Research Article

Investigating Antisense Transcription at the HTT Locus

Eugeny A. Elisaphenko and Anastasia A. Malakhova

Federal Research Center Institute of Cytology and Genetics, the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
ORCID
Anastasia A. Malakhova; 0000-0003-1916-1333

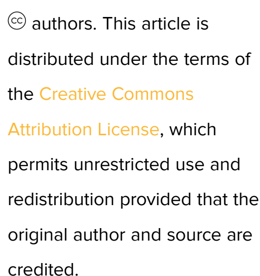
Abstract. Antisense transcription is an important mechanism of gene expression regulation. Antisense RNAs play a role in mRNA processing, translation and epigenetic modifications of DNA and histones in the locus of their origin, leading to gene silencing. *HTT* is a widely expressed gene, the mutation of which causes Huntington's disease. The product of the gene plays an important role in many cell processes, such as intracellular trafficking, cell division, autophagy, and others. An antisense transcription has been found at the *HTT* 5'-region. The *HTT-AS* gene has been reported to affect *HTT* expression in a Dicer-dependent manner. In this study, we analyzed extensive data from RNA-seq experiments for antisense transcription at the *HTT* locus. Antisense transcripts corresponding to the *HTT-AS* gene were not found. However, we revealed a number of antisense transcripts in different parts of the locus that may take part in the regulation and functioning of the *HTT* gene.

Keywords: antisense transcription, *HTT-AS*, *HTT*regulation, Huntington's disease

Corresponding Author: Anastasia A. Malakhova; email: amal@bionet.nsc.ru

Dates
Published 13 January 2022

Publishing services provided by Knowledge E



© authors. This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use and redistribution provided that the original author and source are credited.

Selection and Peer-review under the responsibility of the 8th Scientific and Practical Conference Conference Committee.

OPEN ACCESS

1. Introduction

The *HTT* (*IT-15*) gene was described in 1993 as a high-level transcript in the different brain region and other tissues of human and rodent organisms [1]. Trinucleotide expansion in the first exon of the *HTT* gene causes Huntington's disease. The HTT (Huntingtin) protein plays a role in many cell processes, such as intracellular trafficking, ciliogenesis, cell division, autophagy, axonal transport, and even gene expression regulation. This suggests that the regulation of the *HTT* gene is important for cell functioning and differentiation.

Antisense transcription is one of the mechanisms of gene regulation. ncRNAs play a key role in transcriptional regulation of many important genes, for example, H19/Igf2 locus [2]. The mechanisms of transcriptional repression by antisense RNAs are diverse. Epigenetic mechanisms are associated with DNA and histone modifications. ncRNAs recruit DNA-methylases and histone deacetylases. Impairment in the process causes some human diseases. Another way of antisense RNA function is the regulation of transcription elongation and termination, as well as mRNA splicing. Moreover, antisense

transcripts can affect mRNA translation through blocking RNP complex formation or RNAse H recruiting [3].

An antisense *HTT-AS* gene was discovered in the 5'-region of human *HTT*. According to a recent study, the *HTT-AS* gene has 3 exons [4]. Transcripts start in the *HTT* exon 1 or in the *HTT* promoter region and are alternatively spliced. Authors demonstrated the CAG-repeat length-dependent and Dicer-dependent mechanisms of *HTT* expression regulation by the antisense transcript.

In this study, we have analyzed a large number of RNA-seq experiments on various cells and tissues from the SRA database. We used the annotated *HTT-AS* sequence of EST# DA153759 in GeneBank as a reference for our experiments. After analyzing over 100 RNA-seq samples, we have not found transcripts corresponding to the previously described *HTT-AS*. This is probably due to the difference in methods and low level expression of the antisense gene. However, we found a number of antisense transcripts in different parts of the locus that can take part in the regulation and functioning of the *HTT* gene and, ultimately, can be involved in developing methods of the Huntington's disease therapy.

2. Material and methods

HTT locus sequences of 6 mammal species were extracted from GeneBank (https://www.ncbi.nlm.nih.gov/nuccore/): NC_000004.12:2960681-3293960 *Homo sapiens* chromosome 4, GRCh38.p13; NC_000071.6:34711721-34962534 *Musmusculus* strain C57BL/6J chromosome 5; NC_010450.4:1759768-1994002 *Susscrofa* chromosome 8; NC_037333.1:c115774160-115550426 *Bostaurus* chromosome 6; NC_006585.3:c61271462-61000934 *Canis lupus familiaris* chromosome 3; NC_036883.1:c900136-600991 *Pan troglodytes* chromosome 4.

Computer analysis was conducted using modern versions of the following software packages: BLAST ([5]; http://www.ncbi.nlm.nih.gov) for searching for homologous sequences; RepeatMasker [6] for searching for mobile elements; and PipMaker[7] for genomic analysis of extended loci.

Reads for transcriptome analysis of various tissues and cells were downloaded from NCBI SRA database (http://www.ncbi.nlm.nih.gov/sra). Human run accession numbers: DRR062878; DRR062905; ENCFF000GF; ENCFF000GM; ENCFF002BIR; ENCFF002BJA; ENCFF002BJL; ENCFF002BJS; ERR1243461; ERR1273707; ERR1273713; ERR1590047; ERR1590048; ERR1590049; ERR1590052; ERR1665675; ERR1665677; ERR1665684; ERR1665685; ERR1837048; ERR2704712; SRR1222586;

SRR1530524; SRR1639742; SRR1747151; SRR1747160; SRR1747191; SRR1747205; SRR1747206; SRR1747209; SRR2014233; SRR2040582; SRR2051398; SRR2071348; SRR2126029; SRR2126033; SRR2126036; SRR2130096; SRR2130097; SRR2130100; SRR2130101; SRR2130103; SRR2130104; SRR2130105; SRR2130106; SRR2130110; SRR2130113; SRR2130116; SRR2130119; SRR2130122; SRR2173239; SRR2173256; SRR2428910; SRR2428911; SRR2428918; SRR2453310; SRR3191744; SRR3191745; SRR3192149; SRR3192368; SRR3192369; SRR3192427; SRR3192434; SRR3332539; SRR3332761; SRR3393504; SRR3393506; SRR3543513; SRR3555056; SRR4011889; SRR4040146; SRR4050230; SRR4421334; SRR4421642; SRR4421668; SRR4421689; SRR4421735; SRR4421736; SRR4422588; SRR4422625; SRR4444423; SRR4444436; SRR4444437; SRR445718; SRR445729; SRR490965; SRR490966; SRR490968; SRR490969; SRR490972; SRR490988; SRR491004; SRR491017; SRR5048183; SRR5048184; SRR5171068; SRR5171069; SRR5171072; SRR5189650; SRR5239180; SRR5239181; SRR529646; SRR529647; SRR5626810; SRR5643181; SRR5680558; SRR5680565; SRR5680570; SRR5680573; SRR5680574; SRR5680576; SRR5680581; SRR5680582; SRR5680590; SRR5680598; SRR5897144; SRR5897224; SRR6010596; SRR6010609; SRR6010615; SRR6010627; SRR6010652; SRR6010660; SRR6010681; SRR6010691; SRR6081906; SRR6082000; SRR6253484; SRR6253510; SRR6253520; SRR6348583; SRR6456115; SRR6459189; SRR6459190; SRR6860524; SRR7013742; SRR7013743; SRR7138154; SRR7469676; SRR7469700; SRR8378562; SRR8378593; SRR8575347; SRR8575351; SRR8575392; SRR8942607; SRR9670695. Mouse run accession numbers: SRR567499; SRR453142; SRR530640; SRR453174; SRR453147; SRR567486; SRR10039015. ncRNA analysis was performed using runs SRR5189650; SRR2126029; SRR2126033; SRR2126036; SRR8575347; SRR8575351; SRR8575392; SRR7013742 and SRR7013743.

Transcriptome reads were mapped to corresponding reference sequences using HISAT2 [8]и STAR [9] software. Output SAM files were analysed with Geneious 11.0.2 software (http://www.geneious.com) [10].

Exon-intron structure, splicing sites and read coverage was investigated using STRINGTIE [11]и TRINITY [12] software.

3. Results and Discussion

Antisense transcription is an important mechanism of gene expression regulation. Earlier, an antisense *HTT-AS* gene was discovered in the 5'-region of the human *HTT* gene. The antisense transcription starts either in the promoter region or in the exon 1 of

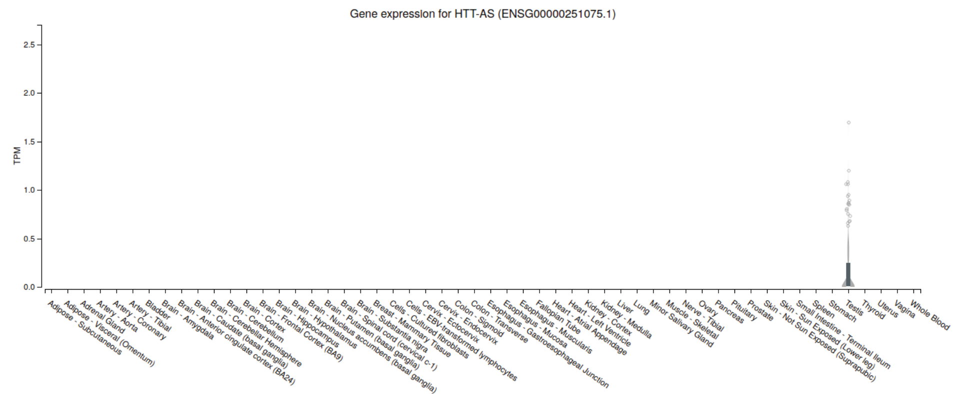


Figure 1: *HTT-AS* expression level in different human tissues and organs according to GTEx portal.

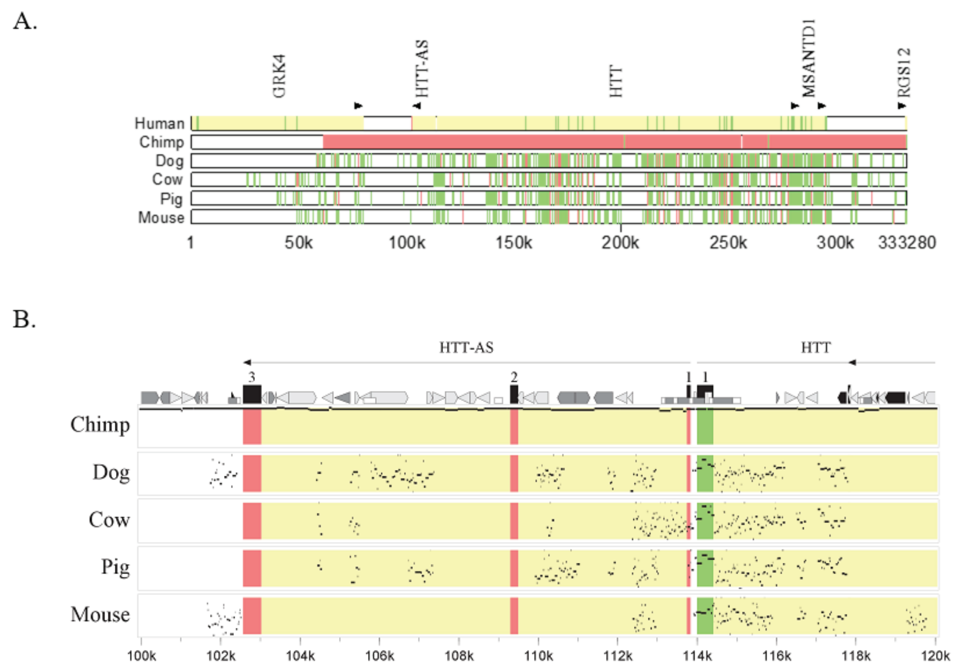


Figure 2: PipMaker analysis of the *HTT* locus of six mammalian species. Human *HTT* locus sequence was used as a reference. (A) — The locus alignment overview. Green colour shows regions with 50-80% homology level. Red colour shows regions with >80% of sequence homology. (B) — Percent Identity Plot of *HTT-AS* and *HTT* 5'-region sequence of six mammal species. Genes are depicted with arrows. Introns are shown in yellow. Exons of *GRK4*, *HTT* and *MSANTD1* are green. *HTT-AS* exons are red. Gray triangles show repear regions and mobile elements. Black rectangles show exons. The Y-axis demonstrates homology level of 50% to 100%.

the *HTT* gene (EST#DA153759; [4]). Considering the *HTT* regulation to be conservative in evolution, we performed a comparative analysis of the *HTT* locus sequence in several mammalian species. PIPMaker analysis revealed highly conservative *HTT* surrounding. All species under investigation harbour*GRK4* (*Grk4*) (G protein-coupled receptor kinase 4) upstream and *MSANTD1* (*Msantd1*) (Myb/SANT DNA binding domain containing 1) downstream of the *HTT* (*Htt*) gene (Figure 1). The structure of the syntenic group is highly

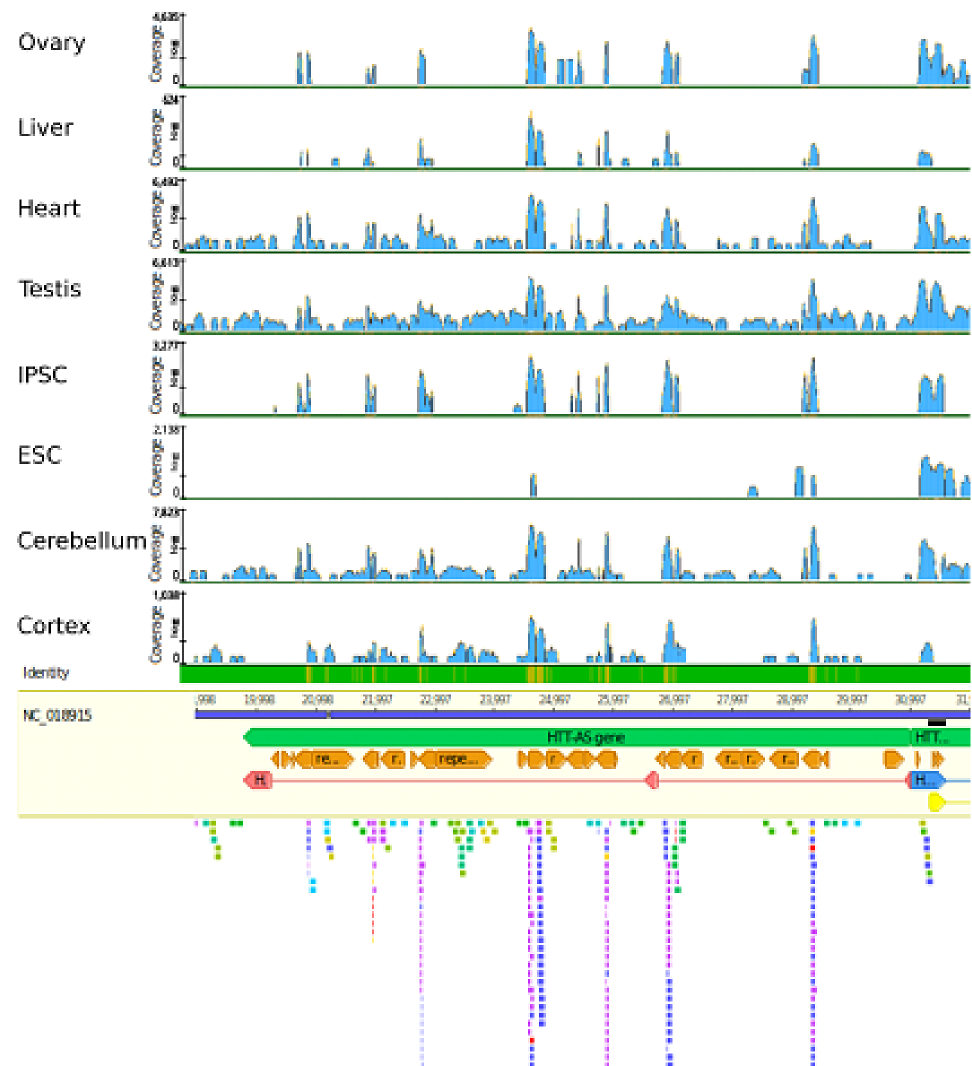


Figure 3: Mapping RNA-seq experiment reads of different human tissues to the HTT locus. Blue peaks demonstrate expression level in the region, its value was determined as the decimal logarithm of the number of reads mapped to the reference sequence. Genes are shown in green. HTT-AS exons are red. Repeat regions and mobile elements are shown in orange. The lower panel demonstrates an example of mapping reads of cortex RNA-seq experiment.

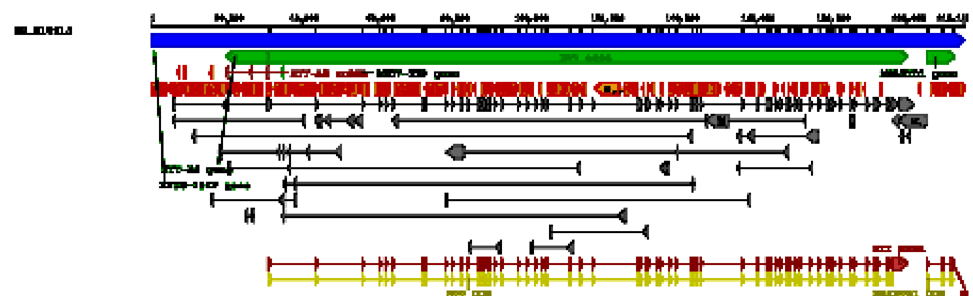


Figure 4: Sense and antisense transcription at the human *HTT* locus based on RNA-seq experiments. Gene sequences are green. Repeats and mobile elements are orange. Grey arrows depict spliced antisense transcripts discovered in the expression analysis. Red arrows depict *HTT* and *MSANTD1* transcripts; yellow are coding exons.

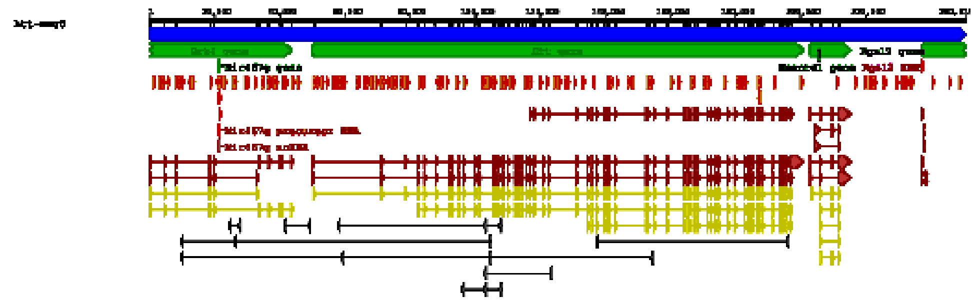


Figure 5: Sense and antisense transcription at the murine *HTT* locus. Genes are depicted in green. Repeats and mobile elements are orange. Grey arrows depict spliced antisense transcripts discovered in the expression analysis based on RNA-seq experiments. Red arrows depict *Grk4*, *Htt* and *Msantd1* transcripts; yellow are coding exons.

conservative in evolution that confirms its importance in the regulation and functioning of the genes.

The PipMaker analysis demonstrates that the exons and partially introns of *HTT*, *GRK4* and *MSANTD1* genes have a high level of homology (>70%) in contrast to *HTT-AS*. Only *HTT-AS* exon 1 is partially homologous between the 6 mammal species in its 5' region adjacent to the *HTT* promoter (or located in the *HTT* promoter region). Human *HTT-AS* exons 2 and 3 show some homology only with chimpanzees, which is quite expected due to the species affinity and similarity of the genomes.

Neither investigation of the SRA of HD patient samples, nor GteX database revealed any *HTT-AS* expression in brain and other tissue samples, except testis (Figure 2).

We performed an analysis of expression patterns in the human *HTT* locus. We used over 100 RNA-seq experiments of different tissues extracted from SRA database (Figure 3). The *HTT* gene expression was found in all samples; the expression level differed between tissues. However, the expression of *HTT-AS*was not discovered. Analysis of the experiments with a high level (over 10x) of coverage revealed a number of reads mapped to the region between *HTT* and *GRK4* genes including *HTT-AS* putative exons. However, the most part of the reads located in the repeat region with mobile elements that are abundant in the human genome. According to a recent study [4], the strand-specific RT-PCR analysis revealed *HTT-AS* expression in most human tissues. The authors described the antisense transcription starting from position +300 of *HTT* start site. But the RNA-seq experiments do not confirm the reported results.

Assuming that *HTT-AS* can be expressed at a low level, we analyzed RNA-seq data for splicing and junctions using two different programs, TRINITY and STRINGTIE. None of the programs revealed any splicing or exon-intron structure in the area between the *HTT* and *GRK4* genes, although, the transcription pattern of the *HTT*, *MSANTD1* and *GRK4* genes fully coincides with previously published results (Figure 4).

It is obvious that the previously reported *HTT-AS* gene does not possess any open reading frame so it belongs to protein non-coding genes. Analysis of ncRNA-seq experiments does not reveal any non-coding RNA transcripts in the locus either.

However, the TRINITY and STRINGTIE programs detected a number of anti-*HTT* transcripts at the locus that matched, did not match, and partially matched its exons across almost the entire length of the gene. Some of them aligned with mobile elements, usually Alu or LINE. A transcript in the region of the *HTT-AS* gene was also detected. One of its exons includes exon 3 of *HTT-AS*, but the start of the transcription, some exons, and all the splicing sites align with mobile elements.

Similarly, we performed a study of the *Htt* locus in mice with the analysis of several RNA-seq experiments. In general, the pattern resembles the expression in human. A number of antisense transcripts are detected throughout the *Htt* gene (Figure 5). Some of them partially coincide with those detected in humans, but it is still difficult to verify the presence of an evolutionarily conservative antisense gene. There are also a number of transcripts in the 5'-region of *Htt* that start in the *Htt* promoter or even in the exons of the gene and continue to the neighbouring*Grk4* gene as an antisense RNA. The intron of the *Grk4* gene also contains *miR-467g* microRNA in the antisense orientation. We assumed that *miR-467g*was formed from one of the identified antisense transcripts. Thus, the role of antisense transcription at the locus in mice closely related to micro RNA expression. In the human genome, this micro RNA may be either not yet annotated or eliminated. Then the detected antisense transcription is either associated with micro RNA or is relict, in the case of its elimination.

In conclusion, based on the analysis of RNA-seq data, an antisense transcription occurs in the *HTT* locus. However, it is very different from the one published earlier. Its significance and functional role are still difficult to assess and require further experiments.

4. Acknowledgments

This work was supported by the Russian Science Foundation project No 16-15-10128.

References

[1] Li SH, Schilling G, Young WS et al. Huntington's disease gene (IT15) is widely expressed in human and rat tissues. Neuron. 1993;11(5):985-993. doi:10.1016/0896-6273(93)90127-d

[2] Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, LevorseJM,Tilghman SM. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature. 2000;405(6785):486-489. doi:10.1038/35013106

[3] Pelechano V, Steinmetz LM. Gene regulation by antisense transcription. Nature Reviews Genetics. 2013;14(12):880-893. doi:10.1038/nrg3594

[4] Chung DW, Rudnicki DD, Yu L, Margolis RL. A natural antisense transcript at the Huntington's disease repeat locus regulates HTT expression. HumanMolecularGenetics. 2011;20(17):3467-3477. doi:10.1093/hmg/ddr263

[5] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. JournalofMolecularBiology. 1990;215(3):403-410. doi:10.1016/S0022-2836(05)80360-2

[6] Smit AFA, Hubley R, Green P. RepeatMasker annotation Open-4.0. 2013/01/01. Available for: http://www.repeatmasker.org

[7] Schwartz S, Zhang Z, Frazer KA et al. PipMaker–A web server for aligning two genomic DNA sequences. Genome Research. 2000;10(4):577-586. doi:10.1101/gr.10.4.577

[8] Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory requirements. Nature Methods. 2015;12(4):357-360. doi:10.1038/nmeth.3317

[9] Dobin A, Davis CA, Schlesinger F et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635

[10] Kearse M, Moir R, Wilson A et al. Geneious. Bioinformatics.2012;28:1647–9.

[11] Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology. 2015;33(3):290-295. doi:10.1038/nbt.3122

[12] Grabherr MG, Haas BJ, Yassour M et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology. 2011;29(7):644-652. doi:10.1038/nbt.1883