Gene 834 (2022) 146609

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Functional indications for transposase domestications – Characterization of the human *piggyBac* transposase derived (PGBD) activities

Orsolya Kolacsek [a], Gerda Wachtl [a,b], Ábel Fóthi [a], Anita Schamberger [a], Sára Sándor [a], Enikő Pergel [a], Nóra Varga [a], Tamás Raskó [c], Zsuzsanna Izsvák [c], Ágota Apáti [a], Tamás I. Orbán [a,*]

[a] Institute of Enzymology, Research Centre for Natural Sciences, Budapest, Hungary
[b] Doctoral School of Biology, Institute of Biology, ELTE Eötvös Loránd University, Budapest, Hungary
[c] Max Delbrück Center for Molecular Medicine in the Helmholtz Society, Berlin, Germany

ARTICLE INFO

Edited by Andre van Wijnen

Keywords:
PGBD
PGBD5
MER75
MER85
Transposon
Cellular stress

ABSTRACT

Transposable elements are widespread in all living organisms. In addition to self-reproduction, they are a major source of genetic variation that drives genome evolution but our knowledge of the functions of human genes derived from transposases is limited. There are examples of transposon-derived, domesticated human genes that lost (SETMAR) or retained (THAP9) their transposase activity, however, several remnants in the human genome have not been thoroughly investigated yet. These include the five human *piggyBac*-derived sequences (PGBD1-5) which share ancestry with the *Trichoplusia ni* originated *piggyBac* (PB) transposase. Since PB is widely used in gene delivery applications, the potential activities of endogenous PGBDs are important to address. However, previous data is controversial, especially with the claimed transposition activity of PGBD5, it awaits further investigations. Here, we aimed to systematically analyze all five human PGBD proteins from several aspects, including phylogenetic conservation, potential transposase activity, expression pattern and their regulation in different stress conditions. Among PGBDs, PGBD5 is under the highest purifying selection, and exhibits the most cell type specific expression pattern. In a two-component vector system, none of the human PGBDs could mobilize either the insect PB transposon or the endogenous human PB-like MER75 and MER85 elements with intact terminal sequences. When cells were exposed to various stress conditions, including hypoxia, oxidative or UV stress, the expression profiles of all PGBDs showed different, often cell type specific responses; however, the pattern of PGBD5 in most cases had the opposite tendency than that of the other *piggyBac*-derived elements. Taken together, our results indicate that human PGBD elements did not retain their mobilizing activity, but their cell type specific, and cellular stress related expression profiles point toward distinct domesticated functions that require further characterization.

1. Introduction

Transposable elements (TEs) are discrete DNA segments that can move into new chromosomal locations either by a "cut and paste" or by a "copy and paste" mechanism. They fall into two classes according to whether their transposition intermediate is RNA (Class I or retrotransposons) or DNA (Class II or DNA transposons). TEs are ubiquitous in living organisms and several remnants of them occupy large portions of genomes (Kidwell, 2002). While they are generally considered as selfish sequences being able to cause harmful mutations, they are also important evolutionary factors (Cosby et al., 2019). Active LINE-1 retrotransposon copies continue to produce genetic diversity in human populations (Kazazian, 1999), and have a potential to contribute to individual somatic mosaicism (Coufal et al., 2009; Singer et al., 2010). One of the most direct contributions of TEs to host genome evolution is the emergence of new genes, also known as "molecular domestication" or "exaptation," where TE sequences acquire new functions. This process can be illustrated by the RAG1 recombinase (Kapitonov and Jurka,

Abbreviations: TE, Transposable Element; TIR, Terminal Inverted Repeat; TSD, Target Site Duplication; PB, *PiggyBac*; PGBD, *PiggyBac*-Derived; SB, *Sleeping Beauty*; MITE, Miniature inverted repeat transposable elements; NPC, neural progenitor cell.

* Corresponding author at: Institute of Enzymology, Research Centre for Natural Sciences, Eötvös Loránd Research Network, 1117 Budapest, Magyar Tudósok körútja 2, Hungary.
E-mail addresses: orban.tamas@ttk.hu, orbant@biomembrane.hu (T.I. Orbán).

https://doi.org/10.1016/j.gene.2022.146609
Received 15 October 2021; Received in revised form 17 May 2022; Accepted 18 May 2022
Available online 21 May 2022

2005) or the SET histone methyltransferase (Cordaux et al., 2006), showing domestication examples of a Transib or a Mariner transposon, respectively. Though domestication is usually associated with the loss of their mobilizing capacities, there are also examples of genes retaining their transposase activity, such as the human THAP9, which is able to transpose the distantly related *Drosophila P-element* (Majumdar et al., 2013),

PiggyBac (PB) DNA transposon elements, first characterized in the cabbage looper moth *Trichoplusia ni* (Fraser et al., 1983), have then been identified in a variety of eukaryotes from protozoa to primates (Lander et al., 2001; Pritham et al., 2005; Bouallegue et al., 2017). A PB element generally contains a 1.8 kb open reading frame (ORF) encoding a 68 kDa transposase; the ends of the transposon sequence consist of 13–15 nt terminal inverted repeats (TIRs), which are flanked by the duplication of the target site (TSD) sequence 'TTAA'. Among DNA transposons, the PB superfamily is peculiar because precise excision of the element restores the pre-integration site (Mitra et al., 2008), making transposition mechanism less genotoxic than in case of other families, where restoration of the excision site is an error prone process (Chen et al., 2020). There are active PB copies even in higher phylogenetic orders like mammals (Mitra et al., 2013), and domestication can also be observed for functions such as DNA binding and excision, or recombination of host DNA in ciliates (Baudry et al., 2009; Cheng et al., 2010) and in *Xenopus* (Hikosaka et al., 2007). The PB of the *T. ni* has been developed as an effective nonviral gene transfer tool (Ding et al., 2005; Yusa et al., 2011), with the potential for human gene therapy applications (Manuri et al., 2010). However, these applications raise concerns about the presence of ~2000 PB-like elements in the human genome (Mandal and Kazazian, 2008), especially for PGBD5 (*piggyBac* derived 5) which was suggested to have transposase activity for *piggyBac* TIR sequences (Henssen et al., 2015) and recently for distantly related *piggyBac*-like elements (Helou et al., 2021b), although these findings have later been challenged (Beckermann et al., 2021). Cross-reactivities of these distant PB-related elements enhance the possibility of genomic instability during transgenesis performed by the *T. ni* PB transposon system (Ivics, 2016).

The five PB derived elements in the human genome (PGBD1-5) are variously conserved among vertebrates; PGBD5 dates back to the cephalochordate/vertebrate split, PGBD1 and PGBD2 are found in mammals, whereas PGBD3 and PGBD4 are present predominantly in primates (Lander et al., 2001; Sarkar et al., 2003; Pavelitz et al., 2013; Bouallegue et al., 2017). Miniature inverted repeat transposable elements (MITEs) are internally deleted versions of invading transposon copies which could be spread effectively as nonautonomous variants throughout the genome (Hartl et al., 1992). They were transposed *in trans* by master copies of the complete autonomous transposon being able to express a transposase. MITEs can be illustrated by their consensus sequence: MER85 and MER75 have been identified belonging to PGBD3 and to PGBD4 respectively, because they TIR sequences have been preserved and is recognizably identical (Lander et al., 2001). In total, there are 1587 copies of them in the human genome, one-fifth of which have intact TIR ends and TSDs (Henssen et al., 2015). PGBD3, PGBD4 and their MITEs are the most recently amplified DNA transposon families in the human genome as they invaded the primate lineage. For the other PGBDs, the corresponding recognition TIR sequences can no longer be identified (Sarkar et al., 2003).

Genetic structures of PGBDs are very different. PGBD1, PGBD2 and PGBD3 are transcribed as extra exons within host genes. The transposase-derived sequences usually present as one single exon except for *PGBD5* which has six introns in vertebrates. PGBD1 members are the result of an ancestral fusion between exons containing SCAN domains (leucine-rich regions) in the N terminal regions and the last exon encoding for the transposase (Sarkar et al., 2003; Raskó et al., 2021); the human PGBD1 fusion product was also called as HUCEP-4 earlier. PGBD2 mRNA is transcribed with three exons, and its translation starts form the last one. PGBD3 is inserted into the fifth intron of the *Cockayne Syndrome group B* gene (CSB), which is a chromatin remodeling factor (Citterio et al., 2000; Newman et al., 2006) and also has a role in transcription coupled nucleotide excision repair (Fousteri and Mullenders, 2008; Gray and Weiner, 2010). Unlike other PGBDs, the PGBD3 transposon potentially functions as a natural exon trap and contains a splicing acceptor site in the 5′ region and a polyadenylation signal in the 3′ region. Thus, an alternative splicing of this region leads to a regular CSB product or to the CSB-PGBD3 fusion protein, but also to an extra transcript starting from the fifth exon and producing a single PGBD3 protein (Newman et al., 2008). Indeed, closest relatives of PGBD3 from *Acyrthosiphum pisum* and *Stegodyphus mimosarum* also share exon trap features (Bouallegue et al., 2017). Moreover, the authors suggest that these RNA processing signals were present in the common ancestor of PGBD1, PGBD2 and PGBD3, and were then lost along the branches leading to PGBD1 and PGBD2. This might be the reason for utilizing host promoter for the expression of all of these elements, which is also a characteristic of PGBD4 but in a different manner. PGBD4 has been inserted near the ORF of *Endoplasmic reticulum Membrane Complex 7* (EMC7), and is expressed via the bidirectional activity of the shared promoter (https://www.ncbi.nlm.nih.gov/gene/161779).

Recently, the potential activity of PGBD5, the most ancestral member of the human PGBD genes has become a very controversial issue. Originally, its transposase activity was considered to be very unlikely (Pavelitz et al., 2013; Saha et al., 2015), however, later papers claimed a transposase (Henssen et al., 2015; Helou et al., 2021b), or rather moderately, a genome "slicing" activity of PGBD5 (Henssen et al., 2016; Henssen et al., 2017). On the other hand, a recent systematic study provided evidence that PGBD5 cannot perform canonical DNA transposition (Beckermann et al., 2021). These results clearly awaited further systematic examinations of all human PGBD variants, and the importance to investigate their potential reactivity either with the MITEs of MER75/MER85, or with the insect derived PB vector system. In the current study, we have characterized all the five PGBDs present in the human genome from various aspects. We have analyzed phylogenies and conservation levels among the primates and investigated their expression profiles in different cell lines, together with some co-regulated transcripts produced from the same loci. Potential transposase activities have been tested with the two-component transposon system, providing evidence for the lack of *bona fide* transposition ability. In addition, we have examined the cross-reactivities between the insect PB transposase and MER75/MER85 elements of the human genome, but these were found to be non-functional. Lastly, to get closer to exploring the functions of *PGBD1-5*, we have tested the regulation of expression levels under different stress conditions. Our results support the hypothesis of human PGBDs being domesticated with a simultaneous loss of their transposase activity, and the use of the insect-derived PB system for potential gene therapy applications.

2. Material and methods

2.1. Phylogeny and conservation analysis by Ka/Ks

Human PGBD1-*5* protein sequences (NCBI Ref Seqs: NP_001171672.1, NP_733843.1, NP_736609.2, NP_689808.2, NP_001245240.1) were analyzed in Pfam 33.1 protein families database (https://pfam.xfam.org/) (El-Gebali et al., 2019) to search for the transposase domain DDE_Tnp_1_7 (PF13843, transposase IS4). The defined sequences and the narrow primate homologous flanking region (+0/1 amino acids (aa) at start and +4 aa at end) were used as basis (nucleotide sequences are shown in Suppl. Table 1). Amino acid sequences were determined by TranslatorX (Abascal et al., 2010), protein alignments were done by Clustal Omega program (Sievers et al., 2011). Phylogenetic tree was visualized by Archaeopteryx (Han and Zmasek, 2009). Nonsynonymous and synonymous substitution rates (denoted as *Ka* and *Ks*, respectively) were estimated using the software KaKs_Calculator (Zhang et al., 2006), with the method of model averaging. *Ka/Ks*

calculations are detailed in Suppl. Table 2. Significance levels (p-values) of differences from the *Ka/Ks* statistics were calculated by the two tailed *t*-test.

2.2. Cell lines, culturing and toxicity assays

Human embryonic kidney cells (HEK-293) and HeLa cells were maintained as described previously (Kolacsek et al., 2014); the HUES9 embryonic stem cell line was cultured as described earlier (Apáti et al., 2008). The establishment and maintenance of the MSCL-2 mesenchymal-like cell line was described in detail previously (Varga et al., 2011). Maintenance of 62F iPSC cells and differentiation into hippocampal neural progenitors (NPCs) were performed as described earlier (Vofely et al., 2018).

For toxicity assays, 6×10^4 of HeLa or 8×10^4 of HEK-293 cells were seeded onto 24-well plates and transfected with the FuGENE® 6 reagent, according to the manufacturer's instructions (Roche Applied Science). For the assays, 150 ng of pPB-CAG-GFP transposon donor plasmid (Kolacsek et al., 2014) and 150 ng of transposase (mPB or PGBD1-5) expression plasmid were co-transfected into the cells. For controls, mock transfections (transfection reagents without DNA) or using the pET41 control plasmid (Merck) instead of the transposase expression plasmid were applied in the assays. At 48 h posttransfection, cells were collected by trypsinization and after the addition of propidium iodide (PI, in the final concentration of 2 μg/ml), cell viability was measured by the percentage of PI-negative cells using the Attune NxT flow cytometer (Thermo Fisher Scientific). The use of a GFP-expressing donor plasmid allowed to determine not only the transfection efficiency, but also to determine and compare the percentage of viability/toxicity among the transfected and the non-transfected cells. All transfections were carried out in duplicates.

2.3. RT-qPCR for gene expression analyses

Total RNA was isolated from cells in near confluent 6-well plates using RNeasy Plus Mini kit (QIAGEN). Isolated RNA was digested with DNaseI (New England BioLabs) according to manufacturer instructions, and 1 μg of DNA-free total RNA was reverse transcribed in the volume of 10 μl using random oligonucleotides with the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific). 5 μl ten-fold diluted cDNAs were used in qPCRs in 20 μl reaction volume, containing 50 nM of primers and Power SYBR® Green PCR Master Mix (Thermo Fisher Scientific). Reactions were performed in triplicates on StepOnePlus™ real-time PCR platform (Thermo Fisher Scientific) with the following thermal profile: 95 °C 10 min, 40 cycles of 95 °C 15 s and 60 °C 1 min, and melting curve profile was added. With the exception of PGBD4 all the qPCR primer pairs were designed for intron spanning with the use of Universal Probe Library Assay Design Center (Roche Applied Science) and listed in the Suppl. Table 3.

Expression levels in different tissue cell lines were determined with the Ct shift method described previously (Kolacsek et al., 2017). Briefly, we have constructed plasmid standards for each target as described, containing PCR amplicon sequences of the target and the Hsp90 endogenous control. Plasmid standards were used as reference samples containing equal copies of the target and the endogenous control. Plasmid standard and cDNA samples were measured in qPCRs using the ΔΔCt formula, determining the exact ratio of the target and the endogenous control. Different targets are comparable by this method as they are measured in the same endogenous control expression units plotted on the y-axis.

RNA variants from the CSB locus were detected by three amplicons (Fig. 1D, left panel). One amplifies the whole CSB-PGBD3 fusion mRNA splice variant expressed from the canonical promoter (tr1svPG3), another amplifies the whole CSB mRNA splice variant also expressed from the canonical promoter (tr1svCSB). The third amplicon detects the PGBD3 exon splice variant, but measures the sum of the fusion and the sole PGBD3 mRNA (tr1 + 2svPG3), the latter being expressed from a cryptic promoter located in exon 5 (tr2svPG3). The single PGBD3 can be calculated by subtracting the appropriate PGBD3 splice variant (tr1svPG3) from the sum of the transcripts (tr1 + 2svPG3). This mRNA gene product is denoted as sPGBD3 and produces a single PGBD3 protein.

Gene expression quantification after stress treatments (see Section 2.6.) were carried out by the standard ΔΔCt formula using the untreated control cells as reference sample and $^{-}$ΔΔCt was plotted. Quantification of p21 and VEGFA were used as positive controls for controlling the efficiency of treatments. In hypoxia and UV stress, analysis were normalized to Hsp90 and HMGB1 as multiple endogenous controls. For oxidative stress condition, HMGB1 and B2M endogenous controls were used for normalization. These endogenous controls were selected from a panel based on their stable expression levels (Hsp90, HMGB1, B2M, PolR2A, and RPLP0).

2.4. Plasmid constructs

All PCR primer pairs used for construction of plasmids are listed in Suppl. Table 3. Structures of the transposon donor plasmids are shown in Fig. 2A. Donor plasmid sequences are exactly the same except TIR sequences and some untranslated regions of the original transposon. Transposon donors contain a phosphoglycerate kinase (PGK) promoter regulated puromycin antibiotic resistance expression cassette. Candidates of MER75B and MER85 sequences with intact TIRs and TSDs were PCR amplified from gDNA with a short flanking region from 4 and 21 chromosomes, respectively. PCR products were inserted into pGEM®-T vector (Promega) and verified by Sanger sequencing. MER candidates were ligated into the excision backbone identical with pPB-puro donor plasmid. HindIII restriction site naturally present in the MER85 candidate sequence was used to insert the PGK-puromycin cassette. For this purpose, HindIII site was inserted into the middle region of MER75B by site directed PCR mutagenesis. As the CSB-PGBD3 protein was shown to bind the palindromic inner sequence previously (Gray et al., 2012), this palindromic sequence was reconstituted with PCR mutagenesis in the cloned MER85 candidate. Maps and sequences of the puromycin resistance cassette expressing plasmids are shown in Suppl. Fig. 1.

For the construction of transposase expressing helper plasmids, the coding sequence of PGBD1, PGBD2, PGBD3, CSB-PGBD3, PGBD4, and PGBD5 proteins were PCR amplified with Phusion™ High-Fidelity DNA Polymerase (Thermo Fisher Scientific) from cDNA and inserted into pGEM®-T vector (Promega). Sequences were verified by Sanger sequencing. Coding sequences were cloned into CMV promoter driven pEGFP-C1 or pEGFP-N1 expression vectors from Clontech by replacing the GFP (pPGBD1, pPGBD2, pPGBD3, pCSB-PGBD3, pPGBD4, pPGBD5). In pPGBD3, the predicted catalytic amino acid triad DND (Keith et al., 2008) was reconstituted to DDD by point mutation of aa 352 from N to D (pPGBD3N/D) by site-directed PCR mutagenesis.

For PGBD5, apart from the amplified, full length protein encoding cDNA (identical to the reference sequence of Gene ID #79605), we also used two other protein species for the transposition experiments. At the beginning of our studies, expression databases for various Primate species listed a PGBD5 protein variant lacking or different in the region encoded by exon 5. Although some of these entries have recently been curated in NCBI we still used this exon 5 deletion protein named PGBD5del as a control. In the pPGBD5del helper plasmid, exon 5 was deleted from pPGBD5 by the FastCloning method (Li et al., 2011) with the amplification of the plasmid by primers with overlapping complementary 5′ ends. As a third PGBD5 protein, we requested the plasmid pRecLV103-GFP-PGBD5 from Addgene which was deposited from the work of Henssen et al. (2015). The coding sequence of this expression plasmid differs from that we have cloned (identical to the reference sequence of Gene ID #79605), as it contains an alternative exon 1 (most likely an earlier cloning artefact which has been permanently suppressed in the NCBI database, see https://www.ncbi.nlm.nih.gov/nucco

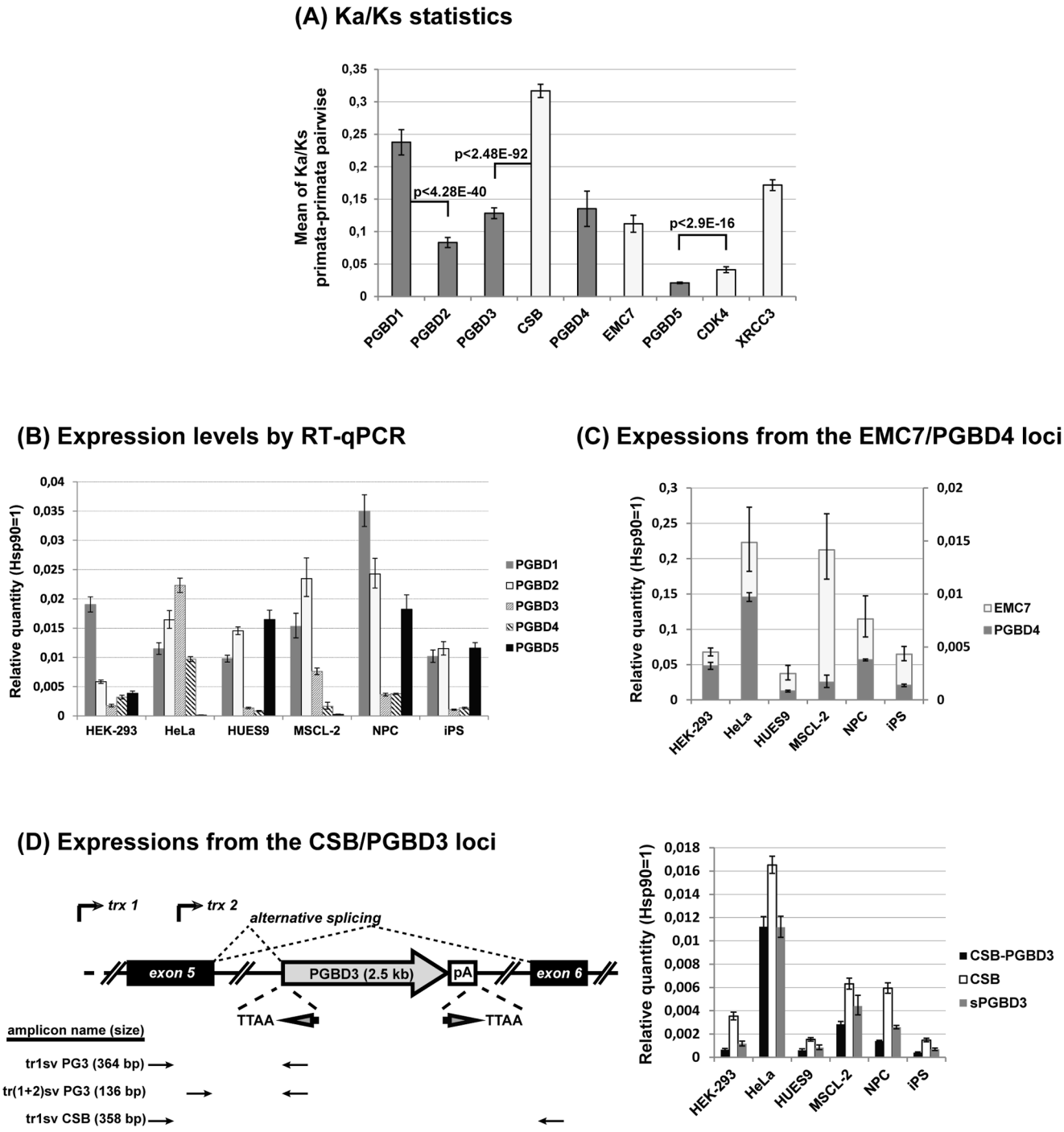

Fig. 1. Evolutionary analyses and the expression profiles of human PGBDs (A) *Ka/Ks* statistics of PGBDs based on comparative primate-primate pairwise analyses. Dark columns represent PGBD1-5, white ones are other endogenous genes; p values are indicated for certain selected cases (see text for explanation). **(B)** Gene expression analysis of PGBDs in different human cell lines by RT-qPCR. **(C)** Comparison of expression levels from the EMC7/PGBD4 loci. Left y-axis scale is for EMC7, right scale is for PGBD4. **(D)** Comparison of expression levels from the CSB/PGBD3 locus. (Left panel) Structure of the locus and primer pairs used for the analysis. (Right panel) Expression analysis of three different mRNA variants. Arrow depicted as "trx 1" indicates transcription from the canonical gene promoter, whereas "trx 2" shows transcription starting from a cryptic promoter located in exon 5; pA: polyA signal; "TTAA" sequences are the original target site duplications of the domesticated PGBD3 transposase; Sv stands for splice variant. Tr1svPG3 primer pair detects mRNA producing CSB-PGBD3; tr1svCSB pair detects mRNA producing CSB; tr(1 + 2)svPG3 detects all PGBD3 splice variants including the one from the cryptic transcript. The latter produces a solo PGBD3 protein (sPGBD3) encoding transcript, which can be calculated from the measurements of tr(1 + 2)svPG3 and tr1svPG3. For the quantification of PGBD3 in Fig. 1B, tr(1 + 2)svPG3 primer pair was used detecting both PGBD3 splice variants. In all graphs, mean values of at least 3 independent measurements are given, and error bars represent 95% confidence intervals.

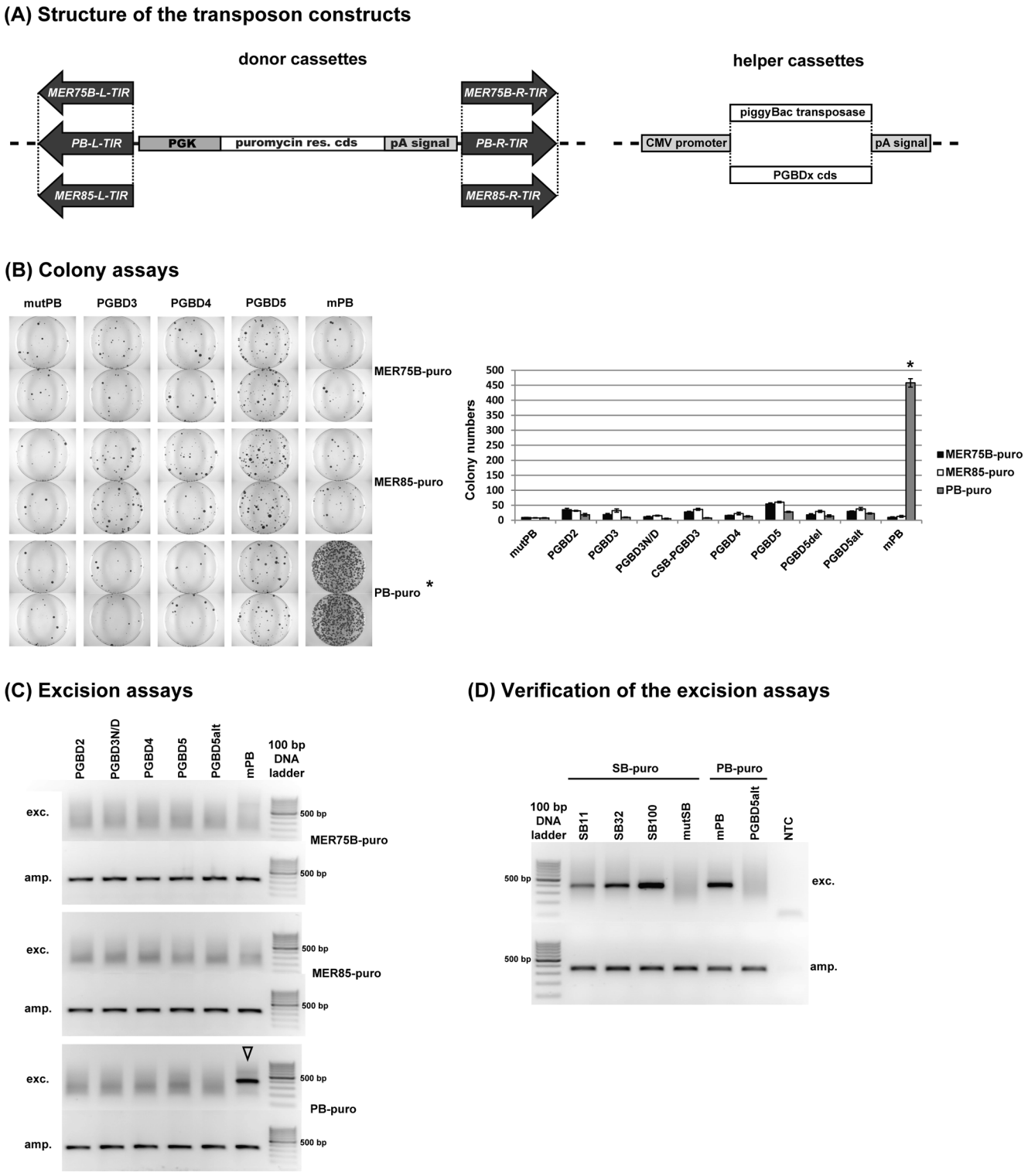

Fig. 2. Transposition assays of PGBDs in HEK-293 cell. (A) Three different transposon donors were tested: MER75B, MER85, and PB. They carry an antibiotic selection cassette, and their structures differ only in the specific transposon TIRs. The helper constructs express either the *piggyBac* transposase or the different PGBDs to be tested (x = 1–5). L-TIR: left-TIR; R-TIR: right-TIR; PGK: phosphoglycerate kinase promoter; CMV: cytomegalovirus promoter; cds: coding sequence; pA signal: polyadenylation signal. **(B)** Cells transfected with the two components (the selected donor and transposase expressing helper plasmids) were then selected by puromycin, and the surviving transgenic cells were stained and counted by the colony assay. (Left panel) Selected representative photos of the colony assays with the appropriate controls, showing two technical replicates for each assay. (Right panel) Quantifications of colony assays represented by the means of at least three independent experiments, error bars indicating standard deviations. ***:** The shown positive control mPB/PB-puro co-transfections were selected from fivefold dilutions. **(C)** Transfected cells were tested for excision events by PCR, where the product (381 bp, 'exc.') indicates the excised and repaired donor plasmid copies. Assay control was the ampicillin sequence, present on the donor constructs, with a 340 bp PCR product. **(D)** Sensitivity of the excision assay was tested by three SB variants with different activities, and compared to the PGBD5alt variant (a GFP-tagged construct used by Henssen et al., 2015) in parallel transfections. A representative result of several independent experiments is shown. NTC: non-template (water) control.

re/NM_024554) and N-terminally fused GFP; hence we named this variant as PGBD5alt (pPGBD5alt). The plasmid also contains a puromycin selection cassette but since our transposon donor system is based on puromycin selection, we partially deleted the puromycin cassette in this plasmid by SacII digestion and re-ligation. This was a necessary modification, otherwise random integration of the pPGBD5alt helper plasmid would increase colony numbers during the transposition assay.

The following transposase protein variants were used as positive and negative controls: for the PB system, the original *piggyBac* in a mammalian codon-optimized form, denoted as mPB (Cadiñanos and Bradley, 2007); for the *Sleeping Beauty* (SB) system, the SB11, SB32, SB100X and mutSB variants (Mates et al., 2009). Transposon donor for SB was identical to pPB-puro except TIR sequences (pSB-puro) (Kolacsek and Orban, 2018). To construct a negative control for the PB system, we generated the Asp268Gly mutant at the DDD catalytic site of mPB by PCR mutagenesis (mutPB).

2.5. Transposition assay and Western blotting

Transfections were carried out in duplicates. 2×10^5 of HEK-293 or HeLa cells were seeded onto 12-well plates, and transfected with FuGENE® 6 according to the manufacturer's instructions (Roche Applied Science). Next day, 200 ng transposon donor and 400 ng transposase expressing helper plasmid and 200 ng GFP expressing pEGFP-C1 transfection control plasmid from Clontech were transfected (the optimal ratios were chosen based on our previous study (Kolacsek et al., 2014)). In the case of pPGBD5alt helper plasmid that expresses GFP as an N-terminal fusion, the transfection control plasmid was replaced with pET-41 (Novagen). At 48 h post-transfection, cells were harvested and one tenth of them were analyzed by Attune NxT flow cytometer (Thermo Fisher Scientific) for GFP expressing cells to control transfection efficiency. Transfection efficiencies were between 53 and 59%. For antibiotic selection, 1% of the transfected cells were seeded onto cell culture Petri dishes, selected for two weeks in 1 μg/ml puromycin (Sigma-Aldrich), then surviving cells were fixed in ice cold methanol and stained with 0,04% Crystal Violet (Sigma-Aldrich) in 25% methanol. Colonies were quantified with the Universal Hood Gel Imager Model # 75S, using the Quantity One 4.4.0 software (BioRad).

After processing the transfected cells for colony assays and for flow cytometry measurements at 48 h post-transfection, the remaining cells were used for excision assay. Duplicated transfections were pooled and plasmids were isolated from the cells using a modified protocol of the QIAGEN plasmid Miniprep Kit, applying 250 μl of 1.2% SDS supplemented with 50 μg of Proteinase K for cell lysis step and protein removal from the DNA. 10 ng of isolated plasmid DNA was used in a two-round nested PCR assay where primers were specific to the plasmid backbone and resulted in a PCR product only from the excised and repaired plasmid copies due to large transposon cassettes. PCRs were carried out in 50 μl reaction volumes using 200 nM primers with 2xPCR master mix (Promega) and with the following thermal profile: 95 °C 5 min, 35 cycles of 95 °C 30 s 60 °C 30 s and 72 °C 1 min, 72 °C 5 min. 5 μl of 200-fold diluted first round product was used for the second round. Ampicillin sequence presenting in all donor construct was used as assay control, PCR conditions were the same as for the first round PCR of excision. Products were separated in a 2% agarose gel and visualized by ethidium bromide staining. PCR primers for excision analysis are listed in Suppl. Table 3.

Separate wells were transfected for Western blotting to detect transient PGBD protein expression. Cells were lysed at 48 h post-transfection by the addition of TE sample buffer (0.1 M TRIS-PO_4, 4% SDS, 40 mM Na-EDTA, 40% glycerol, 0.04% bromophenol blue, and 4% β-mercaptoethanol; materials from Sigma-Aldrich). Samples were sonicated to shear DNA and equal amounts were loaded on a 7.5% acrylamide gel (SDS-PAGE) (Sigma-Aldrich). Proteins were transferred to a 0.2 μm PVDF membrane (BioRad). Blots were probed with the following primary antibodies: anti-PGBD1 (Abgent, cat. AP17268C), anti-PGBD2 (Biorbyt, cat. orb519701), anti-PGBD3 (Sigma-Aldrich, cat. HPA025825), anti-PGBD4 (Sigma-Aldrich, cat. HPA040896), anti-PGBD5 (Biorbyt, cat. orb13159), and anti-β-actin (Sigma, cat. A3854) HRP conjugate. Donkey anti-rabbit IgG (H + L) HRP conjugate secondary antibody (Jackson ImmunoResearch, cat. 711–035-152) were used to visualize the results. Detection was performed with Pierce™ ECL Western Blotting Substrate (BioRad, cat. 32106).

2.6. Stress conditions

1.5×10^5 HEK-293 and 1×10^5 HeLa cells were seeded onto 24-well plates. Treatments were carried out in 6 parallels. Next day medium was replaced with 50 μM deferasirox (DFX) (MedChem Express) or 4 mM $KBrO_3$ (Sigma-Aldrich) for inducing hypoxic or oxidative stress, respectively. Concerning $KBrO_3$, we chose the higher possible concentration without severe cell loss (Luan et al., 2007). After 24 h cells were harvested, and parallels were pooled for total RNA purification, reverse transcription and RT-qPCR (see 2.3). For UV light exposure, the medium was removed and cells were rinsed with 40 μl PBS. UV irradiation was performed using the Stratalinker UV Crosslinker Model 1800, delivering 20 J/m^2 dose of UV light. Cells were recovered by adding back the culture medium, and samples were harvested after 6 h or 18 h posttreatment.

3. Results

3.1. Evolution of human piggyBac-derived sequences

Active TEs spread throughout the genome but protection of genome stability acts as a selective pressure and mutations can eventually inactivate the TE copies. However, several examples reveal that by gaining new functions, domesticated copies can also be preserved in the evolution by purifying selection. While classical TEs are present in multi-copies within and between species, domesticated elements are generally found in single orthologous copies. The PB DNA transposon family is unique in that domestication for useful function exemplifies the retention of some transposase-like aspects such as the targeted removal of endogenous DNA fragments by the cut and paste mechanism (Hikosaka et al., 2007; Baudry et al., 2009; Cheng et al., 2010).

The first part of our study focuses on the evolution of the human PGBD elements. As described earlier (Sarkar et al., 2003), the N-terminal region (positions 1–130, using *T. ni* transposase as reference) is not well conserved, therefore as opposed to other phylogenetic analyses, we have restricted our conservation analysis to primate sequences and to the IS4/5 transposase-like domain (Suppl. Table 1). The phylogenetic tree inferred from 24 species and 111 PGBD sequences (Suppl. Fig. 2) revealed that PGBD1 and PGBD2 are more closely related, and they are found mostly in mammals (Sarkar et al., 2003; Bouallegue et al., 2017; Raskó et al., 2021). PGBD3 is clustered closest to them (Suppl. Fig. 2), invaded into the primate lineage, but relatives of PGBD3 can also be found in the aphid *Acyrthosiphum pisum* and in the spider *Stegodyphus mimosarum,* although they do not appear to be domesticated copies (Bouallegue et al., 2017). PGBD4 is also widespread among primates but clustered separately from PGBD3, PGBD1 and PGBD2 (Suppl. Fig. 2). The closest relatives of PGBD4 are in bat flying fox *Pteropus vampyrus*, in aphid *Acyrthosiphon pisum*, and in moth *Spodoptera frugiperda* (Bouallegue et al., 2017). Inconsistencies between the transposase tree and the species phylogeny suggest that the PB family might be frequently and successfully horizontally transferred. In addition, PGBD4 is more related to domesticated elements in lower phyla: PGM and TPB2 found in ciliates, and KOBUTA in *Xenopus* (Bouallegue et al., 2017). PGBD5 is the most ubiquitous member present in a large spectrum of species and is clustered quite distinctly from the other PGBDs (Suppl. Fig. 2). As previously described (Pavelitz et al., 2013), this reflects an early domestication event of PGBD5 in the cephalochordate lineage. Moreover, no sequence similarity of the PGBD5 flanking regions can be found in

hemichordates, echinoderms, and urochordates, supporting the conservation of the entire region, including the flanking sequences (Bouallegue et al., 2017).

During evolution if a new function has been acquired, low ratio (*Ka/Ks* < 1) of nonsynonymous (*Ka*) to synonymous (*Ks*) nucleotide substitution rates suggests that the sequence is under strong purifying selection, while a high ratio (*Ka/Ks* > 1) indicates that adaptive peak is not yet reached (Hurst, 2002). Earlier calculations of the *Ka/Ks* ratio of PGBDs provide arguments in favor of domestication of these sequences (Newman et al., 2008; Bouallegue et al., 2017). Here we analyzed the level of conservation of the primate PGBD elements, comparing them to some endogenous genes, including those expressed from the same loci (sequences are in Suppl. Table 1, *Ka/Ks* calculations are in Suppl. Table 2). Comparison of *Ka/Ks* statistics revealed that PGBD5 is the most conserved among PGBDs, it is subjected to a higher selective pressure than the conserved cell cycle regulator gene CDK4 (Fig. 1A). Values of other PGBDs were found to be comparable to endogenous genes like the DNA repair factor XRCC3, falling into the category of protein coding genes under strong purifying selection (Hurst, 2002). PGBD4 and EMC7, regulated by the same promoter, are similarly conserved, however, the sequence of PGBD3 is under much higher selective pressure than the coding region of the *CSB* "host" gene. The latter one has been described to have DNA repair activity (Sarker et al., 2005), yet it seems less conserved than the other analyzed endogenous repair factor, XRCC3 (Fig. 1A). Another notable result is that there is a three-fold difference between the conservation of the separately clustered and closely related PGBD1 and PGBD2 (Fig. 1A), suggesting that PGBD2 is also a strongly conserved, domesticated gene with a currently unknown function. Our analysis may differ from the previous ones (Sarkar et al., 2003; Bouallegue et al., 2017; Helou et al., 2021b), because it is restricted to primate lineage and is limited strictly to the DDD transposase-derived domain omitting the N-terminal regions.

3.2. Cell type specific expression profiles of human PGBDs

When obtaining information on the expression profiles of PGBDs in online databases, such as BioGPS (Wu et al., 2013) or the Human Protein Atlas (https://www.proteinatlas.org/), PGBD5 shows strong neural-specific expression, which was confirmed by Pavelitz et al. (2013) by *in situ* hybridization. Concerning the other PGBD genes, however, they show more diverse expression patterns. We have developed a qPCR-based quantification method (the Ct shift method) to accurately compare the levels of different nucleic acids (Kolacsek et al., 2017) (see also section 2.3.). Using this method, we measured expression levels of all PGBDs in six cell lines: HEK-293, HeLa, a human embryonic stem cell line (HUES9), a mesenchymal stem cell-like cell line (MSCL-2, differentiated *in vitro* from HUES9 (Varga et al., 2011)), an induced pluripotent stem cell line (62F), and hippocampal neural progenitor cells (differentiated *in vitro* from the 62F iPSC line (Vofely et al., 2018)). Comparing to the Hsp90 as a reference gene, the PGBD expression levels were consistently found to be two orders of magnitude lower (Fig. 1B). This magnitude may indicate a general expression level that is well tolerated by the cell in regard of this type of external sequences. While PGBD1 and PGBD2 are similarly expressed in all cell lines examined, PGBD3, PGBD4 and PGBD5 show more cell line specific differences (Fig. 1B). Interestingly, the expression profiles of PGBDs in NPCs were similar to those in HUES9 and iPS cell lines: PGBD3 and PGBD4 are expressed at a relatively low level, whereas PGBD5 is similarly abundant as PGBD1 and PGBD2. Our result supports the Primary Cell Atlas dataset of the BioGPS database in relation of PGBD5 (Mabbott et al., 2013), where in addition to neural expression, embryonic stem cells and some iPSs also show elevated expression levels. Our data suggests that PGBD5 might function in pluripotent cells, and also in differentiated cells of the neural lineage.

We compared mRNA levels expressed from shared loci in cases of PGBD4 and PGBD3 (Fig. 1C and D). We have measured EMC7 mRNA as one order of magnitude higher than PGBD4 (Fig. 1C), which confirmed that the promoter shared with PGBD4 induces transcription originally in the direction of EMC7. Expression differences between EMC7 and PGBD4 were ranging from ~ 20- to ~ 50 fold, with the exception of cell line MSCL-2 where difference was ~ 130 fold. These results indicate the regulated promoter directionality but post-transcriptional regulation (e.g. mRNA stability) may also contribute to this high difference between EMC7 and PGBD4 mRNA steady state levels.

Human PGBD3 is of special interest, because its host gene, CSB is mutated in the Cockayne Syndrome (CS) that is a rare genetic disease characterized by neurological problems, growth failure and premature ageing. The study of Newman et al. (2008) showed that the mutation spectrum of CSB gene in CS in general prevents the production of the full length CSB but allows to express the fusion CSB-PGBD3 gene product. However, the complete absence of the CSB locus products does not cause severe progeria of CS, only mild UV sensitivity (Horibata et al., 2004). Thus, the exclusive presence of CSB-PGBD3 may play a causative role in CS (Weiner and Gray, 2013). Nevertheless, it is possible that CSB-PGBD3 fusion protein is important in both health and disease, because it may confer a metabolic advantage presumably in DNA damaging stress, but exerts this effect only if CSB-PGBD3 is co-expressed with CSB (Bailey et al., 2012). The complex nature of the locus results in a complex pattern of mRNA and protein isoforms, with the solo PGBD3 expressed from a cryptic promoter located in exon 5 (Fig. 1D, left panel) (Newman et al., 2008; Kolacsek et al., 2017). Examining all the gene products separately by the Ct shift method, we could reveal that both PGBD3 splice variants are expressed in a significant amount as compared to the canonical CSB mRNA (Fig. 1D, right panel). These PGBD3 splice variants are present in lower amounts than the CSB transcript in all cell types, but the extent is different, suggesting tissue specific differences in splicing regulation. Post-transcriptional regulation cannot be ruled out, but correlation between the fusion and solo PGBD3 mRNA product rather indicates the regulation of splicing (Fig. 1D, right panel). Overall, when compared to the CSB transcript, the dose of the PGBD3 containing mRNAs is the highest in HeLa and MSCL-2 cell lines which may confer the metabolic advantage described earlier for the co-presence of CSB and CSB-PGBD3 (Bailey et al., 2012).

Taken together, our data indicate that all human PGBDs are regulated distinctly from each other and in a cell type specific manner. These results are in favor of the hypothesis that these transposase-originated genes represent individual domestication events. Nevertheless, it is a crucial question whether they retained their transposase activity, therefore we continued to systematically investigate this issue for all PGBDs.

3.3. Lack of transposase activity of all PGBDs

Concerning the human of the PB superfamily, the predicted DDD catalytic aspartate amino acids at positions of 268, 346, and 447 of the *T. ni* PB transposase are not conserved among the domesticated PGBD sequences (Sarkar et al., 2003; Newman et al., 2008; Pavelitz et al., 2013). However, mutation analysis revealed a fourth aspartate amino acid (D450) that is also essential for transposition of *T. ni* PB, but it has a tolerance for a glutamate substitution (Keith et al., 2008). Intriguingly, we have found the fourth D to be conserved among all the primate PGBD orthologs and paralogs (data not shown), so the position is likely to play an essential role in the function of these PB family members. PGBD3 and PGBD4 are the youngest domesticated elements, and presumably have acquired only a few inactivating mutations. Nevertheless, only PGBD4 retained all the essential catalytic amino acids DDDD, so it cannot be excluded to have transposase activity (Mitra et al., 2008). PGBD3 has gained a D to N point mutation at position 352, so it has a DNDD motif (Sarkar et al., 2003; Newman et al., 2008). On the other hand, Gray et al. observed that CSB-PGBD3 is capable of binding to the TIR of PGBD3 transposon and of its residual MER85 elements *in vitro* and *in vivo*, providing a hint that MER85 could have been the integral part of an

ancient PGBD3-like functional transposon (Gray et al., 2012). This prompted us to test the presumably active, "reconstituted" PGBD3 mutant having the expected DDDD motif. Apart from testing the reactivity with the insect-derived PB system, we also tested the MER75 and the MER85 elements of recognizable PB origin present in the human genome, the potential substrates of ancient PGBD4 and PGBD3 transposases, respectively. We have cloned representative copies of these MER sequences from the human genome, and constructed transposon donor plasmids by inserting a selection marker of puromycin resistance into the internal region (Fig. 2A). The two component transfection system was used in HEK-293 cells, where helper plasmids expressed the appropriate human PGBD variant proteins. After two weeks of antibiotic selection, transgenic efficiencies were tested by colony counting. As a result, neither PGBD3 nor PGBD4 showed any indication of transposition when compared to the background control (the insect PB catalytic mutant, Fig. 2B).

To continue with the systematic evaluation of all *piggyBac*-related transposition activity in the human genome, we tested all human PGBD proteins for their potential interaction with either the insect PB vector system or the two MER elements. This was also urged by a somewhat stunning previous study showing that PGBD5 has significant classical transposase activity for the TIR of *T. ni* PB transposon, in spite of this ortholog being the oldest among human PGBDs, clearly lacking recognizable flanking TIR sequences. The authors also claimed that it is not the components of the originally predicted 'DDD' motifs but rather amino acids in different positions were the ones responsible for this unexpected activity (Henssen et al., 2015). However, due to the ambiguity of the encoded protein prediction this group used, we tested not only their GFP-tagged version (PGBD5alt), but also the currently available updated sequence variant (PGBD5, Gene ID: 79605), as well as an alternative variant lacking exon 5 (PGBD5del) which also appeared earlier in expression databases (see section 2.4.). Successful transient exogenous expressions of all human PGBD proteins were confirmed by Western blot (Suppl. Fig. 3). As a result of the colony assays, none of the PGBDs showed reactivity with either the insect PB vector system or the MER elements: the colony counts produced by all PGBDs and their variants were far below the mPB activity (Fig. 2B). On the other hand, we detected somewhat elevated colony numbers in case of PGBD1 (see details in our concurrent publication, (Raskó et al., 2021)), and to a lower extent, for PGBD5. This prompted us to test whether the first step of transposition occurs in case of any combinations, therefore we performed excision assays detecting the capability of the human PGBDs to cut out the transposon cassettes. However, with the exception of the positive control (the insect mPB), no excision events were detectable by a two round nested PCR (Fig. 2C), indicating the lack of transposition mediated by the human PGBDs or their variants. To test the sensitivity of our excision PCR, we have made parallel transfections by the *Sleeping Beauty* transposon variants with different activity (Kolacsek et al., 2014) and the PGBD5alt used by Henssen et al. (2015), for which authors detected transposase activity. Based on this, it can be concluded that even the low excision activity of the SB11 variant is still well-detectable by this PCR assay but we were unable to reproduce the result of Henssen et al. (2015) for their used PGBD5alt variant (Fig. 2D).

Another potential explanation for not detecting transposition activity would be the toxicity caused by the expression of PGBDs: in that case, the death of the transfected cells would prevent transposition to occur. To exclude that, we performed a cell viability assay comparing the overexpression of all PGBDs to the effect of mock transfection and to the expression of other unrelated expression vectors in both HeLa and HEK-293 cells. As a result, no elevated toxicity could be detected for overexpressing any of the human PGBD proteins as compared to the controls or to the insect mPB transposase (Suppl. Fig. 4). Taken together, all these results indicate the lack of *bona fide* transposition activity for all human PGBDs, including all the tested PGBD5 variants.

3.4. Expression changes of PGBDs in stress conditions

It is likely that *PGBD* transposases have been conserved due to their interactions with various cellular factors, thus we can assume that they also retained their expression pattern regulation. Several transposases interact with DNA repair factors to increase their efficiency, or to modulate the cell cycle, often depending on the environmental conditions (Zayed et al., 2003; Izsvak et al., 2004; Walisko et al., 2006). We hypothesized that PGBDs may still have these cellular connections, therefore we investigated the main cell cycle regulating circumstances, the DNA damaging stress conditions, whether they can influence the expression patterns of the domesticated primate PGBD proteins. What is common in hypoxia, oxidative stress and UV stress is that they cause DNA damage, and by activating p53 responsive genes like *p21*, they regulate cell cycle progression (Bunz et al., 1998; Pan et al., 2004; Pires et al., 2010). Expression studies were carried out in two cell lines, HEK-293 and HeLa, and hypoxic conditions resulted in similar induction of all domesticated PGBD elements in both cell lines (Fig. 3A, left panel). Expressions of mRNAs from the shared loci of *PGBD3* and *PGBD4* were concordantly induced, but to different extents (Fig. 3A, right panel). Induction of CSB was two times higher than PGBD3, which again confirms strong regulation difference between the two splicing products expressed from the same promoter. Induction of EMC7 also differed from PGBD4 in HEK-293 cells, but not in HeLa cells, suggesting that posttranscriptional regulation may also contribute to their expression profiles.

In oxidative stress conditions, PGBD3 and PGBD4 expressions responded concordantly in the two cell lines, both being reduced (Fig. 3B, left panel), but the effect on PGBD3 in HeLa cells was more pronounced, producing an eight-fold reduction ($-\Delta\Delta Ct \sim -3$). Measuring the effect on CSB revealed that the entire locus seems to be downregulated in HeLa (Fig. 3B, right panel), but not in HEK-293 cell line where only the level of the PGBD3 splice variant is reduced. Concerning the other locus with a shared promoter, the expression level of EMC7 in oxidative stress conditions was not significantly changed in either cell line, in contrast to PGBD4 (Fig. 3B, right panel). However, in UV stress conditions, only PGBD5 showed a significant response of 4–5-fold downregulation (Fig. 3C), that was an immediate effect after 6 h of treatment (data not shown). The distinct expression patterns indicate that these domesticated genes may be involved in completely different cellular processes.

4. Discussion

There are several *piggyBac*-related elements present in the human genome and both the transposase-derived protein coding genes and the repetitive TIR sequences raised concerns about the safe applications of the widely used insect *piggyBac*-based genetic tools. An intense debate was therefore initiated when the PGBD5 protein was recently claimed to have transposase activity (Henssen et al., 2015; Ivics, 2016). Based on this unexpected result, we have decided to systematically investigate this issue by examining whether these *piggyBac*-related elements still have transposase activity or to the contrary, whether we can provide indications for them being domesticated endogenous genes.

Focusing on the PGBD copies present in primates, our phylogenetic analysis supports the hypothesis about the frequent horizontal transfer of PB elements also in this lineage (Pagan et al., 2010), with PGBD5 being the most ancient homolog without obvious flanking TIR sequences, pointing towards a domesticated function, rather than towards a conserved transposase activity. Our results are in line with the study of Pavelitz *et al.*, concluding that PGBD5 protein is a neuron specific endogenous element, and also claiming that it is unlikely that it can bind DNA in brain nuclei, which would be necessary for transposition (Pavelitz et al., 2013). Analyzing the gene structures of the other homologs, the lack of TIR sequences and the examples of fusion transcripts with endogenous protein coding genes are also in favor of the

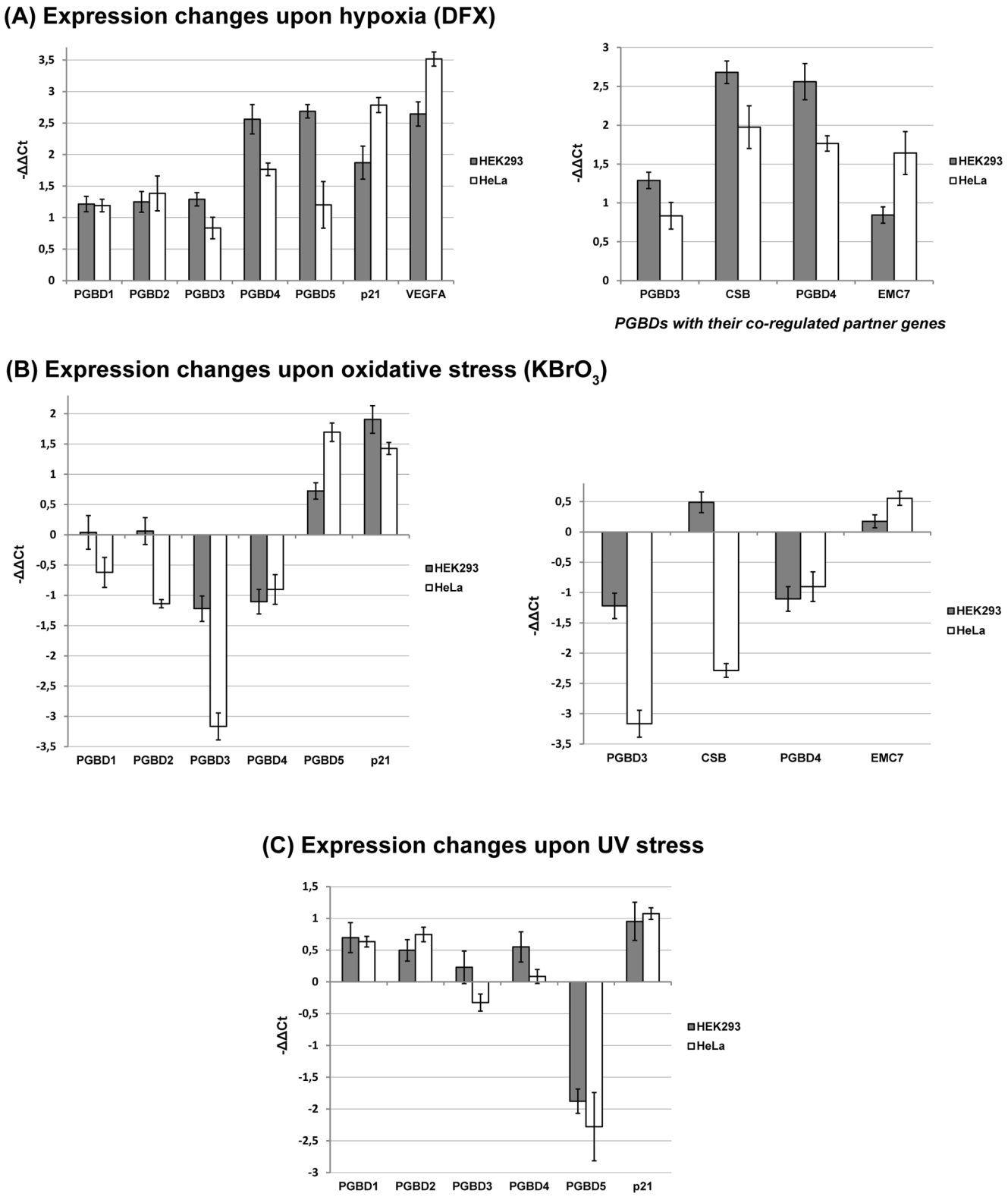

Fig. 3. Expression changes of PGBDs upon different stress conditions. Stress treatments were carried out in HEK-293 and HeLa cells, mRNA levels were measured by RT-qPCR. **(A)** Cells were treated with deferasirox (DFX) to induce hypoxic response. (Left panel) PGBD1-5 expression changes were measured, p21 and Vascular Endothelial Growth Factor A (VEGFA) were used as stress condition controls. (Right panel) Expression changes shown for the shared loci of CSB/PGBD3 and EMC7/PGBD4. **(B)** Cells were treated with KBrO3 to induce oxidative stress. (Left panel) PGBD1-5 expression changes were measured, p21 was used as stress condition control. (Right panel) Expression changes shown for the shared loci of CSB/PGBD3 and EMC7/PGBD4. **(C)** Cells were radiated with UV light, and PGBD1-5 expression changes were measured 18 h after treatment; p21 was used as stress condition control. For the analysis of PGBD3, the tr(1 + 2)svPG3 primer pair was used to detect both PGBD3 splice variants (see Fig. 1D, left panel). Mean values of at least three independent measurements are given, error bars represent 95% confidence intervals.

domestication hypothesis. Only PGBD3 seems to have preserved the structure of an active ancient transposon with MER85 TIR sequences present on both flanking regions; however, the mutation present in the DDD motif indicates the inability to perform the "canonical" cut-and-paste transposition reaction. When examining the expression profiles of human PGBDs in several types of cell lines, the selective and carefully regulated expression patterns also argue for specific endogenous functions. Our conclusion was that *PGBDs* have most likely been domesticated in primates but their functions are unlikely to be associated with canonical transposase activity.

To directly examine the ability of human PGBDs to perform transposition, we tested each transposase-like proteins for mobilization of the insect transposon sequence, as well as for the MER85 and the MER75 sequences. The most striking result was that we could show the lack of transposase activity for all five human PGBD genes. Even the best candidate, PGBD3 and CSB-PGBD3 fusion product gave negative results: although both were shown to bind to MER85 sequences (Bailey et al., 2012; Gray et al., 2012), our "restored" catalytic domain variant (PGBD3N/D) was not able to perform transposition either. This result shows that even the "youngest" PGBD homolog has lost the canonical DNA transposase activity. In addition and of most importance, we could not detect transposase activity for any of the tested PGBD5 variants, even in various cell lines (e.g. in HeLa, data not shown). Although in a later publication, Henssen et al. presents a more careful interpretation of PGBD5 having "only" some genome-cutting ability, possibly causing oncogenic mutations (Henssen et al., 2017), our results are in contrast with two recent papers claiming a lower but detectable transposase activity of PGBD5 (Helou et al., 2021a; Helou et al., 2021b). On the other hand, a more recent study is in agreement with our results, presenting the lack of cross-reactivity among the insect *piggyBac*, the bat *piggyBat* transposase and human PGBD5 (Beckermann et al., 2021). They found no evidence of PGBD5 transposase activity, moreover, they showed that the two other transposases have restricted activity toward their own cognate sequences, showing no cross-mobilization of each other's transposon elements (Beckermann et al., 2021). Similarly to our approach, this group also failed to reproduce the results of Henssen et al., either using the original GFP-tagged PGBD5 or the corrected PGBD5 transposase sequence variant. Our and their results are all in favor of PGBD5 being a domesticated endogenous gene, having lost the ancient transposase activity.

In an earlier study, Saha *et al.* have found a few solo 14–16 bp sequences in the human genome that are identical with the end of the TIR of *T. ni* PB transposon sequence (Saha et al., 2015). Authors examined whether the insect PB transposase used in gene delivery is able to mobilize elements with these genomic sequences. Their result showed that the PB vector system is safe in this aspect. Here we examined the MER75 and the MER85 elements in a similar manner and our MER donor constructs do not show reactivity with the mPB transposase. Our results confirm the safety of the application of PB system for gene transfer in all aspects: neither endogenous PGBDs can remobilize PB vectors integrated in the genome, nor the PB transposase can react with endogenous residual PB-like sequences of the human genome. Both scenarios would pose a risk of genetic instability, however, increasing number of observations suggests that it should not be feared of.

In spite of their inability to perform a cut-and-paste reaction, we still detected somewhat elevated colony numbers resulting from the co-transfection of PGBDs and donors, as compared to the background values obtained with the inactive variant, mutPB (Fig. 2B, right panel). We hypothesize that high exogenous PGBD expression may aid the random integration of donor sequences, especially given the number of MER sequences in the human genome. This view is supported by the fact that PGBDs with the most elevated background (PGBD1 and PGBD5) lack the C-terminal cysteine rich domain (CRD) responsible for sequence specific DNA binding of the *T. ni* PB transposase (Keith et al., 2008; Morellet et al., 2018; Chen et al., 2020), making it unlikely that these would mobilize any transposon in a sequence-specific manner (concerning PGBD1, see our concurrent paper (Raskó et al., 2021)). The two recently published works on PGBD5 showed similar results (Helou et al., 2021a; Helou et al., 2021b): cotransfecting the human PGBD5 and PB transposon substrates, they obtained a few fold elevation in the background level of colony numbers; however, the authors evaluate these results as transposase activity of the domesticated PGBD5. On the other hand, these may be explained alternatively: as a general behavior of transposases, sequence specific or nonspecific DNA binding, together with oligomerization may facilitate the formation of transposase-like synaptic complexes, resulting in enhanced repair of randomly damaged regions by integrating unrelated DNA sequences, including transfected plasmid templates.

In the last part of our study, we aimed to investigate the potential cellular functions of the human PGBDs. In general, transposon activity and expression are sensitive to and regulated by stress conditions, therefore we tested if this response have been retained for the domesticated PGBDs. There were some indications for that: in the case of the co-regulated CSB and PGBD3, the former one plays a central role in the cellular response to stress as a chromatin remodeling factor through modulating transcriptional changes, e.g. response to hypoxia, DNA damage, and also response to insulin-like growth factor-1 (Filippi et al., 2008; Velez-Cruz and Egly, 2013). Also, in the promoter of the CSB gene there are two hypoxia response elements where the HIF-1α transcription factor can bind (Filippi et al., 2008). In addition, the elevated random integration of donor constructs with co-transfection of certain PGBDs raises the possibility that the transposase originated proteins could interact with DNA repair factors during processing of randomly damaged plasmid DNA. We tested stress conditions causing potential DNA damage and in response to hypoxia, we detected elevated expression levels for all PGBDs, although to a different extent. On the other hand, no common expression changes were found under oxidative stress and UV stress conditions: as opposed to the other homologs, PGBD3 and PGBD4 responded with a decrease to oxidative stress in all cell types examined, while PGBD5 was upregulated in oxidative stress but downregulated in UV stress conditions. These opposite behaviors indicate that these domesticated genes may be involved in different cellular processes, however, the distinct and often cell type specific expression responses support the hypothesis that all PGBDs are domesticated genes with precise endogenous regulations.

In conclusion, our systematic investigations revealed that none of the human PGBD elements retained their original ability for transposition. On the other hand, several pieces of evidence support the hypothesis that these elements have been "co-opted" during evolution and serve distinct domesticated functions. The low Ka/Ks values for many of them indicate that they are under a strong purifying selection, and in some cases, they seem more conserved that the co-regulated transcripts of the same locus (e.g., PGBD3 vs CSB). In addition, for such co-regulated transcripts for PGBD3 and PGBD4, the splicing and the transcription pattern is markedly different and actively regulated in a cell type specific manner, indicating endogenous functions. Finally, the expression profile changes of all human PGBDs in various stress conditions also point to the fact that the roles of these genes can likely be placed in different cellular pathways. Taken together, although the exact and distinct domesticated functions of human PGBDs still require further investigations, the lack of their transposition ability provides a necessary safety background for using the *piggyBac* gene delivery system for potential gene therapy applications in the future.

CRediT authorship contribution statement

Orsolya Kolacsek: Investigation, Methodology, Data curation, Validation, Writing – original draft, Writing – review & editing. **Gerda Wachtl:** Investigation, Methodology. **Ábel Fóthi:** Data curation, Visualization. **Anita Schamberger:** Investigation. **Sára Sándor:** Investigation. **Enikő Pergel:** Investigation. **Nóra Varga:** Investigation. **Tamás Raskó:** Validation, Formal analysis, Writing – review & editing.

Zsuzsanna Izsvák: Validation, Formal analysis, Writing – review & editing. **Ágota Apáti:** Investigation, Validation, Funding acquisition. **Tamás I. Orbán:** Conceptualization, Supervision, Data curation, Validation, Funding acquisition, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors are indebted to Kornélia Némethy and Beáta Haraszti for their excellent technical help. The human embryonic stem cell line (HUES9) and iPSC line (62F) were kindly provided by Douglas Melton (HHMI) and Fred H. Gage (Salk Institute), respectively. This study was supported by grants from the National Research, Development and Innovation Fund of Hungary: VEKOP-2.1.1-15-2016-00156, VEKOP-2.3.3-15-2017-00014, 2018-1.2.1-NKP-2018-00005 grants to T.I.O., NKFIH-OTKA grants K112112 to T.I.O., PD121287 and FK124661 to A. S., and by the National Brain Research Program (NAP) of Hungary (grant number: 2017-1.2.1-NKP-2017-00002, to Á.A.).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2022.146609.

References

Abascal, F., Zardoya, R., Telford, M.J., 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucl. Acids Res. 38, W7–W13.

Apáti, Á., Orbán, T.I., Varga, N., Németh, A., Schamberger, A., Krizsik, V., Erdélyi-Belle, B., Homolya, L., Várady, G., Padányi, R., Karászi, É., Kemna, E.W.M., Német, K., Sarkadi, B., 2008. High level functional expression of the ABCG2 multidrug transporter in undifferentiated human embryonic stem cells. Biochim. Biophys. Acta 1778 (12), 2700–2709.

Bailey, A.D., Gray, L.T., Pavelitz, T., Newman, J.C., Horibata, K., Tanaka, K., Weiner, A. M., 2012. The conserved Cockayne syndrome B-piggyBac fusion protein (CSB-PGBD3) affects DNA repair and induces both interferon-like and innate antiviral responses in CSB-null cells. DNA Repair (Amst) 11 (5), 488–501.

Baudry, C., Malinsky, S., Restituito, M., Kapusta, A., Rosa, S., Meyer, E., Bétermier, M., 2009. PiggyMac, a domesticated piggyBac transposase involved in programmed genome rearrangements in the ciliate Paramecium tetraurelia. Genes Dev. 23 (21), 2478–2483.

Beckermann, T.M., Luo, W., Wilson, C.M., Veach, R.A., Wilson, M.H., 2021. Cognate restriction of transposition by piggyBac-like proteins. Nucl. Acids Res. 49, 8135–8144.

Bouallegue, M., Rouault, J.D., Hua-Van, A., Makni, M., Capy, P., 2017. Molecular Evolution of piggyBac Superfamily: From Selfishness to Domestication. Genome Biol. Evol. 9, 323–339.

Bunz, F., Dutriaux, A., Lengauer, C., Waldman, T., Zhou, S., Brown, J.P., Sedivy, J.M., Kinzler, K.W., Vogelstein, B., 1998. Requirement for p53 and p21 to sustain G2 arrest after DNA damage. Science 282 (5393), 1497–1501.

Cadinanos, J., Bradley, A., 2007. Generation of an inducible and optimized piggyBac transposon system. Nucl. Acids Res 35, e87.

Chen, Q., Luo, W., Veach, R.A., Hickman, A.B., Wilson, M.H., Dyda, F., 2020. Structural basis of seamless excision and specific targeting by piggyBac transposase. Nat. Commun. 11, 3446.

Cheng, C.-Y., Vogt, A., Mochizuki, K., Yao, M.-C., Bloom, K.S., 2010. A domesticated piggyBac transposase plays key roles in heterochromatin dynamics and DNA cleavage during programmed DNA deletion in Tetrahymena thermophila. Mol. Biol. Cell 21 (10), 1753–1762.

Citterio, E., Van Den Boom, V., Schnitzler, G., Kanaar, R., Bonte, E., Kingston, R.E., Hoeijmakers, J.H.J., Vermeulen, W., 2000. ATP-dependent chromatin remodeling by the Cockayne syndrome B DNA repair-transcription-coupling factor. Mol. Cell. Biol. 20 (20), 7643–7653.

Cordaux, R., Udit, S., Batzer, M.A., Feschotte, C., 2006. Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. Proc. Natl. Acad. Sci. USA 103 (21), 8101–8106.

Cosby, R.L., Chang, N.-C., Feschotte, C., 2019. Host-transposon interactions: conflict, cooperation, and cooption. Genes Dev. 33 (17-18), 1098–1116.

Coufal, N.G., Garcia-Perez, J.L., Peng, G.E., Yeo, G.W., Mu, Y., Lovci, M.T., Morell, M., O'Shea, K.S., Moran, J.V., Gage, F.H., 2009. L1 retrotransposition in human neural progenitor cells. Nature 460 (7259), 1127–1131.

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.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C., Qureshi, M., Richardson, L.J., Salazar, G.A., Smart, A., Sonnhammer, E.L.L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S.C.E., Finn, R.D., 2019. The Pfam protein families database in 2019. Nucl. Acids Res. 47, D427–D432.

Filippi, S., Latini, P., Frontini, M., Palitti, F., Egly, J.-M., Proietti-De-Santis, L., 2008. CSB protein is (a direct target of HIF-1 and) a critical mediator of the hypoxic response. EMBO J. 27 (19), 2545–2556.

Fousteri, M., Mullenders, L.HF., 2008. Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. Cell. Res. 18 (1), 73–84.

Fraser, M.J., Smith, G.E., Summers, M.D., 1983. Acquisition of Host Cell DNA Sequences by Baculoviruses: Relationship Between Host DNA Insertions and FP Mutants of Autographa californica and Galleria mellonella Nuclear Polyhedrosis Viruses. J. Virol. 47 (2), 287–300.

Gray, L.T., Fong, K.K., Pavelitz, T. and Weiner, A.M., 2012. Tethering of the conserved piggyBac transposase fusion protein CSB-PGBD3 to chromosomal AP-1 proteins regulates expression of nearby genes in humans. PLoS Genet, 8, e1002972.

Gray, L.T., Weiner, A.M., 2010. Ubiquitin recognition by the Cockayne syndrome group B protein: binding will set you free. Mol. Cell. 38 (5), 621–622.

Han, M.V., Zmasek, C.M., 2009. phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinf. 10, 356.

Hartl, D.L., Lozovskaya, E.R., Lawrence, J.G., 1992. Nonautonomous transposable elements in prokaryotes and eukaryotes. Genetica 86 (1-3), 47–53.

Helou, L., Beauclair, L., Dardente, H., Arensburger, P., Buisine, N., Jaszczyszyn, Y., Guillou, F., Lecomte, T., Kentsis, A. and Bigot, Y., 2021a. The C-terminal Domain of piggyBac Transposase Is Not Required for DNA Transposition. J. Mol. Biol., 433, 166805.

Helou, L., Beauclair, L., Dardente, H., Piegu, B., Tsakou-Ngouafo, L., Lecomte, T., Kentsis, A., Pontarotti, P., Bigot, Y., 2021b. The piggyBac-derived protein 5 (PGBD5) transposes both the closely and the distantly related piggyBac-like elements Tcr-pble and Ifp2. J. Mol. Biol., 433, 166839.

Henssen, A.G., Henaff, E., Jiang, E., Eisenberg, A.R., Carson, J.R., Villasante, C.M., Ray, M., Still, E., Burns, M., Gandara, J., Feschotte, C., Mason, C.E., Kentsis, A., 2015. Genomic DNA transposition induced by human PGBD5. Elife 4.

Henssen, A.G., Jiang, E., Zhuang, J., Pinello, L., Socci, N.D., Koche, R., Gonen, M., Villasante, C.M., Armstrong, S.A., Bauer, D.E., Weng, Z., Kentsis, A., 2016. Forward genetic screen of human transposase genomic rearrangements. BMC Genomics 17, 548.

Henssen, A.G., Koche, R., Zhuang, J., Jiang, E., Reed, C., Eisenberg, A., Still, E., MacArthur, I.C., Rodríguez-Fos, E., Gonzalez, S., Puiggròs, M., Blackford, A.N., Mason, C.E., de Stanchina, E., Gönen, M., Emde, A.-K., Shah, M., Arora, K., Reeves, C., Socci, N.D., Perlman, E., Antonescu, C.R., Roberts, C.W.M., Steen, H., Mullen, E., Jackson, S.P., Torrents, D., Weng, Z., Armstrong, S.A., Kentsis, A., 2017. PGBD5 promotes site-specific oncogenic mutations in human tumors. Nat Genet 49 (7), 1005–1014.

Hikosaka, A., Kobayashi, T., Saito, Y., Kawahara, A., 2007. Evolution of the Xenopus piggyBac transposon family TxpB: domesticated and untamed strategies of transposon subfamilies. Mol. Biol. Evol. 24, 2648–2656.

Horibata, K., Iwamoto, Y., Kuraoka, I., Jaspers, N.G.J., Kurimasa, A., Oshimura, M., Ichihashi, M., Tanaka, K., 2004. Complete absence of Cockayne syndrome group B gene product gives rise to UV-sensitive syndrome but not Cockayne syndrome. Proc. Natl. Acad. Sci. USA 101 (43), 15410–15415.

Hurst, L.D., 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. 18, 486.

Ivics, Z., 2016. Endogenous Transposase Source in Human Cells Mobilizes piggyBac Transposons. Mol. Ther. 24, 851–854.

Izsvak, Z., Stuwe, E.E., Fiedler, D., Katzer, A., Jeggo, P.A., Ivics, Z., 2004. Healing the wounds inflicted by sleeping beauty transposition by double-strand break repair in mammalian somatic cells. Mol. Cell. 13, 279–290.

Kapitonov, V.V., Jurka, J., 2005. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol. 3, e181.

Kazazian Jr., H.H., 1999. An estimated frequency of endogenous insertional mutations in humans. Nat. Genet. 22, 130.

Keith, J.H., Schaeper, C.A., Fraser, T.S., Fraser Jr., M.J., 2008. Mutational analysis of highly conserved aspartate residues essential to the catalytic core of the piggyBac transposase. BMC Mol. Biol. 9, 73.

Kidwell, M.G., 2002. Transposable elements and the evolution of genome size in eukaryotes. Genetica 115, 49–63.

Kolacsek, O., Erdei, Z., Apati, A., Sandor, S., Izsvak, Z., Ivics, Z., Sarkadi, B., Orban, T.I., 2014. Excision efficiency is not strongly coupled to transgenic rate: cell type-dependent transposition efficiency of sleeping beauty and piggyBac DNA transposons. Hum. Gene .Ther. Methods 25, 241–252.

Kolacsek, O., Orban, T.I., 2018. Transcription activity of transposon sequence limits Sleeping Beauty transposition. Gene 676, 184–188.

Kolacsek, O., Pergel, E., Varga, N., Apati, A., Orban, T.I., 2017. Ct shift: A novel and accurate real-time PCR quantification model for direct comparison of different nucleic acid sequences and its application for transposon quantifications. Gene 598, 43–49.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K.,

Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A. T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Li, C., Wen, A., Shen, B., Lu, J., Huang, Y., Chang, Y., 2011. FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. BMC Biotechnol. 11, 92.

Luan, Y., Suzuki, T., Palanisamy, R., Takashima, Y., Sakamoto, H., Sakuraba, M., Koizumi, T., Saito, M., Matsufuji, H., Yamagata, K., Yamaguchi, T., Hayashi, M., Honma, M., 2007. Potassium bromate treatment predominantly causes large deletions, but not GC>TA transversion in human cells. Mutat. Res. 619, 113–123.

Mabbott, N.A., Baillie, J.K., Brown, H., Freeman, T.C., Hume, D.A., 2013. An expression atlas of human primary cells: inference of gene function from coexpression networks. BMC Genomics 14, 632.

Majumdar, S., Singh, A., Rio, D.C., 2013. The human THAP9 gene encodes an active P-element DNA transposase. Science 339, 446–448.

Mandal, P.K., Kazazian Jr., H.H., 2008. SnapShot: Vertebrate transposons. Cell 135 (192–192), e1.

Manuri, P.V., Wilson, M.H., Maiti, S.N., Mi, T., Singh, H., Olivares, S., Dawson, M.J., Huls, H., Lee, D.A., Rao, P.H., Kaminski, J.M., Nakazawa, Y., Gottschalk, S., Kebriaei, P., Shpall, E.J., Champlin, R.E., Cooper, L.J., 2010. piggyBac transposon/transposase system to generate CD19-specific T cells for the treatment of B-lineage malignancies. Hum. Gene. Ther. 21, 427–437.

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. Nat. Genet. 41, 753–761.

Mitra, R., Fain-Thornton, J., Craig, N.L., 2008. piggyBac can bypass DNA synthesis during cut and paste transposition. EMBO J. 27, 1097–1109.

Mitra, R., Li, X., Kapusta, A., Mayhew, D., Mitra, R.D., Feschotte, C., Craig, N.L., 2013. Functional characterization of piggyBat from the bat Myotis lucifugus unveils an active mammalian DNA transposon. Proc. Natl. Acad. Sci. USA 110, 234–239.

Morellet, N., Li, X., Wieninger, S.A., Taylor, J.L., Bischerour, J., Moriau, S., Lescop, E., Bardiaux, B., Mathy, N., Assrir, N., Betermier, M., Nilges, M., Hickman, A.B., Dyda, F., Craig, N.L., Guittet, E., 2018. Sequence-specific DNA binding activity of the cross-brace zinc finger motif of the piggyBac transposase. Nucl. Acids Res. 46, 2660–2677.

Newman, J.C., Bailey, A.D., Fan, H.Y., Pavelitz, T., Weiner, A.M., 2008. An abundant evolutionarily conserved CSB-PiggyBac fusion protein expressed in Cockayne syndrome. PLoS Genet 4, e1000031.

Newman, J.C., Bailey, A.D., Weiner, A.M., 2006. Cockayne syndrome group B protein (CSB) plays a general role in chromatin maintenance and remodeling. Proc. Natl. Acad. Sci. USA 103, 9613–9618.

Pagan, H.J., Smith, J.D., Hubley, R.M., Ray, D.A., 2010. PiggyBac-ing on a primate genome: novel elements, recent activity and horizontal transfer. Genome Biol. Evol. 2, 293–303.

Pan, Y., Oprysko, P.R., Asham, A.M., Koch, C.J., Simon, M.C., 2004. p53 cannot be induced by hypoxia alone but responds to the hypoxic microenvironment. Oncogene 23, 4975–4983.

Pavelitz, T., Gray, L.T., Padilla, S.L., Bailey, A.D., Weiner, A.M., 2013. PGBD5: a neural-specific intron-containing piggyBac transposase domesticated over 500 million years ago and conserved from cephalochordates to humans. Mobile DNA 4, 23.

Pires, I.M., Bencokova, Z., Milani, M., Folkes, L.K., Li, J.L., Stratford, M.R., Harris, A.L., Hammond, E.M., 2010. Effects of acute versus chronic hypoxia on DNA damage responses and genomic instability. Cancer Res. 70, 925–935.

Pritham, E.J., Feschotte, C., Wessler, S.R., 2005. Unexpected diversity and differential success of DNA transposons in four species of entamoeba protozoans. Mol. Biol. Evol. 22, 1751–1763.

Raskó, T., Pande, A., Radscheit, K., Zink, A., Singh, M., Sommer, C., Wachtl, G., Kolacsek, O., Inak, G., Szvetnik, A., Petrakis, S., Bunse, M., Bansal, V., Selbach, M., Orbán, T.I., Prigione, A., Hurst, L.D., Izsvák, Z., 2021. PiggyBac Transposable Element-derived 1 controls Neuronal Progenitor Identity, Stress Sensing and mammal-specific paraspeckles. bioRxiv. https://doi.org/10.1101/2021.05.19.444448v1.

Saha, S., Woodard, L.E., Charron, E.M., Welch, R.C., Rooney, C.M., Wilson, M.H., 2015. Evaluating the potential for undesired genomic effects of the piggyBac transposon system in human cells. Nucl. Acids Res. 43, 1770–1782.

Sarkar, A., Sim, C., Hong, Y.S., Hogan, J.R., Fraser, M.J., Robertson, H.M., Collins, F.H., 2003. Molecular evolutionary analysis of the widespread piggyBac transposon family and related "domesticated" sequences. Mol. Genet. Genomics 270, 173–180.

Sarker, A.H., Tsutakawa, S.E., Kostek, S., Ng, C., Shin, D.S., Peris, M., Campeau, E., Tainer, J.A., Nogales, E., Cooper, P.K., 2005. Recognition of RNA polymerase II and transcription bubbles by XPG, CSB, and TFIIH: insights for transcription-coupled repair and Cockayne Syndrome. Mol. Cell. 20, 187–198.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539.

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 Neurosci. 33, 345–354.

Varga, N., Vereb, Z., Rajnavolgyi, E., Nemet, K., Uher, F., Sarkadi, B., Apati, A., 2011. Mesenchymal stem cell like (MSCl) cells generated from human embryonic stem cells support pluripotent cell growth. Biochem. Biophys. Res. Commun. 414, 474–480.

Velez-Cruz, R., Egly, J.M., 2013. Cockayne syndrome group B (CSB) protein: at the crossroads of transcriptional networks. Mech. Ageing Dev. 134, 234–242.

Vofely, G., Berecz, T., Szabo, E., Szebenyi, K., Hathy, E., Orban, T.I., Sarkadi, B., Homolya, L., Marchetto, M.C., Rethelyi, J.M., Apati, A., 2018. Characterization of calcium signals in human induced pluripotent stem cell-derived dentate gyrus neuronal progenitors and mature neurons, stably expressing an advanced calcium indicator protein. Mol. Cell. Neurosci. 88, 222–230.

Walisko, O., Izsvak, Z., Szabo, K., Kaufman, C.D., Herold, S., Ivics, Z., 2006. Sleeping Beauty transposase modulates cell-cycle progression through interaction with Miz-1. Proc. Natl. Acad. Sci. USA 103, 4062–4067.

Weiner, A.M., Gray, L.T., 2013. What role (if any) does the highly conserved CSB-PGBD3 fusion protein play in Cockayne syndrome? Mech. Ageing Dev. 134, 225–233.

Wu, C., Macleod, I., Su, A.I., 2013. BioGPS and MyGene.info: organizing online, gene-centric information. Nucl. Acids Res. 41, D561–D565.

Yusa, K., Zhou, L., Li, M.A., Bradley, A., Craig, N.L., 2011. A hyperactive piggyBac transposase for mammalian applications. Proc. Natl. Acad. Sci. USA 108, 1531–1536.

Zayed, H., Izsvak, Z., Khare, D., Heinemann, U., Ivics, Z., 2003. The DNA-bending protein HMGB1 is a cellular cofactor of Sleeping Beauty transposition. Nucl. Acids Res. 31, 2313–2322.

Zhang, Z., Li, J., Zhao, X.Q., Wang, J., Wong, G.K., Yu, J., 2006. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genom. Proteom. Bioinform. 4, 259–263.