The role of dyskerin mediated pseudouridylation in development and disease

NKFIH-FK124230 – Final report

Pseudouridylation, the isomerisation of uridine into pseudouridine (pseU), is one of the most abundant post-transcriptional modifications in the eukaryotic cell. The process of pseudouridylation is catalyzed by pseudouridine synthases (PUSs) which can target either consensus sequences in the transcriptome (stand-alone PUSs) or can act in a guided manner, like the enzyme dyskerin, which uses small nucleolar RNAs (snorRNAs) to find its target sites. Before the beginning of this project, as part of a collaborative work to reveal the etiology of a novel dyskerin-related human syndrome, we created an allelic series for *dkc1*, the zebrafish ortholog of the gene that encodes dyskerin.

In the original Research Plan for the project we set out to achieve two major aims: 1.) to characterize the phenotype of these *dkc1* mutant lines and 2.) to observe and describe the changing patterns of pseudouridylation of zebrafish RNAs during development. While the COVID-19 pandemic and related restrictions made our research much more difficult (e.g. international collaborations became much more challenging and sometimes grinded to a halt, see below), ultimately we were able to achieve these goals with minor delays and modifications.

Characterization of *dkc1* mutant zebrafish lines and pathogenic DKC1 and NOP10 variants

In the first two years of the project we completed the characterization of two null-alleles of *dkc1* (*elu1* and *elu8*). In these studies we have complemented histological analysis of 4 days post fertilization (dpf) homozygous mutant larvae (in which we already see a strong phenotype) with whole-transcriptome analysis of 1.5 dpf mutants (in which morphologically differences are already present). Our results revealed the similarity between the phenotypic characteristics of the novel human syndrome and that of the zebrafish mutants: cataracts, smaller eyes, underdeveloped pineal gland and kidney (Balogh, Chandler, Varga et al., 2020). We also demonstrated that homozygous *dkc1* mutant zebrafish have impaired differentiation in multiple tissues. The interpretation of these results was greatly aided by the gene ontology (GO) analysis of the transcriptomic datasets that revealed the mutant phenotype as a *de facto* ribosomopathy, where the disruption of dyskerin function resulted in erroneous ribosome biogenesis (Balogh, Chandler, Varga et al., 2020). This conclusion is also supported by some of our unpublished, preliminary observations showing that the ribosomes of the mutants are not able to interpret internal ribosomal entry site (IRES) sequences correctly.

Our transcriptome analysis also shed light on a significant earlier observation: histochemical analysis of the mutants suggested that while cells in many tissues are stuck in the cell cycle and fail to undergo differentiation, surprisingly, they do not undergo programmed cell death. In several other ribosomopathies an upregulation of Tp53 can be observed due to the dysfunction of the translational machinery. Accordingly, *dkc1*(*elu1*) null-mutants show an upregulation in the number of *tp53* transcripts. However, instead of the canonical, full-length, pro-apoptotic isoform, a shorter, anti-apoptotic *Δ113tp53* isoform is increased in abundance. Importantly, lack of *tp53* by itself is not able to rescue the hallmarks of the *dkc1* loss-of-function phenotype, such as early lethality, smaller eyes and tectum, shorter body axis, as $dkc1^{-/-}$, $tp53^{-/-}$ double mutants also display these phenotypic marks. This finding also suggests that upregulation of the *Δ113tp53* isoform is not causative to the observed impairment in differentiation (Balogh, Chandler, Varga et al., 2020).

As the set of the symptoms shown by the patients carrying the X-linked *DKC1(E206K)* allele is significantly different from those previously described for Dyskeratosis Congenita (DC), the syndrome generally associated with dyskerin-deficiency, we were pleasantly surprised to find out that another group at University College London Great Ormond Street Institute of Child Health has identified patients with similar symptoms carrying mutations in *NOP10*, a gene that encodes one of the main interaction partners of dyskerin. Furthermore, this novel mutation, NOP10(T16M), happened to be at a position that is situated in a complementary position to our mutation, DKC1(E206K), in the PUS ribonucleoprotein complex. As part of an international collaborative work, members of our group – Dr. Gusztáv Scháy (SE) and Dr . Eszter Balogh (SE) – performed *in vitro* pressure perturbance analysis on the human DKC1(E206K) allele and the NOP10(T16M) mutant isoforms. Their results show that the interaction between dyskerin and its binding partner NOP10 is impaired in both cases and, more importantly, structural modelling shows that these mutations cause slight but significant alterations in the active centre of dyskerin. These results have been published in the *Proceedings of the National Academy of Sciences* (D1, IF: 11.2), and I am both a co-first and co-corresponding author on the manuscript (Balogh, Chandler, Varga et al., 2020).

In parallel with our in-depth analysis of the *dkc1* null-alleles *elu1* and *elu8* we also started to analyse a hypomorphic allele of the gene, *elu2.* The *dkc1(elu2)* allele carries a Val inserted in the catalytic domain and unlike the null-mutants is not larval-lethal. Interestingly, homozygous *dkc1*$^{elu2/elu2}$ fish are not only much smaller and pigmentation deficient (hinting at a possible disruption of the iridiophore biogenesis), but are also present in non-mendelian ratios (5-10% instead of 25%) among the adult offspring of heterozygous carriers. Later analysis suggested that these homozygous fish are also extremely sensitive for overcrowding and have significantly shorter life-span, partly because they appear to be prone to malignant transformations, especially melanomas. Perhaps most unexpectedly, the behavior of mutants is also altered: in an open-tank paradigm they show only extreme behavioral types (they spend their time either at the bottom or at the top of the tank.)

We wanted to understand if some of these phenotypic effects are due to excess apoptosis, therefore, we crossed these fish into a *tp53* loss-of-function background. So far we were able to identify only a single *dkc1*$^{elu2/elu2}$*; tp53*$^{-/-}$ double mutant, but interestingly, both the size and the pigmentation defects have been rescued to some extent in this individual.

The structural analysis of this p.Leu189_Arg190insVal insertional mutant, which was performed by Dóra K. Menyhárd (ELTE), suggested that the catalytic pocket of the enzyme is altered significantly due to the Val insertion. This change most likely changes the efficiency of pseudouridylation, which might provide a structural explanation for the observed phenotypes. We are performing experiments to test this prediction and once those are finished we will publish our observations related to the hypomorphic disease model in a specialized journal. (We have already presented these data at several meetings: 11th European Zebrafish Meeting (2020), Developmental Disorders: From Mechanism to Treatment (2021), IV. Annual Meeting of Cell, Developmental and Stem Cell Biologists in Hungary (2021).)

Detection of pseU sites in zebrafish samples

In accordance with the original Research Plan, we also made significant efforts to characterize the dynamics of pseudouridylation during development. We started two independent (complementary) collaborations to achieve this aim, through somewhat different approaches.

With the group of Wendy Gilbert at Yale University we wanted to adapt their original methodology used to detect pseU sites in yeast and human cells (Carlile et al. 2014 Nature) to zebrafish samples. In parallel we also established another collaboration with Oxford Nanopore Technologies (ONT) to use their innovative technology to detect pseudouridylation in zebrafish total RNA samples.

The "classic" methodology, which uses carbodiimide (CMC) to detect pseU, failed during the library preparation step for unknown reasons, unfortunately, and the Gilbert group was not able to overcome this problem. Therefore, we decided to use a more traditional, targeted approach to detect particular, conserved pseU sites taking advantage of our *dkc1* null mutant strains. Using the snoGPS script (Schattner et al., 2005 *NAR*) we screened the zygotic zebrafish 18S rRNA for unique sites (i.e. thus not present in the maternal rRNA, see below) that could be recognized by the H/ACA-box snoRNA guides annotated in the zebrafish genome. We identified two such sites, 'U661' and 'U937', with orthologous positions in the human 18S rRNA and six further sites, which do not have human orthologs, but could still be *bona fide* targets of zebrafish snoRNAs ('U259', 'U1745', 'U1747', 'U1320', 'U895', 'U1032'). Our short-term goal was to design specific primers for the 'U661' and 'U937' sites and test these specific positions for pseudouridylation signal. I planned to take advantage of a Fulbright Fellowship to visit the Gilbert lab in early 2020 to do these experiments, but due to the COVID-19 pandemic and related regulations I had to postpone this visit till January 2022.

We also had high hopes from the very beginning for the nanopore-based detection of pseU in the transcriptome. The principles of such analysis has been established before (Garalde et al. 2018 *Nat Meth*) and the first experiments performed by the Nanopore team comparing rRNA from *dkc1* null-mutant and wild type fish gave promising results (i.e. they could train their algorithms to detect pseU sites). Unfortunately, just as we were starting to get promising results, the cooperation with ONT hit a dead-end, as with the pandemic colleagues at ONT have concentrated all their efforts on viral sequencing.

Luckily, however, at the end of 2020 we were also able to initiate a new collaboration with Eva Maria Novoa Pardo's group, located at the Centre for Genomic Research (CRG) in Barcelona, Spain. Dr. Novoa was creating algorithms that can reliably detect posttranscriptional RNA modifications using nanopore sequencing (Liu et al., 2019 *Nat Comm*) and their new computational pipelines were finetuned to detect pseudouridylation events in the sequencing datasets (Begik et al., 2021 *Nat Biotech*).

Analysing 1.5 dpf and 5 dpf samples from both our null-mutant fish and their wild type counterparts the Novoa group was able to identify conserved and species-specific pseU sites in both the zygotic and the maternal rRNAs with high confidence. Using two computational pipelines, with different algorithms and stringency, they were able to detect reliably 32 pseU sites in the 18S rRNA of zebrafish. Of these sites 19 are conserved in human rRNA as well, and, more interestingly, only 9 are present at both developmental stages. We planned to expand these studies to other RNA species as well, however pandemic-related personnel restrictions at CRG made this collaborative work tedious during the past year. We hope that as the pandemic situation becomes more tenable we can continue the collaboration at a higher frequency. Nevertheless, interestingly, the initial results of the Novoa group confirmed a change in the dynamics between maternal and zygotic rRNA species in our null-mutants larvae.

Zebrafish preload the eggs with maternal rRNA, which later gets replaced with zygotic rRNA. Interestingly, these two rRNA species are not identical and are transcribed from different loci

(Locati et al., 2017 *RNA*). Originally we hypothesized that *dkc1* null mutants degrade the maternal rRNA, but are unable to process the zygotic pre-rRNA in the absence of Dkc1-function. Interestingly, when we tested this idea using RT-PCR we found that mutants indeed contain very little zygotic rRNA, but the maternal rRNA is not degraded in the observed timeframe. The nanopore sequencing data provided an independent confirmation to this observation. We would like to pursue this analysis later as these observations give further proof to the "specialized ribosome" hypothesis, i.e. the translation of some mRNAs could be specific for the zygotic ribosome as the maternal ribosome is unable to process them, hence the phenotype.

Development of a zebrafish-specific snoRNAome database

When we realized that the detection of pseU in the transcriptome is more challenging than we initially thought, we also started to pursue a more indirect method as well, to reveal the dynamics of pseudouridylation during zebrafish development. With the help of a talented PhD student, Renáta Hamar, we started an *in silico* approach to map the zebrafish "snoRNAome" using both pre-existing and newly created small-RNA and transcriptomic datasets,

Pseudouridylation by the dyskerin enzyme is guided by H/ACA-box snoRNAs, therefore, identifying these snoRNAs and studying their expression during development would give us a better understanding of the dynamics of pseudouridylation. Initially we have searched the NCBI database for raw sequencing datasets that included the snoRNA populations. Our initial results suggested that there were several snoRNAs with dynamic expression during development. Furthermore, the analysis of datasets from female and male adult zebrafish suggested the presence of not only tissue-specific snoRNA-expression profiles, but also sex-dependent expression of certain snoRNAs even in non-gonadal male and female tissues (e.g. gut, liver and brain). Sequencing the snoRNA fraction (60-200 nt range) in multiple new transcriptomic datasets we were able to confirm these differences. Our targeted sequencing results, combined with some previously analysed datasets, have also offered us the possibility to assemble the zebrafish snoRNAome. Similar datasets have been published for other species, but for zebrafish only limited data was available and several lines of evidence (e.g. number of annotated snoRNA genes vs. full genome size) suggested that many zebrafish snoRNAs have not been identified before. Even using stringent detection criteria, we were able to increase to number of the known snoRNA genes from 247 to 314 and with more relaxed criteria we identified about 1500 other, putative snoRNA genes. As we isolated total RNA from multiple tissues for our studies, we are also able to describe tissue-specific expression for many of these new snoRNAs as well.

Our work has resulted in a valuable database of the zebrafish snoRNAome, where the specific expression patterns and putative targets of specific snoRNAs were described (https://renata-h.shinyapps.io/own-database/). As similar datasets have been published for other species in high impact journals (e.g. Jorjani et al., 2016 *NAR*; Canzler et al., 2018 *RNA*) we are in the midst of preparing a manuscript of our results (planned submission 2021 December, the target journal is *RNA* – Q1, IF: 4.32).

Importantly, the *in silico* pipelines we have created to analyse snoRNA expression from transcriptomic dataset are "species-agnostic", thus could be used in other species. Indeed we are already planning a cooperation with researchers at KU Leuven to analyze snoRNA expression dynamics in human patient samples.

Mechanistic insights into the enzymatic activity of the dyskerin enzyme

Finally, we also made important discoveries in the more theoretical part of the collaborative work aiming to understand the mechanistic molecular details of dyskerin-driven pseudouridylation. This work, accomplished by Dr. György Ferenczy (ELKH-RCNS), Dóra Judit Kiss (ELKH-RCNS) and Dóra K. Menyhárd (ELTE) combined quantum and molecular modeling and has been able to identify the structural changes that occur in the enzymatic pocket of dyskerin during pseudouridylation. We proposed a detailed reaction mechanism for the uridine to pseudouridine transformation catalyzed by human box H/ACA pseudouridine synthase complex. A key feature of the mechanism is that a part of the substrate binding energy is stored in uridine conformational strain, and this reaction barrier lowering effect is missing in inactive mutants. Since neither uridine transformation nor substrate binding presents sequence requirements, we suggest that engineered guide snoRNAs that can associate with the pseudouridine synthase complex inherently and endogenously present may be applied to facilitate the pseudouridylation of nearly any substrate. A manuscript describing these results in detail is under review at *ACS Catalysis* (D1, IF:13.08).

Finally, as detailed in the original plans, during the first year of the project we were able to complete the purchase of a new, automated Stand Alone zebrafish maintenance-system (Tecniplast). This system is now installed in our fish facility and houses all the relevant fish lines for the experimental work described above.

Relevant publications during the four years of the project:

Máté Varga, Dorottya Ralbovszki, Eszter Balogh, Renáta Hamar, Magdolna Keszthelyi, Kálmán Tory (2018) Zebrafish Models of Rare Hereditary Pediatric Diseases. *DISEASES* 6**:**(2) p. 43 https://doi.org/10.3390/diseases6020043

Miguel Godinho Ferreira, Catarina Martins Henriques, Kathleen Claes, **Máté Varga**, Maria Luisa Cayuela, Maria Caterina Mione (2019) The zebrafish as an emerging model to study DNA damage in ageing, cancer and other diseases. *FRONT CELL AND DEV BIO* https://doi.org/10.3389/fcell.2018.00178 (Q1, IF: 5.186)

Eszter Balogh, Jennifer C. Chandler, **Máté Varga**, Mona Tahoun, Dóra K. Menyhárd, Gusztáv Schay, Tomas Goncalves, Renáta Hamar, Regina Légrádi, Ákos Szekeres, Olivier Gribouval, Robert Kleta, Horia Stanescu, Detlef Bockenhauer, Andrea Kerti, Hywel Williams, Veronica Kinsler, Wei-Li Di, David Curtis, Maria Kolatsi-Joannou, Hafsa Hammid, Anna Szőcs, Kristóf Perczel, Erika Maka, Gergely Toldi, Florentina Sava, Christelle Arrondel, Magdolna Kardos, Attila Fintha, Ahmed Hossain, Felipe D'Arco, Mario Kaliakatsos, Jutta Koeglmeier, William Mifsud, Mariya Moosajee, Ana Faro, Eszter Jávorszky, Gábor Rudas, Marwa H. Saied, Salah Marzouk, Kata Kelen, Judit Götze, George Reusz, Tivadar Tulassay, François Dragon, Géraldine Mollet, Susanne Motameny, Holger Thiele, Guillaume Dorval, Peter Nürnberg, András Perczel, Attila J. Szabó, David A. Long, Kazunori Tomita, Corinne Antignac, Aoife M. Waters, Kálmán Tory (2020) Pseudouridylation defect due to *DKC1* and *NOP10* mutations causes nephrotic syndrome with cataracts, hearing impairment, and enterocolitis, *PROC NAT ACAD SC*I, 117 (26): 15137-15147 https://doi.org/10.1073/pnas.2002328117 (D1, IF: 11.2) (*co-first and co-corresponding author*)

Submitted publications (under revision):

Kiss, Dóra; Olah, Julianna; Stirling, András; Toth, Gergely; **Varga, Máté**; Menyhard, Dora; Ferenczy, György (2021) The structure-derived mechanism of box H/ACA pseudouridine synthase offers a plausible paradigm for programmable RNA editing. *ACS CATALYSIS* (under revision) (D1, IF: 13.08)

Manuscripts in preparation:

Renáta Hamar and **Máté Varga**: The snoRNAome of zebrafish. (in preparation, planned submission to *RNA* (Q1, IF: 4.32))