Skoltech
Skolkovo Institute of Science and Technology

Thesis Changes Log

Name of Candidate: Evgeniia Shcherbinina

PhD Program: Life Sciences

Title of Thesis: Role of lncRNA LL35 in hepatocyte function

Supervisor: Dr. Timofei Zatsepin

The thesis document includes the following changes in answer to the external review process.

Pavel Ivanov

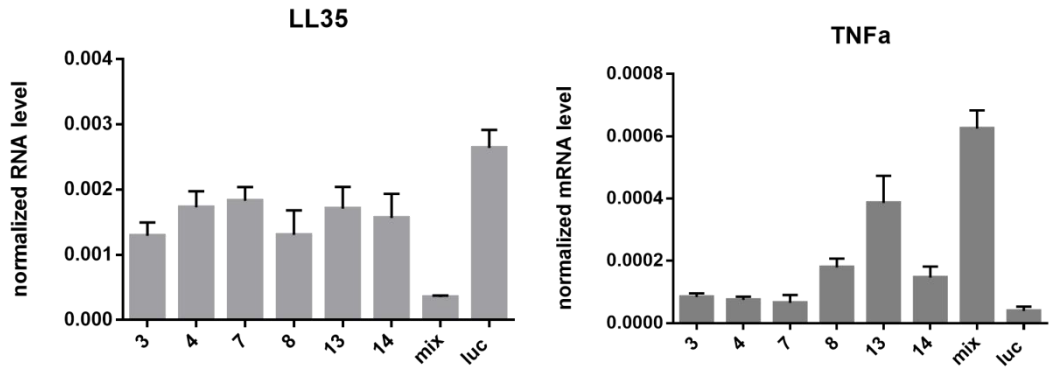
Specific questions about the work:

1) The RACE analysis identified 1193nt transcript that is different to the annotated sequences. What could be the reasons for it?
It is a common story for lncRNA transcripts due to the issues in annotation of transcriptome data. Another reason can be a diversity of lncRNA transcripts in different cell lines and tissues. Taking into account that NCBI provides 42 transcripts for LL35 and Ensembl only 3 transcripts, we can conclude that annotation of LL35 transcripts and especially the ends of the transcripts in public databases may be unreliable. Using RACE analysis, we identified the transcript with the same core sequence as an annotated one, but 148 nt shorter at 5'- end and 279 nt shorter at 3'- end. Since inaccurate annotation of transcript ends using RNA-seq is a common problem, RACE is always needed for accurate detection of splicing products. Thus, we were able to determine the exact sequence of one of the LL35 transcripts in the AML12 cells, which is important for investigation of its function.

2) Loss-of-function approach has been chosen. Would overexpression studies complement these data, what are the limitations of these approaches?
Yes, overexpression studies would complement our studies and can serve as a confirmation of effects we observed during loss-of-function studies. We used LL35 transcript sequence identified by RACE to create a vector for LL35 insertion into the genome. We checked LL35 expression at day 7 after transfection with the vector and it was ~9 times higher than in control cells. Unfortunately, after two months of selection we could not confirm the overexpression of LL35 in the selected cell line. Among possible reasons, one can be silencing the transcript's promotor after random insertion into the genome. Moreover, difficulties with lncRNA overexpression are rather frequent. With limitations of overexpression approach, I faced in my study, when I didn't succeed with the creation of a stable cell line with LL35 overexpression. At the same time ASO- or siRNA- knockdown techniques require proper delivery inside the cell and usually leave 10-20% expression of target gene. Loss-of-function methods may also result in off-target effects and immunogenic toxicity.

3) How would you control unspecific side effects of ASO-mediated depletion studies? Why exactly 5 ASOs were chosen for these LOF studies?

We modeled a secondary structure of LL35 and designed more than 10 ASO specific to different parts of LL35 hairpins to cover all annotated transcripts of LL35. To exclude possible off-targets we checked ASOs complementarity to all possible murine transcript and miRNAs by BLAST. After that we checked the ASO efficacy separately, then we tested different combinations of the most effective ASO. Additionally, to check the possible off-targets we measured TNFa mRNA expression level and found that ASO #13 cause its upregulation. This ASO was removed from the mix. The chosen mix of 5 ASO (#3,4,7,8,14) used in this study showed the best efficacy for LL35 knockdown without significant off-target effects.



4) A cut-off for differentially expressed genes (|log2foldchange|>0.8,adjustedp-value<0.1) has been chosen. How would justify this choice? Why only 5 differentially expressed genes were common between in vitro and in vivo sets?

Based on the number of differentially expressed genes with each p-value and FC values, we considered that in this case more changes and pathways were captured using filter by |log2foldchange|>0.8 and adjusted p-value<0.1. This cut-off is acceptable (https://www.nature.com/articles/srep32460) and resulted in the optimal numbers of both DEGs and enriched pathways. Low intersection between differentially expressed genes *in vitro* and *in vivo* may be explained by more complex cellular communications in the liver than in cell culture. Moreover, GalNAc-conjugates provide targeted delivery to hepatocytes only, while there are other cell types in the liver such as hepatic stellate cells, Kupffer cells and liver sinusoidal endothelial cells, which cannot be excluded from RNA-seq analysis. Moreover, it is not rare that RNA-seq data for cell culture and tissue have unexpectedly low similarities (https://www.mdpi.com/1422-0067/22/13/6958, https://pubmed.ncbi.nlm.nih.gov/33106934/). Small intersection between sets of DEGs highlights the importance of conducting *in vivo* studies. Probably, usage of another murine hepatocyte cell line or more complex organoids would give more similarities with liver transcriptome than AML12. I also added the discussion of this point to the page 149.

5) Putative hits from RNA-seq data were validated by qRT-PCR. Why western blotting was not used to validate hits on the protein level?

The aim of RT-qPCR for hits from RNA-seq was to prove that obtained transcriptome data is valid and that we can use this data for further pathway enrichment analysis. At this level, we were not interested in evaluation of protein expression, especially since we could hardly expect that we would detect similar changes at proteins level by Western-blotting as we observed in transcriptome. Western-blotting was used in this study to look at the specific changes we were interested in.

6) There is a good correlation between pathways found in «–omics» data. How many/what

the fraction of individual genes/ proteins overlap between RNA-seq and proteomics analyses?
Proteomic data was obtained only for *in vitro* studies. Changes in the transcriptome do not always lead to changes in the proteome (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3111477/). Moreover, the sensitivity of RNA-seq and LC-MS is very different. While there is not big overlap between individual proteins and genes, major pathways overlap, and we think this is the most important.

7) Sponging of microRNAs: how similar human miR-222-3p and murine miR-22-3p and 23a-3p?
Human miR-222-3p and murine miR-222-3p have 100% identity, murine sequence is 2 nt longer than human. miR23a-3p is also 100% identical in human and mice.

8) It is not clear whether total lysate or nuclear extracts were used for RIP and/or biotinylated LL35 pull down analyses.
I've added this point to Methods and Materials section to the pages 110-111. For RIP the total lysate was used. The data presented in the Thesis text for LL35 pull-down analysis were obtained with total lysates. But we also performed pull down assay using only nuclear lysate, which didn't lead to acceptable results.

Additional points and discussion:

I feel that Discussion section should be revised. The following points should be discussed: 1) What are the limitations of the study?; 2) Why specific cell line was chosen for in vitro studies?; 3) What are possible cytoplasmic functions of LL35, and whether cytoplasmic pool of LL35 is affected during LOF studies?; 4) Why RIP approach was chosen? Are any of putative binding proteins (GC1α, STAT3, PKM1, CTNNB1, SIRT1, IGF2BP2) are known RNA-binding proteins?; and 5) Future directions
All suggested points were added to the Discussion section (pages 148, 152, 153, 156).

Minor points:

* NMD is a "nonsense-mediated mRNA decay" (and not 'nonsense-mediated decay').
It was corrected (pages 11, 24).
• Figure 26. It would be useful to give more information in the figure legend (e.g., what are green boxes). Also exon numbering should be added.
I corrected the figure and the figure legend (pages 116-117).
• Fig 35C and its quantification: how many repeats were done for this analysis?
For this analysis we used 3 replicas per condition, information was added to the figure legend (page 138).

Petr Sergiev

1) Why there are so small overlap between differentially expressed genes in vitro and in vivo? Might the smaller number of differentially expressed genes in the liver be explained by less efficient LL35 downregulation?
Low intersection between differentially expressed genes *in vitro* and *in vivo* may be explained by more complex cellular communications in the liver than in cell culture. Moreover, GalNAc-conjugates provide targeted delivery to hepatocytes only, while there are other cell types in the liver such as hepatic stellate cells, Kupffer cells and liver sinusoidal endothelial cells, which cannot be excluded from RNA-seq analysis. Moreover, it is not rare that RNA-seq data for cell culture and tissue have unexpectedly low similarities

(https://www.mdpi.com/1422-0067/22/13/6958, https://pubmed.ncbi.nlm.nih.gov/33106934/). Small intersection between sets of DEGs highlights the importance of conducting *in vivo* studies. Probably, usage of another murine hepatocyte cell line or more complex organoids would give more similarities with liver transcriptome than AML12. Possibly, the smaller number of differentially expressed genes in the liver may be explained by less efficient LL35 downregulation in whole liver because we had a delivery only in the hepatocytes. I also added the discussion of this point to the page 149.

2) How one could explain opposite effects of LL35 downregulation on several genes in vivo and in vitro (e.g. Chek1, Cenpi, Gtg6, Gstm2 etc.)?
I think that the opposite effects could be explained similar to low intersection between sets of differentially expressed genes. In tissues, cells receive a signal from neighboring cells of a different type, so the regulation of their internal processes differs from that of a cell line. We used targeted delivery to hepatocytes only, while for expression analysis we used a part of liver, which contained several cell types.

3) Metabolome analysis and changes in the glucose metabolism suggested to check differential expression for a number of genes. How well the results correlate with transcriptome analysis?
The results for differential expression of metabolism related genes obtained by RT-qPCR partially correlate with transcriptomic data. Some genes don't pass the threshold for adjusted p-value <= 0.1.

4) How proteins were selected for pull-down analysis of LL35?
For RIP analysis proteins were selected based on literature review (those, which were shown to interact with lncRNA, participate in regulation of glucose and lipid metabolism, have relation to DEANR1 and on the intermediate hypotheses that appeared in the course of our study (https://jeccr.biomedcentral.com/articles/10.1186/s13046-019-1470-y, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2790597/, https://www.nature.com/articles/s41388-020-01525-3). Moreover, all these proteins except of SIRT1 were shown to be RNA-binding proteins, in particularly, STAT3, IGF2BP2 and PKM1 are able to bind different lncRNAs. I added this point to the discussion section.

5) Many RNP complexes are formed efficiently in the cell, rather than in extracts. Would it be better to fish out naturally formed LL35 RNPs by e.g. complementary biotinylated oligonucleotide?
Before we did pull-down for LL35 protein partner identification, we tried to use CHART (Capture Hybridization Analysis of RNA Targets) - the most common method for the identification of lncRNA protein partners, rather than FISH. In CHART we used biotinylated oligonucleotides specific to LL35 and formaldehyde, which serves to covalently connect the RNA to its biological targets at the time of cross-linking. Unfortunately, we didn't manage to identify LL35 partner by CHART and repeat it several times with the same result, probably, because of low concentration of LL35-protein complexes and high level of non-specific RNA binding with oligonucleotides. After we used pull-down and managed to identify at least one possible LL35 partner, which is needed to be further validated.

6) Micro RNA sponge function was assumed to lower concentration of miRNAs. Why is it so? Sponge should lower the concentration of free miRNA, but presumably not total.
In some publications, including these two related to DEANR1 (https://iubmb.onlinelibrary.wiley.com/doi/10.1002/iub.2416, https://link.springer.com/article/10.1007/s10863-021-09874-8) authors observed negative

correlation between expression of miRNA and lncRNA which has ability to sponge this miRNA. This negative correlation usually serves as a first evidence in favor of possible interaction between lncRNA and miRNA. This interaction should be further proved by dual luciferase assay and using RIP for AGO2 protein. Due to the predominantly nuclear localization of LL35, we do not think that the sponge is its main function, that is why we touched on the study of this sponge hypothesis only superficially and performed only RT-qPCR after LL35 depletion to check the correlation between expression levels of LL35 and two miRNA of interest. Because we didn't observed changes in miRNAs expression we didn't proceed with sponge hypothesis. But accurate study of possible ability of LL35 to bind different miRNAs may be a future direction of this work and should start with bioinformatics prediction, analysis of miRNA targets expression levels, RIP for AGO2 and finally dual luciferase assay.

Richard Lathe

Critique

Using a single experimental strategy based on LL35 knockdown, the candidate has investigated multiple parameters that might be affected using an impressive array of analytical techniques. The broad outcome of this study is as follows:
(i) Small changes were seen in almost all parameters investigated, from gene expression, to pathway analysis, to proteomics and metabolomics and lipidomics, to glucose metabolism and insulin tolerance, to cell division.
Although many of the changes are small, they are statistically significant, so we consider them important. Obtained data correlate with published data about LL35 role in murine lungs. Authors also didn't observe crucial effect from LL35 knockout in murine lungs, they expected to see impairments in lungs development, but observed only impairment in recovery after epithelial injury in lungs. Taking together these data, I think that it might be meaningful and more effective to study LL35 function in models of various diseases rather than under normal physiological conditions. This could be a direction for further study of LL35.

(ii) The statistical analysis was purely by pairwise t-testing.
The statistical analysis was performed by multiple t-test because it is common for such data and there were no requirements to apply different statistical tests during publication of main results of this study. But I will keep this point in mind for my future research.

(ii) Results from in vivo and in vitro knockdown were different.
Different results *in vivo* and *in vitro* may be explained by more complex cellular cellular communications in the liver than in cell culture. Moreover, GalNAc-conjugates provide targeted delivery to hepatocytes, while liver contains other cell types, such as hepatic stellate cells, Kupffer cells and liver sinusoidal endothelial cells, which cannot be excluded from RNA-seq analysis. Moreover, it is not rare than RNA-seq data for cell culture and tissue have unexpectedly low similarities (https://www.mdpi.com/1422-0067/22/13/6958, https://pubmed.ncbi.nlm.nih.gov/33106934/). Small intersection between sets of DEGs highlights the importance of conducting *in vivo* studies. Probably, usage of another murine hepatocyte cell line or more complex cell culture studies would give more similarities with liver transcriptome than AML12.

(iii) In some cases the results obtained argue against the contention that LL35 is a tumor-suppressor.

I agree that the fact that LL35 knockdown causes decrease in AML12 cell viability and migration is controversial to the decreased expression of LL35 in HCC and after partial hepatectomy. But our research is focused on LL35 function in normal hepatocyte and we didn't describe the role of LL35 as a tumor suppressor. To discuss the role of LL35 as a potential tumor suppressor more experiments, such as loss-of-function and gain-of-function, need to be done in hepatoma cell lines and HCC mouse models, which were not the focus of current research.

The overall picture is that none of the results identify a major role of LL35, and none of the effects of knockdown were remarkable. It is therefore unclear whether the central conclusion of the thesis – that murine lncRNA LL35 is the functional analog of human lncRNA DEANR1 – is substantiated by the data.
The fact that LL35 is a functional analogue of DEANR1 was proposed by Swarr et al. during investigation of LL35 functions in murine lungs. Authors emphasized this point because LL35 and DEANR1 shares similar locus downstream FOXA2 gene, have conservative sequences in promoter region and first exon, both participate in the regulation of FOXA2. But due to the low sequences conservation and small effect of LL35 downregulation on FOXA2 expression (both in murine lungs and liver) compared with DEANR1, we decided to double-check the fact of their analogy in murine liver. We conclude that LL35 is a functional analogue of DEANR1 in murine hepatocytes based on the following data: they share similar locus downstream FOXA2 gene, they are downregulated in HCC in comparison with normal liver, they are involved in the regulation of glycolysis, they are involved in the regulation of Notch and NF-kB pathways in similar way.

Suggested revisions and amendments

1. The grammar of the title should be amended to read: Role of lncRNA LL35 in hepatocyte function (singular case for hepatocyte).
I changed the title of my Thesis from "Role of lncRNA LL35 in hepatocytes function" to "Role of lncRNA LL35 in hepatocyte function".

2. The introduction should be slightly expanded to include an additional (or supplementary) figure addressing sequence similarity between LL35, DEANR1, and corresponding sequences in other species. It is important to answer the question of how conserved the lncRNA sequence is, and whether similar sequences are present (or not) in other related metazoan species including marsupials, birds, amphibians, and fish, and possibly even insects, and to look at the divergence of LL35/DEANR1 sequences versus control sequences (e.g., FOXA2, other lncRNAs).
In Literature review part 2.4 (page 82) I described that the promoter region, transcriptional start site and first exon of LL35 have strong conservation across all primates and placental mammals. However, the rest LL35 sequence is poor conservative between mouse LL35 and human DEANR1 (illustrated in Figure 22). I also added to the paragraph that LL35 conservation extends to Xenopus tropicalis, zebrafish, and tetraodons but not lampreys. According to your suggestion I added this point also to the introduction (page 14) (not only to literature review). Also it is rather common that lncRNA sequences have low conservation among species, making it's difficult to speak about their homology and analogy. I'm not sure that divergence calculation will be helpful here, because canonical formulas for divergence are based on codons and sites with synonymous substitutions, which is not applicable to lncRNAs. Moreover, the secondary structure may be prevalent in the question of lncRNAs homology. I've checked the similarity between DEANR1 and LL35 genes and found 16% similarity of their sequences (calculated by DEANR1), while for FOXA2 it is 72%. Moreover, FOXA2 ortholog is presented even in *Drosophila*.

3. The conclusion of the thesis (LL35 is the functional analog of DEANR1) should be rephrased to reflect the uncertainties encountered in this work. The candidate should also give thought to what experiments might be required to more accurately evaluate the relationship between LL25 and DEANR1.
For lncRNAs, there are no strict criteria for determining their analogy in different species. For example, in the paper about LL35 function in lungs https://pubmed.ncbi.nlm.nih.gov/30923168/ authors state that LL35 is a functional analog of DEANR1 based only on their similar genomic location, conservation of short sequence in the beginning and similarity of FOXA2 regulation. We have described their functional similarity more broadly, that is why we concluded that LL35 is the functional analog of DEANR1.

Konstantin Lukyanov

Minor points:

1. Page 95: “Libraries were sequenced by HiSeq4000 (Illumina, San Diego, USA) instrument in 5 nt single read mode”. What does "5 nt" mean? Were only 5 nucleotides sequenced?
I apologize for this typo, the sentence was corrected to “Libraries were sequenced by HiSeq4000 (Illumina, San Diego, USA) instrument in a single read mode”.

2. Materials and Methods: Company name, city, state and country is mentioned every time. This clutters up the text; for example, see page 101, where “Avanti Lipids, Birmingham, AL, USA” repeats 10 times in one sentence. It is generally accepted to note city, state and country of a company for the first time only.
I corrected the Materials and Methods according to your comment: I left the country, state and city only at the first mention of the manufacturer's company, with repeated references to the company, I left only its name.

3. Page 113: “Center for Molecular and Cell Biology”. It should be Center for Molecular and Cellular
Biology. “Institute of Biomedica Chemistry” -> Institute of Biomedical Chemistry.
Thank you for these comments, these points were corrected.

4. In Fig. 29B,C, Fig. 31 and Suppl. Fig. S3 bars have different shades of gray. Is it meaningful?
No, this is just because each scale bar represents a particular pathway, molecular function, etc., and since all pathways are different, bars have different colors, and not to make them colored, they have different shades of gray.

5. Page 131: “mean fluorescent intensity for LL35 ASO = 104.9 ±26.1 pixels, mean for Luc ASO = 76.0 ±20.1 pixels”. Usually, fluorescence intensity in images are not measured in pixels.
I changed pixels to the relative fluorescent intensity (RFI) (page 131).

6. Formatting of Supplementary Figures and their legends could be better if the legends are placed on the same page below the image.
I corrected this point.

Yury Popov

Specific comments:

1. Section 3.8 the principle of qRT-PCR (TaqMan?) should be specified
I this work qRT-PCR was performed using SYBR green. I added specification to Materials and Methods section 3.8 (page 91).

2. All legends to the figures should specify the number of biological replicates studied/used to build the graphs and calculate p value. Statistical tests and p value nature should be specified as well.
This point was corrected in the text: I added the replicates number and specified tests for p-value in all legends to the figures.

3. Fig. 28. ALT/AST and other LFTs units should be checked against the normal ranges, possibly assay calculation error
We compared blood parameters from mice with LL35 knockdown with those from mice injected with luciferase ASO. We believe that in this case, this comparison allows us to more reliably evaluate the effect of LL35 knockdown, and not of conjugates injection, for example.

4. Since the hexokinase seems regulated by LL35, please briefly comment on possible involvement in dietary fructose metabolism, the question highly relevant to public health and liver disease
I added a discussion of this issue to the page 151:
"LL35 knockdown led to significant upregulation of hexokinase 2 (Figure 35B), which is involved not only in glucose, but also in dietary fructose metabolism. Excess of fructose leads to elevated TAG, LDL cholesterol, apolipoprotein B, which are risk factors of cardiovascular disease, as well as it is associated with high blood pressure, nonalcoholic fatty liver disease (NAFLD) and fructose malabsorption [229]. In liver hexokinase predominantly converts glucose to glucose-6-phospate, but can also convert fructose to fructose-6-phospate to a lower degree [230]. Possibly, hexokinase upregulation after LL35 knockdown may also lead to an imbalance in fructose metabolism."

5. Important: interpretation of data in 4.7 should be refined to be more specific to what was studied – cell proliferation appears to be referred to interchangeably with that of cell survival. This needs to be critically revised.
I have replaced «cell proliferation» to cell viability wherever this term was used to interpret the results of the MTS assay. Changes were made on pages 3, 143, 145, 146, 154, 156, 157.

6. Discussion may benefit from brief specific paragraph elaborating potential implications of results – for future research, therapeutic and diagnostic applications, specific diseases, etc.
The paragraph about clinical relevance and possible future directions of my study was added to the end of the Discussion section (page 156).
"Our study was focused on LL35 function in normal hepatocytes, but the role of human DEANR1 is actively studied in different pathologies, mainly in cancer. Therefore, for greater clinical relevance of our study, one of the future directions from the one hand is the study of LL35 in HCC, from the other hand – investigation of DEANR1 in normal hepatocytes based on the obtained for LL35 data. Overexpression of LL35 in HCC will demonstrate its functions in cancer progression and metastasis and also possible involvement in the response to anticancer therapy. We demonstrated that depletion of LL35 in hepatocytes interferes with normal glucose and lipid metabolism, and impair response to insulin treatment. Disruption of these features in the liver leads to various diseases including diabetes, steatosis, NAFLD (non-alcoholic fatty liver disease) and others [238]. Thus, one more future direction of the research may be the study of the LL35 and DEANR1

functions in metabolic liver disorders".