



## Data Article

## RNA-Seq data of ALKBH5 and FTO double knockout HEK293T human cells



Egor A. Smolin, Andrey I. Buyan, Dmitry N. Lyabin,  
Ivan V. Kulakovskiy\*, Irina A. Eliseeva\*

Russian Academy of Sciences, Institute of Protein Research, Pushchino 142290, Russia

## ARTICLE INFO

## Article history:

Received 9 March 2022

Revised 7 April 2022

Accepted 12 April 2022

Available online 20 April 2022

Dataset link: [Transcriptome \(RNA-Seq\) of HEK293T, HEK293T ΔALKBH5ΔFTO cells \(Original data\)](#)

## Keywords:

m6A methylation  
m6A demethylation  
Epitranscriptomics  
RNA modifications  
ALKBH5  
FTO  
Demethylase  
Transcriptome

## ABSTRACT

N6-methyladenosine (m6A) is the most abundant, highly dynamic mRNA modification that regulates mRNA splicing, stability, and translation. The m6A epigenetic mark is erased by RNA demethylases ALKBH5 (AlkB Homolog 5) and FTO (Fat mass and obesity-associated protein). The ALKBH5 and FTO RNA demethylases recognize m6A in similar nucleotide contexts. Therefore, these proteins can partially substitute for each other. To assess the impact of total m6A demethylation failure we performed high-throughput sequencing of cytoplasmic RNA from ALKBH5 and FTO double knockout and wild type HEK293T cells. The RNA-Seq libraries were sequenced on Illumina NextSeq 500, trimmed, and mapped to the human genome. The consequent read counting and differential expression analysis in the R environment detected 5871 differentially expressed and 166 alternatively spliced genes comparing double knockout against wild type HEK293T cells. Raw data are deposited in NCBI Gene Expression Omnibus (GEO) repository under GEO accession GSE198050.

© 2022 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

\* Corresponding authors.

E-mail addresses: [ivan.kulakovskiy@gmail.com](mailto:ivan.kulakovskiy@gmail.com) (I.V. Kulakovskiy), [yeliseeva@vega.protres.ru](mailto:yeliseeva@vega.protres.ru) (I.A. Eliseeva).

## Specifications Table

Subject	Biological sciences
Specific subject area	Bioinformatics, Transcriptomics, RNA modifications
Type of data	Tables and figures
How the data were acquired	RNA-Seq libraries were obtained from cytoplasmic RNA of HEK293T and HEK293T $\Delta$ FTO $\Delta$ ALKBH5 cells. Sequencing was performed using Illumina NextSeq 500.
Data format	Raw and analyzed
Description of data collection	Cytoplasmic RNA from wild type and $\Delta$ ALKBH5 $\Delta$ FTO HEK293T cells was isolated using TRIzol LS Reagent and purified on Direct-zol RNA Miniprep columns. Libraries were prepared using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB) and RiboMinus Eukaryote System v2 (Thermo Fisher Scientific) for rRNA depletion. Statistical analysis was performed with R/Bioconductor.
Data source location	Institute of Protein Research (Pushchino, Russia)
Data accessibility	Analyzed data and raw RNA-Seq data were deposited in the NCBI GEO database under GEO accession GSE198050 ( <a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE198050">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE198050</a> )

## Value of the Data

- N6-methyladenosine (m6A) is the most abundant and dynamic mRNA modification that affects mRNA splicing, stability, and translation [1]. The double knockout  $\Delta$ ALKBH5 $\Delta$ FTO cells lack two m6A demethylases and provide suitable model to study m6A demethylation failure.
- The presented RNA-Seq data will be beneficial to study the impact of the impaired m6A RNA demethylation on gene expression.
- The data will facilitate understanding of individual and shared functional roles of FTO and ALKBH5 if analyzed along with existing RNA-Seq datasets from single demethylase KO cells [e.g. 2,3] and m6A-methylation profiles [e.g. 4,5].

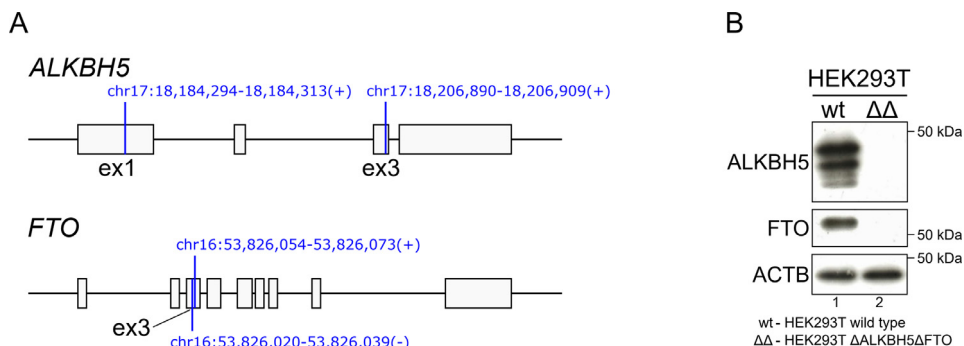
## 1. Data Description

Here we present the RNA-Seq data for ALKBH5 and FTO double knockout HEK293T cells. The ALKBH5 and FTO RNA demethylases have similar m6A substrate specificity [6] and likely can partially compensate for the absence of each other in the living cell. While there are existing RNA-Seq datasets for individual knockouts and knockdowns of ALKBH5 [e.g. 3] and FTO [2], this is the first transcriptomic profiling of the double knockout cells. We created the knockout cells using the CRISPR/Cas9 approach. The position of gRNAs and the lack of ALKBH5 and FTO in knockout cells are illustrated in Fig. 1.

The RNA-Seq was performed in three replicates for HEK293T and double knockout cells. The libraries were sequenced in a 76 bp single-end run on Illumina NexSeq 500. In all samples, quality-trimming (Q20) was required for less than 1% of reads. The total sequenced library size and mapping statistics are provided in Table 1, with more than 60% of total reads uniquely mapped and included in gene-level read counts. The raw reads (fastq) and gene-level read counts (tab-separated plain text) are available in the NCBI GEO database under GEO accession GSE198050.

### 1.1. The presented RNA-Seq data are suitable for analysis of differential gene expression in FTO and ALKBH5 double knockout cells

Gene-level read counting showed good concordance between experimental replicates and separation of wild type HEK293T from double knockout samples, see results of the principal



**Fig. 1. Evaluation of ALKBH5 and FTO knockout.** (A) Schematic representation of guide RNA (gRNA) locations. Two gRNAs targeting the third exon of FTO (for nickase CRISPR/Cas9) and two gRNAs targeting the first and third exons of ALKBH5 (for CRISPR/Cas9) are shown. The coordinates are according to the hg38 genome assembly. (B) Western blot of wild type and  $\Delta$ ALKBH5 $\Delta$ FTO HEK293T cells using anti-ALKBH5, anti-FTO, and anti-ACTB as a loading control.

**Table 1**

RNA-Seq library size and mapping statistics.

Sample name	Total library size	Uniquely mapped reads	Total gene-level read counts
HEK293T, replicate 1	10897264	8956475	7400887
HEK293T, replicate 2	11306936	9195340	7552615
HEK293T, replicate 3	8406417	6862227	5635235
$\Delta$ ALKBH5 $\Delta$ FTO, replicate 1	9226839	7324822	5924003
$\Delta$ ALKBH5 $\Delta$ FTO, replicate 2	10443280	8153351	6434276
$\Delta$ ALKBH5 $\Delta$ FTO, replicate 3	9633924	7664124	6137469

component analysis in Fig. 2A based upon 13718 genes that passed two counts per million in at least 3 of 6 samples. With these genes, we performed differential expression (DE) analysis to estimate overall RNA composition changes and detected 774 genes passing  $FDR < 0.05$  and  $|\log_2(\text{Fold Change})| > 1$ . Regarding the knockout targets, FTO ( $\log_2(\text{Fold Change}) = -0.9$ ,  $FDR < 10^{-40}$ ) and ALKBH5 ( $\log_2(\text{Fold Change}) = -0.4$ ,  $FDR < 10^{-18}$ ) were significantly downregulated (Fig. 2B,C). Thus, the RNA-Seq data confirm the knockout effect, and are suitable for differential expression analysis.

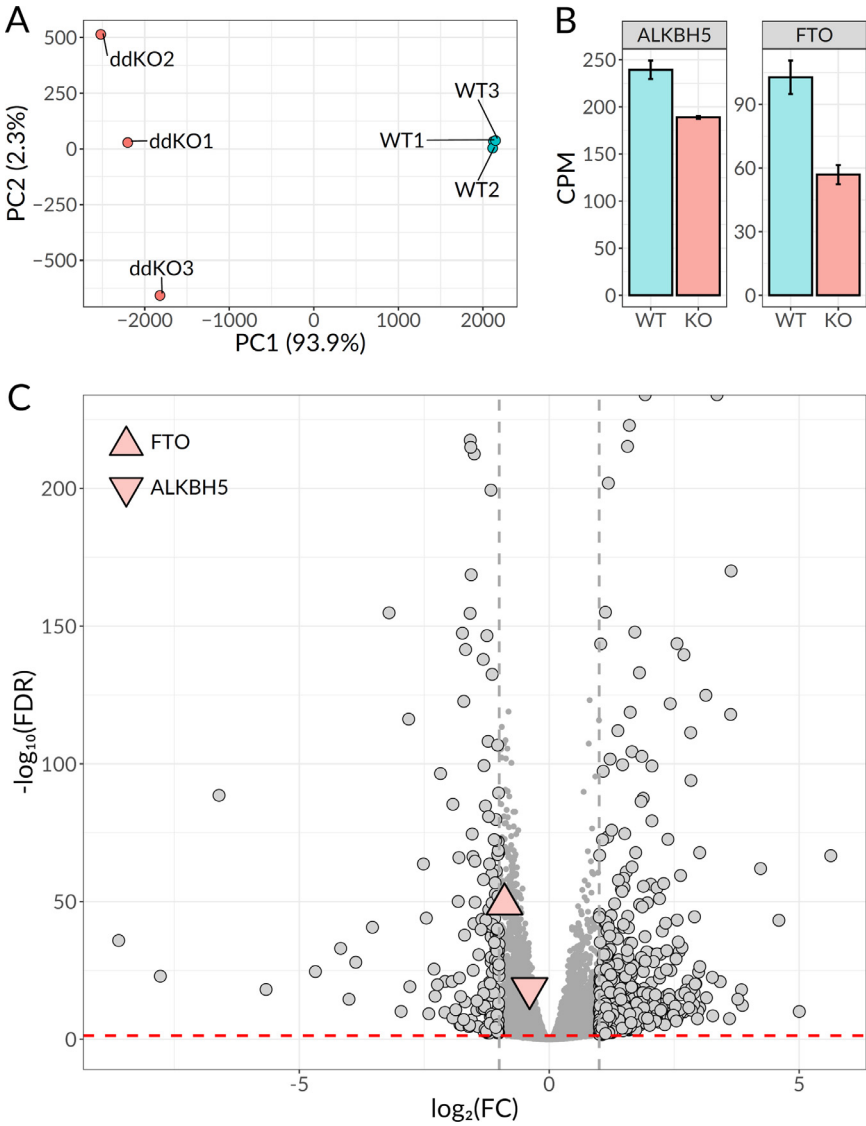
### 1.2. The presented RNA-Seq data are suitable for detection of alternative splicing events occurring upon FTO and ALKBH5 knockout

RNA N6-methyladenosine demethylation by FTO and ALKBH5 has been shown to affect splicing [2,7]. The presented RNA-Seq data are suitable for the detection of alternative splicing events, as differential exon usage analysis of 8887 multi-exonic genes found 166 cases passing  $FDR < 0.05$  (Fig. 3A). The illustrative examples of isoform switching under FTO and ALKBH5 knockout include RPS24, SETD7, ASPH, KRT8, and SEPT9 genes, the genomic views are shown in Fig. 3B–F.

## 2. Experimental Design, Materials and Methods

### 2.1. Cell culture and FTO/ALKBH5 knockout

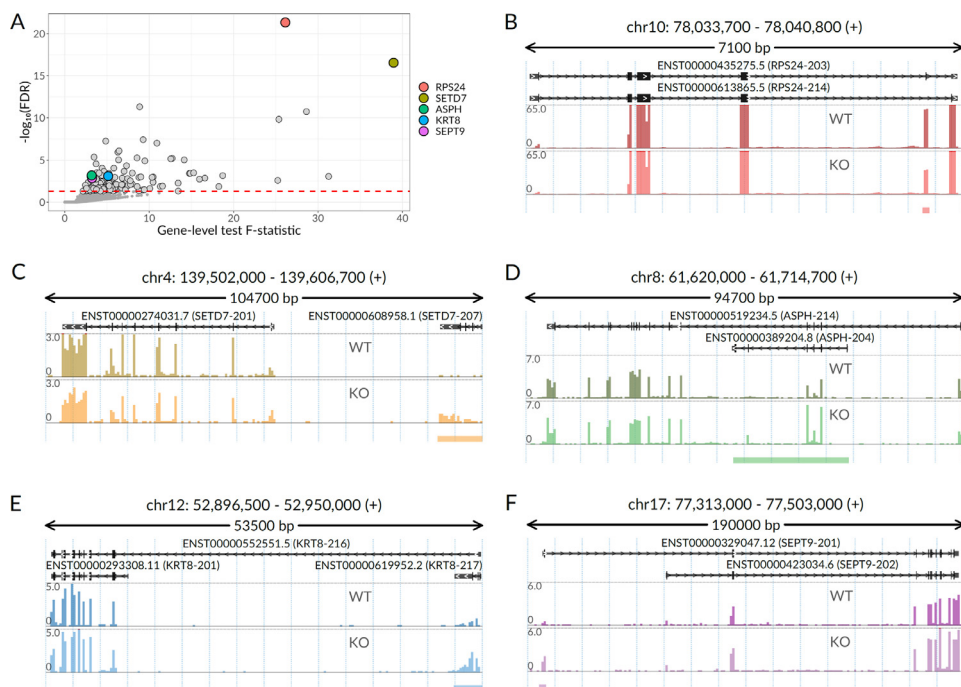
HEK293T cell line (originally obtained from ATCC, American Type Culture Collection) was kindly provided by Dr. Elena Nadezhkina (IPR RAS). The cells were cultivated in Dulbecco's



**Fig. 2. The data are suitable for analysis of differential gene expression under FTO and ALKBH5 knockout. (A)** Principal component analysis for three replicates of HEK293T (WT) and three double knockout (KO) samples, X and Y axes correspond to the first two principal components. **(B)** Barplot illustrating FTO and ALKBH5 expression in counts per million (CPM). **(C)** Volcano plot of differential expression analysis. The red line indicates 5% FDR, gray lines indicate -1 and 1  $\log_2$ (Fold Change).

modified eagle medium (DMEM), supplemented with 10% fetal bovine serum (FBS, Capricorn Scientific), 2 mM glutamine, and 1x Antibiotic-Antimycotic (Gibco) at 37°C in a humidified atmosphere containing 5% CO<sub>2</sub> and passaged by standard methods.

The gRNAs were designed using CHOPCHOP software [8]. For the FTO knockout, two gRNAs targeting the third exon (top strand: 5'-CTACAAGTACCTGAACACC-3' and bottom strand: 5'-CCAATGAGGATGCGAGATAC-3') were inserted into pSpCas9n(BB)-2A-Puro(PX462) V2.0 (Addgene #62987, Feng Zhang lab [9]). For ALKBH5 deletion, the gRNAs targeting the first and the third



**Fig. 3.** The data are suitable for analysis of differential exon usage under FTO and ALKBH5 knockout. (A) Volcano plot of differential splicing analysis. The red line indicates 5% FDR, several illustrative examples of alternative splicing are highlighted. (B-F) Genomic views of RNA-Seq profiles for the selected genes with alternative splicing events.

**Table 2**

Primers used for the ALKBH5 donor DNA cloning and amplification. Down – downstream homology arm, up – upstream homology arm, RE – restriction enzyme (bold). For full-length donor DNA amplification, For\_up and Rev\_down primers were used.

Primer name	Sequence	RE	Cloning strategy
For_down	AGGAAACAACAGGGACCTGCTCTGAAACC		Inserted into pJet1.2 at EcoRV in the forward orientation
Rev_down	TACCCACTTTCTCTCTTGGGCTCC		
For_up	AAGCGGCCCGCGGCACGCGCACTGTGG	NotI	Inserted into pJet1.2 with downstream homology arm at NotI/XhoI
Rev_up	TTCTCGAGACGCAGGTCCGTAGCCGCTGG	XhoI	

exon (5'-GACGTCCCGGACAATA-3' and 5'-TCGCTGTGAGAAACA-3') were inserted into pSpCas9(BB)-2A-Puro(pX459) V2.0 (Addgene #62988, Feng Zhang lab [9]). To obtain the ALKBH5 donor vector, upstream and downstream homology arms (about 1000-bp each) were amplified and inserted into the pJet1.2 vector (Table 2).

Cells were transfected with the gRNA expression vectors (for FTO) or with the amplified donor sequence and the gRNA expression vectors (for ALKBH5) using Lipofectamine 3000 (Invitrogen). After puromycin treatment (1  $\mu$ g/ml) for three days, the cells were allowed to recover for two days, and then the pool of survived cells was cloned by limiting dilution and screened by immunoblotting with anti-ALKBH5 (#ab195377, Abcam), anti-FTO (#31687, CST), and anti-ACTB (loading control, #ab8227, Abcam) antibodies as was described previously [10]. The double knockout  $\Delta$ ALKBH5 $\Delta$ FTO cells were obtained by consecutive knockout of FTO and ALKBH5, correspondingly.

## 2.2. RNA isolation and library preparation

The cells were lysed in buffer containing 20 mM Tris-HCl (pH 7.4), 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM DTT, 1% Triton X-100, 100 µg/ml cycloheximide (Sigma-Aldrich), 25 U/ml TURBO DNase (Ambion). Cell lysates were incubated on ice for 10 min, triturated ten times through a 26-G needle, and centrifuged at 20,000 g at +4°C for 10 min. Total RNA was isolated using TRIzol LS Reagent (Thermo Fisher Scientific) and was subsequently purified on columns Direct-zol RNA Miniprep (Zymo Research). The 1 µg of the total RNA supplemented with *in vitro* transcribed spike-in RNAs [10] (7 pg GFP mRNA and 1.2 pg FLUC mRNA) was used for library preparation. The experiment was performed in three replicates for wild type HEK293T and double knockout cells.

The RNA-seq libraries were constructed according to the manufacturer's protocol using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB) and RiboMinus Eukaryote System v2 (Thermo Fisher Scientific) for rRNA depletion.

## 2.3. Sequencing and raw data processing

The RNA-Seq libraries were sequenced using Illumina NexSeq 500 (76 bp single-end reads). The reads were trimmed with cutadapt v2.10 [11] with "-j 24 -minimum-length 20 -nextseq-trim=20" parameters. Read mapping was performed using the hg38 genome assembly and GENCODE v29 genome annotation (modified to include spike-ins as additional contigs) was performed with STAR v2.7.6a [12] using default parameters.

## 2.4. Gene counting, differential expression, and exon usage analyses

Differential expression (DE) analysis was performed to estimate overall transcriptomic changes. The read counting was performed by R featureCounts from RsubRead (v.2.0.1) [13] using GENCODE v29 [14] basic annotation with GTF.featureType = 'gene', and countMultiMappingReads = FALSE to leave only uniquely mapped reads that overlap exons or introns. Next, we used edgeR (v.3.28.1) [15] to identify differentially expressed genes among 13718 genes that passed 2 cpm for at least 3 of 6 samples for DE (default TMM-normalization with calcNormFactors, dispersion estimation with estimateDisp, and test for differential expression with exactTest).

To test the genes for differential exon usage (DEU), the exon-level reads counting was performed with featureCounts allowing multiple feature overlaps (allowMultiOverlap = TRUE). F-test for differential splicing (diffSpliceDGE) was used for detection of significant alternative splicing events.

RNA-Seq profiles for visualization were generated using samtools (v.1.10) [16] view with -f/-F 16 and -q 255 flags to leave only correctly stranded and uniquely mapped reads followed by bedtools (v.2.27.1) [17] genomcov with -split and -scale [10<sup>6</sup> / total number of uniquely mapped reads]. The genomic views were generated with svist4get (v.1.2.24) [18], the other plots were generated using ggplot2 (v.3.3.5).

## Ethics Statements

This work does not contain any studies with humans, animals or social media platforms.

## 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.

## Data Availability

Transcriptome (RNA-Seq) of HEK293T, HEK293T €ALKBH5€FTO cells (Original data) (NCBI Gene Expression Omnibus (GEO)).

## CRedit Author Statement

**Egor A. Smolin:** Investigation, Writing – review & editing, Visualization; **Andrey I. Buyan:** Writing – original draft, Software, Visualization, Investigation; **Dmitry N. Lyabin:** Writing – review & editing, Investigation; **Ivan V. Kulakovskiy:** Writing – original draft, Formal analysis, Data curation, Supervision; **Irina A. Eliseeva:** Conceptualization, Writing – original draft, Supervision, Funding acquisition, Data curation.

## Acknowledgments

This work was supported by Russian Science Foundation [grant number 19-74-10079]. We are grateful to the Skoltech Genomics Core Facility and A.A. Kharkevich Institute for Information Transmission Problems RAS for assistance with high-throughput sequencing.

## References

- [1] P.C. He, C. He, m 6 A RNA methylation: from mechanisms to therapeutic potential, *EMBO J* 40 (2021), doi:[10.15252/embj.2020105977](https://doi.org/10.15252/embj.2020105977).
- [2] M. Bartosovic, H.C. Molaes, P. Gregorova, D. Hrossova, G. Kudla, S. Vanacova, N6-methyladenosine demethylase FTO targets pre-mRNAs and regulates alternative splicing and 3'-end processing, *Nucleic Acids Res.* 45 (2017) 11356–11370, doi:[10.1093/nar/gkx778](https://doi.org/10.1093/nar/gkx778).
- [3] J. Wang, Y. Li, P. Wang, G. Han, T. Zhang, J. Chang, R. Yin, Y. Shan, J. Wen, X. Xie, M. Feng, Q. Wang, J. Hu, Y. Cheng, T. Zhang, Y. Li, Z. Gao, C. Guo, J. Wang, J. Liang, M. Cui, K. Gao, J. Chai, W. Liu, H. Cheng, L. Li, F. Zhou, L. Liu, Y. Luo, S. Li, H. Zhang, Leukemogenic chromatin alterations promote AML leukemia stem cells via a KDM4C-ALKBH5-AXL signaling axis, *Cell Stem Cell* 27 (2020) 81–97 .e8, doi:[10.1016/j.stem.2020.04.001](https://doi.org/10.1016/j.stem.2020.04.001).
- [4] M.A. Garcia-Campos, S. Edelheit, U. Toth, M. Safra, R. Shachar, S. Viukov, R. Winkler, R. Nir, L. Lasman, A. Brandis, J.H. Hanna, W. Rossmanith, S. Schwartz, Deciphering the "m6A Code" via antibody-independent quantitative profiling, *Cell* 178 (2019) 731–747 .e16, doi:[10.1016/j.cell.2019.06.013](https://doi.org/10.1016/j.cell.2019.06.013).
- [5] Z. Zhang, L.Q. Chen, Y.L. Zhao, C.G. Yang, I.A. Roundtree, Z. Zhang, J. Ren, W. Xie, C. He, G.Z. Luo, Single-base mapping of m 6 A by an antibody-independent method, *Sci. Adv.* 5 (2019), doi:[10.1126/sciadv.aax0250](https://doi.org/10.1126/sciadv.aax0250).
- [6] S. Zou, J.D.W. Toh, K.H.Q. Wong, Y.G. Gao, W. Hong, E.C.Y. Woon, N 6 -Methyladenosine: a conformational marker that regulates the substrate specificity of human demethylases FTO and ALKBH5, *Sci. Rep.* 6 (2016) 25677, doi:[10.1038/srep25677](https://doi.org/10.1038/srep25677).
- [7] C. Tang, R. Klukovich, H. Peng, Z. Wang, T. Yu, Y. Zhang, H. Zheng, A. Klungland, W. Yan, ALKBH5-dependent m6A demethylation controls splicing and stability of long 3'-UTR mRNAs in male germ cells, *Proc. Natl. Acad. Sci. U. S. A.* 115 (2017) E325–E333, doi:[10.1073/pnas.1717794115](https://doi.org/10.1073/pnas.1717794115).
- [8] K. Labun, T.G. Montague, M. Krause, Y.N. Torres Cleuren, H. Tjeldnes, E. Valen, CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing, *Nucleic Acids Res.* 47 (2019) W171–W174, doi:[10.1093/nar/gkz365](https://doi.org/10.1093/nar/gkz365).
- [9] F.A. Ran, P.D. Hsu, J. Wright, V. Agarwala, D.A. Scott, F. Zhang, Genome engineering using the CRISPR-Cas9 system, *Nat. Protoc.* 8 (2013) 2281–2308, doi:[10.1038/nprot.2013.143](https://doi.org/10.1038/nprot.2013.143).
- [10] D.N. Lyabin, I.A. Eliseeva, E.A. Smolin, A.N. Doronin, K.S. Budkina, I.V. Kulakovskiy, L.P. Ovchinnikov, YB-3 substitutes YB-1 in global mRNA binding, *RNA Biol.* 17 (2020) 487–499, doi:[10.1080/15476286.2019.1710050](https://doi.org/10.1080/15476286.2019.1710050).
- [11] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, *EMBnet. J.* 17 (2011) 10, doi:[10.14806/ej.17.1.200](https://doi.org/10.14806/ej.17.1.200).
- [12] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: Ultrafast universal RNA-seq aligner, *Bioinformatics* 29 (2013) 15–21, doi:[10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635).
- [13] Y. Liao, G.K. Smyth, W. Shi, The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads, *Nucleic Acids Res.* 47 (2019) e47 –e47, doi:[10.1093/nar/gkz114](https://doi.org/10.1093/nar/gkz114).
- [14] A. Frankish, M. Diekhans, A.-M. Ferreira, R. Johnson, I. Jungreis, J. Loveland, J.M. Mudge, C. Sisu, J. Wright, J. Armstrong, I. Barnes, A. Berry, A. Bignell, S. Carbonell Sala, J. Chrast, F. Cunningham, T. Di Domenico, S. Donaldson, I.T. Fiddes, C. García Girón, J.M. Gonzalez, T. Grego, M. Hardy, T. Hourlier, T. Hunt, O.G. Izuogu, J. Lagarde, F.J. Martin, L. Martínez, S. Mohanan, P. Muir, F.C.P. Navarro, A. Parker, B. Pei, F. Pozo, M. Ruffier, B.M. Schmitt, E. Stapleton, M.-M. Suner, I. Sycheva, B. Uszczyńska-Ratajczak, J. Xu, A. Yates, D. Zerbino, Y. Zhang, B. Aken, J.S. Choudhary, M. Gerstein, R. Guigó, T.J.P. Hubbard, M. Kellis, B. Paten, A. Reymond, M.L. Tress, P. Flicek, GENCODE reference annotation for the human and mouse genomes, *Nucleic Acids Res.* 47 (2019) D766–D773, doi:[10.1093/nar/gky955](https://doi.org/10.1093/nar/gky955).
- [15] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, *Bioinformatics* 26 (2010) 139–140, doi:[10.1093/bioinformatics/btp616](https://doi.org/10.1093/bioinformatics/btp616).

- [16] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome project data processing subgroup, the sequence alignment/map format and SAMtools, *Bioinformatics* 25 (2009) 2078–2079, doi:[10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352).
- [17] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, *Bioinformatics* 26 (2010) 841–842, doi:[10.1093/bioinformatics/btq033](https://doi.org/10.1093/bioinformatics/btq033).
- [18] A.A. Egorov, E.A. Sakharova, A.S. Anisimova, S.E. Dmitriev, V.N. Gladyshev, I.V. Kulakovskiy, svist4get: a simple visualization tool for genomic tracks from sequencing experiments, *BMC Bioinf.* 20 (2019) 113, doi:[10.1186/s12859-019-2706-8](https://doi.org/10.1186/s12859-019-2706-8).