos1 mariner transposons are hyperactive in Aedes aegypti. Insect Biochemistry and Molecular Biology, 35(10), 1199-1207.

Ririe, K. M., Rasmussen, R. P., & Wittwer, C. T. (1997) Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. Analytical Biochemistry, 245(2), 154-160.

Rowe, H. M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V., Layard-Liesching, H., Verp, S., Marquis, J., Spitz, F., Constam, D. B., & Trono, D. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature, 463(7278), 237-240.

Rubin, G. M. & Spradling, A. C. (1982). Genetic transformation of Drosophila with transposable element vectors. Science, 218(4570), 348-53.

Shen, S., Lin, L., Cai, J. J., Jiang, P., Kenkel, E. J., Stroik, M. R., Sato, S., Davidson, B. L., & Xing, Y. (2011). Widespread establishment and regulatory impact of Alu exons in human genes. Proceedings of the National Academy of Sciences of the United States of America, 108(7), 2837-42.

Singer, T., McConnell, M. J., Marchetto, M. C., Coufal, N. G., & Gage, F. H. (2010) LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? Trends in Neurosciences, 33(8), 345-54.

Sivalingam, J., Krishnan, S., Ng, W. H., Lee, S. S., Phan, T. T., & Kon, O. L. (2010). Biosafety assessment of site-directed transgene integration in human umbilical cord-lining cells. Molecular Therapy, 18(7), 1346-1356.

Solyom, S. & Kazazian, H. H. Jr. (2012) Mobile elements in the human genome: implications for disease. Genome Medicine, 4(2), 12.

Spradling, A. C. & Rubin, G. M. (1982). Transposition of cloned P elements into Drosophila germ line chromosomes. Science, 218(4570), 341-7.

Uren, A. G., Kool, J., Berns, A., & van Lohuizen, M. (2005). Retroviral insertional mutagenesis: past, present and future. Oncogene, 24(52), 7656-7672.

Vigdal, T. J., Kaufman, C. D., Izsvak, Z., Voytas, D. F., & Ivics, Z. (2002). Common physical properties of DNA affecting target site selection of sleeping beauty and other Tc1/mariner transposable elements. Journal of Molecular Biology, 323(3), 441-452.

Wang, A. M., Doyle, M. V., & Mark, D. F. (1989). Quantitation of mRNA by the polymerase chain reaction. Proceedings of the National Academy of Sciences of the United States of America, 86(24), 9717-9721.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics, 8(12), 973-982.

Wicks, S. R., de Vries, C. J., van Luenen, H. G., & Plasterk, R. H. (2000). CHE-3, a cytosolic dynein heavy chain, is required for sensory cilia structure and function in Caenorhabditis elegans. Developmental Biology, 221(2), 295-307.

Williams, D. A. (2008). Sleeping beauty vector system moves toward human trials in the United States. Molecular Therapy, 16(9), 1515-1516.

Wilson, M. H., Coates, C. J., & George, A. L., Jr. (2007). PiggyBac transposon mediated gene transfer in human cells. Molecular Therapy, 15(1), 139-145.

Wittwer, C. T., Herrmann, M. G., Moss, A. A., & Rasmussen, R. P. (1997). Continuous fluorescence monitoring of rapid cycle DNA amplification. Biotechniques, 22(1), 130-138.

Woltjen, K., Michael, I. P., Mohseni, P., Desai, R., Mileikovsky, M., Hämäläinen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., Kaji, K., Sung, H. K., & Nagy, A. (2009) piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature, 458(7239), 766-70.

Yant, S. R., Wu, X., Huang, Y., Garrison, B., Burgess, S. M., & Kay, M. A. (2005). High-resolution genome-wide mapping of transposon integration in mammals. Molecular and Cellular Biology, 25(6), 2085-2094.

Zayed, H., Izsvak, Z., Walisko, O., & Ivics, Z. (2004). Development of hyperactive sleeping beauty transposon vectors by mutational analysis. Molecular Therapy, 9(2), 292-304.

Zhu, J., Park, C. W., Sjeklocha, L., Kren, B. T., & Steer, C. J. (2010). High-level genomic integration, epigenetic changes, and expression of sleeping beauty transgene. Biochemistry, 49(7), 1507-1521.