Potentially active DNA transposons in the human genome: genomic parasites or domesticated residents?

Aims of the project

By investigating *piggyBac*-derived elements (PGBDs) as examples, we examined the evolutionary fate of DNA transposons in the human genome. We studied the potential preservation of transposition ability, addressed the process of domestication, and also investigated genome defense mechanisms against transposons as invasive genetic elements. Apart from providing new insights on our genetic regulations and evolutionary inheritance, we believe that the results have strong influence on future gene therapy applications, since the insect derived *piggyBac* transposon system is widely used as a gene delivery method.

Investigating the preservation of "canonical" DNA transposition

By genomic sequence annotations, approximately 2000 *piggyBac*-related elements are predicted in the human genome, the majority of which are repetitive sequences derived from the inverted repeat elements of putative ancient transposons. There are only 5 protein coding sequences that show evolutionary relationship to *piggyBac*-like DNA transposase proteins named PGBD1-5. We established plasmid vectors expressing all 5 of these human PGBD elements, and checked that their overexpression is not toxic to the reporter cell lines (HEK-293, HeLa) commonly used in our laboratory. Since PGBD1 is predicted to be a fusion protein, we expressed and tested this fusion version (previously named as "HUCEP4"), as well as the C-terminal transposase like domain encoded by the last independent exon. By bioinformatics approach, we attempted to search for potential target sequences of the PGBD elements but could not identify ones other than those predicted previously: the MER85 sequences (predicted substrates of PGBD3) and the MER75 sequences (predicted substrates of PGBD4). These latter sequences seem to be intact enough to act as a "canonical" inverted repeat sequences and potentially participate in DNA transposition reactions. However, they are present in several copies in the human genome so in both cases, we chose one consensus sequence with the least number of mutations, and used those for further testing.

Having all sequences identified and cloned, we performed transposition assays by systematically combining and transfecting one of the putative PGBD transposases together with a vector containing a puromycin resistance gene cassette flanked by one of the assumed inverted repeat sequences. As a positive control reaction, we also performed transposition assays with the insect piggyBac system: a plasmid expressing the transposase co-transfected with a vector containing a puromycin resistance expression cassette flanked by its terminal inverted repeat sequences. In addition, we combined and studied the insect system with all PGBDs, transposases and MER sequences alike. We tested the occurrence of transposition at two levels: 1) at the first step of transposon excision from the plasmids; and 2) at the

integration of the transgene into the host cell's genome by providing antibiotic resistance (at a level higher than random integration). We could not detect transposition ability in any of the PGBD genes; moreover, no cross reactivity was detected between the insect transposon system and any of the PGBD genes or MER sequences. In the case of PGBD3, however, sequence analysis revealed that the conserved "DDE" catalytic motif of the putative transposase is mutated which could be the reason why the protein is inherently inactive. By performing a mutagenesis on the expression vector, we changed the amino acids "back" to the potentially active ones, and repeated the transposition assays. However, no "canonical" transposition activity was detected in this case either. We concluded that none of the human PGBD genes preserved DNA transposition ability and as an important aspect, the lack of cross mobilization with the insect piggyBac transposon confirmed that the insect system is a safe gene delivery method in human cells and fully applicable in future human gene therapy applications.

We were about to publish our results this time, when the *piggyBac* transposition was a very hot topic in the scientific literature and several papers were published with contradictory result. Although not testing all PGBDs systematically, the results by Saha et al. were in agreement with our findings and they described no cross-reactivity with the endogenous elements and the insect *piggyBac* system (Saha et al., 2015). On the other hand, the Kentsis-lab showed in their publication that the human PGBD5 is active and shows cross reaction with the insect PB system used for gene delivery in human cell lines (Henssen et al., 2015). We wanted to verify these results but in spite of asking for and using the same expression plasmids as they had, we could not reproduce PGBD5 activity: we could not detect either excision activity or transposon mediated integration by PGBD5 in human cell lines. Discussing this issue in several conferences, and with various research groups, it turned out that no research groups could reproduce the results. Later on, the Kentsis-group "tuned down" their results in their subsequent publications, stating that PGBD5 is a low activity non-canonical transposase, having the "DDD" catalytic protein motif by unusually positioned amino acids, but can initiate genomic rearrangements. However, the reproducibility problem still remained. Collaborating with Zsuzsanna Izsvák's and Zoltán Ivics's research group, we revealed that there was an annotation problem of PGBD5 by the Kentsis-group but that did not explain the putative false positive results. We are currently discussing the problem of how to publish our results, most likely in collaboration with other groups, and incorporating all other results on domestication studies, as discussed later (manuscript in preparation).

Investigating DNA transposition mechanisms

In the meantime, we continued our research on the mechanistic insights of DNA transposition, testing various transposon systems (*piggyBac* or *Sleeping Beauty*) in several cell lines and model organisms (e.g. rats). We used the DNA transposon based gene delivery method to establish homozygous transgenic rat strains expressing calcium sensor proteins, optimizing the method for the most efficient conditions to create low copy integrations (Szebenyi et al., 2015). We also optimized the transposon based delivery for application in

human embryonic stem cells: we generated stem cells overexpressing the ABCG2 multidrug transporter which cells could thereby differentiate towards various multidrug resistant tissue types (Erdei et al., 2018). In addition, we also utilized DNA transposons for neuron specific research: we established human cell lines expressing GABA transporters (Ma et al., 2016) and we generated calcium indicator (GCaMP6) expressing neuronal progenitors from genetically modified human induced pluripotent stem cells (Vofely et al., 2018). DNA transposons were used to create stable cell lines expressing ABCG2 noncoding exon variants for gene regulation studies (Sandor et al., 2016), and some of this knowledge was also summarized in a review article (Apati et al., 2016).

The most important new finding on the molecular mechanisms of transposition was that promoter usage and thereby the expression intensity of the transgene can limit the excision efficiency of certain DNA transposons. We could provide evidence that transgenes regulated with high activity promoters can reduce the overall success of gene delivery and our experiments revealed inverse correlation between transcription activity and the SB100X hyperactive *Sleeping Beauty* variant's excision efficiency. This competition between transcription and transposition was independent of the transgene coding sequence and did not alter the transgenic efficiency in general. However, promoters applied in the transgene cassette can produce different average copy numbers depending on the transcriptional activity of the transposon. Unlike the insect *piggyBac* transposon system, where this phenomenon could not be detected, this mechanism allows a fine balance of expression using the high copy potential *Sleeping Beauty* system that adjusts the copy number of lower activity promoter driven transgenes to a higher expression level. All this contributes to a well-tolerated transgenesis at the cellular level, and would be important to consider in gene therapy applications (Kolacsek and Orban, 2018).

Evidence for transposon domestication of human PGBD genes

According to the ENCODE data, to several expression databases, and to our own investigations, all human PGBD genes are transcribed in various tissues. Following our previous results showing that they are not capable of "canonical" DNA transposition, these observations pointed towards domesticated cellular functions. To start the functional characterization of the PGBD elements, we established and performed a systematic real-time PCR based expression profiling of all human PGBD genes, including their predicted and potential splice variants. We could see indication of distinct cellular functions, as these profiles are markedly different in various tissue types, including embryonic and mesenchymal stem cells. In addition, we could identify a complex splicing pattern of the human PGBD3 gene which is located as a "natural exon trap" in the 5th intron of the human Cockayne Syndrome B gene (Kolacsek et al., 2017). Considering the fact that the PGBD3 transcription unit is flanked by two surprisingly intact MER85 sequences, it raised the possibility that PGBD3 might be involved in splicing regulation.

Based on these previous findings, by using various artificial splicing reporter constructs, we addressed the question whether PGBD3 is involved in its own splicing autoregulation. The

catalytic "DDD" motif of the human PGBD3 is mutated, so it cannot initiate the endonucleolytic cleavage required for transposition, but it can still bind to its putative ancient IRDR motif, MER85 (Newman et al., 2008). Our results with the splicing reporters so far produced conflicting results: when PGBD3 was expressed, some flanking sequence containing GFP-based reporters showed enhanced splicing efficiency but it seemed to vary according to the PGBD3 expression level. Moreover, some strong artificial promoters seem to work better than the others, suggesting that there might also be some interference between the transcription apparatus and PGBD3, although previously we could detect such phenomenon in case of the mariner type DNA transposon, *Sleeping Beauty*, but not for *piggyBac*-related mobile elements (Kolacsek and Orban, 2018). Other experiments with the PGBD4 protein showed that it has an influence on splicing efficiency in case of GFP reporters, and it might regulate both constitutive and alternative splicing. These examples indicate that some human DNA transposons might have been "tamed" and domesticated for splicing regulation; nevertheless, further experiments are needed to clarify this issue (manuscript in preparation).

To further examine the potential functions of human PGBDs, we focused our efforts on PGBD1 which is expressed as a fusion protein ("HUCEP4") in the human genome. It is transcribed in various tissues, and its loss of function seems to be associated with human diseases: various SNPs in PGBD1 are described to be associated with certain neurodegenerative diseases or hemochromatosis. To address these issues, we first studied a PGBD1 SNP (rs1997660 A/G), located in the last exon encoding for the transposase-derived domain. We could show that the frequency of one allele (A) is significantly higher in Hungarian patients with hereditary hemochromatosis, similarly to that described in other (e.g.: Portuguese) populations (Cruz et al., 2008). We could also confirm that this variant is part of a longer haplotype sequence, and analyzing the clinical parameters of carrier and non-carrier patients, the data suggests that the 'A' allele containing haplotype is associated with a more severe phenotype. Some further analysis is required to confirm this finding but it seems that this allele could serve as a biomarker in assessing the disease phenotype. In addition, as PGBD1 is expressed in blood cells, we are currently investigating what function this gene might have in different leukocytes (manuscript in preparation).

Genomic defense against DNA transposons: some aspects of RNA interference mechanisms

In the final part of our research, we investigated if there are any evolutionary footprints in the human genome indicating cellular defense mechanisms against DNA transposons as invasive genetic elements. Several pathways of RNA interference have been proven to be involved in such protective regulation: the Piwi-piRNA system especially in the germline, the exogenous and endogenous siRNA pathways in somatic cells, but also some microRNAs have been shown to regulate transposable elements (Golden et al., 2008; Shalgi et al., 2010; Ishizu et al., 2012). Searching in small RNA sequencing databases, especially from data on Ago and Piwi protein immunoprecipitation experiments, we could identify some piRNA sequences corresponding to PGBD genes. Trying to validate these results, we performed knock down

experiments of some of the four human Piwi proteins, but no effect were seen on expression of PGBD genes. However, several issues are still need to be clarified (e.g. which human Piwis are expressed in somatic tissues) and the experiments need to be carefully adjusted and repeated in various cell lines before valid conclusions can be drawn in this respect.

On the other hand, our investigations on the microRNA pathways revealed new insights. We examined human embryonic stem cells and placenta cells, as they are shown to control transposable elements strictly, and investigated their microRNA profiles to see if we could reveal any connection with transposon control. Our focus turned to a long microRNA cluster on human chromosome 19, which is expressed exclusively in stem cells and in placenta, and which locus also have several transposon-derived sequences but mainly originating from Alu elements, so from retrotransposons. We are currently investigating the complex regulation of that region which is planned to be published later; nevertheless, some microRNA results turned out to be medically relevant issues. When examining human placenta samples and related cell lines, the level of hsa-miR-210 was significantly higher in preeclampsia placentas, which could cause a minor increase of exosomal and a high elevation of Ago-bound miR-210 in circulation. Hypoxia lead to intracellular hsa-miR-210 upregulation in trophoblast cell lines: in extravillous cell (HTR-8) media, only the level of exosomal hsa-miR-210 was increased but no change in Ago-bound hsa-miR-210 level was observed. In contrast, in villous cell (JAR) media, the level of exosomal hsa-miR-210 was increased and enhanced release of Ago-bound hsa-miR-210 was also observed. Based on our data, we postulate that in preeclampsia, exosomal hsa-miR-210 is secreted actively from the trophoblast, and by intercellular communication, it may have a role in disease etiology. In addition, there is a passive release of Ago-bound hsa-miR-210 into the circulation, which may represent by-products of cell-death and is thereby a possible consequence of the disease (Biro et al., 2019).

Dissemination of our scientific results and other aspects of our research

With the financial support of this NKFIH-OTKA project, two PhD dissertations were submitted and successfully defended during the research period. My former PhD student, Orsolya Kolacsek defended her thesis on DNA transposon regulation in mammalian cells in 2016, and she is currently working in my laboratory as a postdoctoral fellow. My other PhD student, Sára Sándor defended her thesis in 2017 and she is currently doing a postdoctoral research in another group in Eötvös Loránd University. She was working on gene regulatory roles of non-coding exons but she used the transposon-based gene delivery system to establish reporter cell lines.

We also disseminated our research results in form of oral and poster presentations in several national and international conferences and meetings, including the International Congress on Transposable Elements (ICTE 2016) in Saint Malo, France, and transposon conferences in Heidelberg, Germany, and in Santa Fe, New Mexico, USA. Tamás Orbán also gave a promoting lecture on DNA transposons in the “Élő Adás” lecture series in Eötvös Loránd University, Budapest.
(https://www.youtube.com/watch?v=XNrkYB6_enY&feature=youtu.be).

References

(publications in **bold** were financed from this research project)

Apati, A., Szebenyi, K., Erdei, Z., Varady, G., Orban, T.I. and Sarkadi, B., 2016. The importance of drug transporters in human pluripotent stem cells and in early tissue differentiation. Expert Opin Drug Metab Toxicol 12, 77-92.

Biro, O., Fothi, A., Alasztics, B., Nagy, B., Orban, T.I. and Rigo, J., Jr., 2019. Circulating exosomal and Argonaute-bound microRNAs in preeclampsia. Gene 692, 138-144.

Cruz, E., Whittington, C., Krikler, S.H., Mascarenhas, C., Lacerda, R., Vieira, J. and Porto, G., 2008. A new 500 kb haplotype associated with high CD8+ T-lymphocyte numbers predicts a less severe expression of hereditary hemochromatosis. BMC Med Genet 9, 97.

Erdei, Z., Schamberger, A., Torok, G., Szebenyi, K., Varady, G., Orban, T.I., Homolya, L., Sarkadi, B. and Apati, A., 2018. Generation of multidrug resistant human tissues by overexpression of the ABCG2 multidrug transporter in embryonic stem cells. PLoS One 13, e0194925.

Golden, D.E., Gerbasi, V.R. and Sontheimer, E.J., 2008. An inside job for siRNAs. Molecular cell 31, 309-12.

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. and Kentsis, A., 2015. Genomic DNA transposition induced by human PGBD5. Elife 4.

Ishizu, H., Siomi, H. and Siomi, M.C., 2012. Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. Genes & development 26, 2361-73.

Kolacsek, O. and 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. and 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.

Ma, X., Lubin, H., Ioja, E., Kekesi, O., Simon, A., Apati, A., Orban, T.I., Heja, L., Kardos, J. and Marko, I.E., 2016. Straightforward and effective synthesis of gamma-aminobutyric acid transporter subtype 2-selective acyl-substituted azaspiro[4.5]decanes. Bioorg Med Chem Lett 26, 417-423.

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

Saha, S., Woodard, L.E., Charron, E.M., Welch, R.C., Rooney, C.M. and Wilson, M.H., 2015. Evaluating the potential for undesired genomic effects of the piggyBac transposon system in human cells. Nucleic Acids Res 43, 1770-82.

Sandor, S., Jordanidisz, T., Schamberger, A., Varady, G., Erdei, Z., Apati, A., Sarkadi, B. and Orban, T.I., 2016. Functional characterization of the ABCG2 5' non-coding exon variants: Stem cell specificity, translation efficiency and the influence of drug selection. Biochim Biophys Acta 1859, 943-51.

Shalgi, R., Pilpel, Y. and Oren, M., 2010. Repression of transposable-elements - a microRNA anti-cancer defense mechanism? Trends in genetics : TIG 26, 253-9.

Szebenyi, K., Furedi, A., Kolacsek, O., Pergel, E., Bosze, Z., Bender, B., Vajdovich, P., Tovari, J., Homolya, L., Szakacs, G., Heja, L., Enyedi, A., Sarkadi, B., Apati, A. and Orban, T.I.,

2015. Generation of a Homozygous Transgenic Rat Strain Stably Expressing a Calcium Sensor Protein for Direct Examination of Calcium Signaling. Sci Rep 5, 12645.

Vofely, G., Berecz, T., Szabo, E., Szebenyi, K., Hathy, E., Orban, T.I., Sarkadi, B., Homolya, L., Marchetto, M.C., Rethelyi, J.M. and 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.