

**Supplemental information**

**Sites of transcription initiation drive mRNA**

**isoform selection**

**Carlos Alfonso-Gonzalez, Ivano Legnini, Sarah Holec, Laura Arrigoni, Hasan Can Ozbulut, Fernando Mateos, David Koppstein, Agnieszka Rybak-Wolf, Ulrike Bönisch, Nikolaus Rajewsky, and Valérie Hilgers**

## Supplemental Tables

**Table S1. Summary of the total number of reads collected per tissue and LRS method. Related to Figures 1 and 6.**

Organism	Drosophila				Human
Tissue	Heads	Embryos 14-16h	Embryos 18-20h	Ovaries	Brain organoids
<b>Nanopore (ONT cDNA)</b>					
Standard	828,444				
BluePippin	39,614,269	5,091,510	3,713,022	4,366,849	5,287,083
<b>Direct RNA (ONT DRS)</b>					
Standard	1,028,763	5,111,208		4,623,486	
<b>FLAM-seq</b>					
Standard	941,167				1,046,039*
<b>Iso-seq</b>					
Standard	20,807				
BluePippin	13,837				

\* Data from (Rybak-Wolf et al., 2021).

**Table S2. Number of isoforms recovered from ONT cDNA and other long-read sequencing methods in each step of the CIA assembly. Related to Figures 1 and 6.**

Method and tissue	Number of isoforms found	Number of isoforms also found with other methods	Number of unique isoforms
<b>Embryos 14-16h ONT cDNA</b>	13,326	10,516	2,810
<b>Embryos 14-16h ONT DRS</b>	3,165	2,942	223
<b>Embryos 18-20h ONT cDNA</b>	11,930	9,602	2,328
<b>Heads ONT cDNA</b>	48,115	15,475	32,640
<b>Heads FLAM-seq</b>	4,845	4,344	501
<b>Heads Iso-seq</b>	3,940	3,608	332
<b>Heads ONT DRS</b>	1,090	1,016	74

Ovaries ONT cDNA	10,187	8,204	1,983
Ovaries ONT DRS	4,043	3,628	415
Human brain organoids ONT cDNA	58,059	n/a	n/a

**Table S3. Number of genes and distinct transcripts (isoforms) recovered from ONT cDNA data in each step of the CIA assembly. Related to Figures 1 and 6.**

Filtering steps. N genes (N isoforms)			
Tissue	Assembly without correction	Assembly after 5' filtering	Assembly after 5' filtering and 3' filtering
Heads	26,807 (138,425)	15,873 (74,482)	9,010 (51,613)
Embryos 14-16h	14,868 (36,107)	8,497 (20,701)	6,775 (15,027)
Embryos 18-20h	12,896 (27,191)	8,877 (19,262)	6,261 (12,356)
Ovaries	9,225 (19,890)	7,255 (13,398)	6,351 (11,376)
Organoids	44,786 (126,135)	19,782 (85,855)	12,470 (58,059)

**Table S5. Drosophila CRISPRa lines tested. Related to Figure 2.**

List of 23 used TRiP-OE lines (Ewen-Campen et al., 2017; Zirin et al., 2020) and result of TSS induction. Each sgRNA line targets the upstream TSS of the indicated genes of interest. sgRNA lines were crossed with flies expressing dCas9-VPR under control of Tubulin-Gal4 (Tub>dCas9-VPR) to induce expression from TSSs linked to a specific 3' end. Of twelve crosses designed to induce nervous-system enriched TSSs, induction was achieved for six (successful), induction was not achieved for five (failed), and one was lethal. Of eleven crosses designed to induce ovary-enriched TSSs, induction was not achieved for either, and three were lethal.

#	BDSC number	Targeted gene	sequence sgRNA1*	sequence sgRNA2*	Phenotype	Induction status
<b>Nervous-system enriched TSS</b>						
1	78119	<i>Mvl</i>	AAATCTTAGGCG GTATTTCTGGG	ACATTTCTAACGG TAACCTGCGG	linked 3' end induced	successful
2	78207	<i>ttv</i>	GGGAAAGCGAAT ATGGTGGAAGG	GAAGATTGGAATG CATCGCTGGG	linked 3' end induced	successful
3	78287	<i>ttk</i>	CAAGAATCGAAC ATTTGAATGGG	CAAGAATCGAACA TTTGAATGGG	linked 3' end induced	successful
4	78649	<i>csu</i>	GCACAGAATGCT CAGAATGATGG	CACAACGTTTCGA AAATAGCGGG	no	failed

5	79440	<i>Fatp1</i>	GCACAGAATGCT CAGAATGATGG	CACAACGTTTCGA AAATAGCGGG	linked 3' end induced	successful
6	79461	<i>wun</i>	ACTGTTTACAGTT ATGGACTTGG	AACAGTCTGCAAA TTGTAGGAGG	linked 3' end induced	successful
7	79798	<i>zfh1</i>	GCCCGAATCACC TCGTACATCGG	ACGGATTCGGTAG AGGACGTCGG	no	failed
8	79871	<i>chn</i>	ACGAAATATGGT AAAACTAAGG	TGCTTAATTTAGG AGCTCAATGG	linked 3' end induced	successful
9	79903	<i>sbb</i>	ACATACATAACTG CGCAATATGG	GAGCGGCATTTCT TTAATGGCGG	no	failed
10	79908	<i>twin</i>	TTAAAAAAGGTT AATATGTCGG	TTAAGGGCATTCTG CGGCTAATGG	no	failed
11	80271	<i>jing</i>	ACTTGCCACCGC GGACACATCGG	CCCCACGCGACC CTCTCTCTTGG	lethal	n/a
12	82755	<i>psq</i>	TTTGCTACGGAC TTTGATATGGG	TTTAATCCTGGAA GCACTTATGG	no	failed
<b>Ovary-enriched TSS</b>						
13	78127	<i>CASK</i>	CTTAGATACCAA CGGGTTTCTGG	CGGAACCTGGAGT CGGTTACCGG	no	failed
14	78295	<i>sky</i>	ACACTCTCCAC AAAACGCGG	CAGGAAAACCGAA CGAACGTTGG	no	failed
15	78595	<i>Pka-R1</i>	GGTAAATAGTCC GAAAAAGAGGG	AGATAAAGCAGT AGCAATGTGG	no	failed
16	79516	<i>Pdp1</i>	TTAGTCAGAGTC GTTTCGTAGTGG	CTGCGACTCTCCG TTACATAAGG	no	failed
17	79673	<i>SPoCk</i>	GTGTGGCCACGA TTGGGCCGAGG	ATATTGGCAACGT GTGATCACGG	no	failed
18	79863	<i>Mef2</i>	CTTTAATTTAAAA AATTTCCAGG	TTCTACAACCATA AGTACATTGG	no	failed
19	79900	<i>brat</i>	GTCAATGTTGGT ATGACACTTGG	TCGATTTAAGTAG CTAAGGACGG	no	failed
20	79987	<i>REPTOR</i>	CAGGTGGCACGT CGATTTGGCGG	GGAAAATAAAGAA TGTAATAGG	no	failed
21	80516	<i>E2f1</i>	CGTGAGCCGATG TTATTTAGTGG	n/a	lethal	n/a
22	80517	<i>Stat92E</i>	ATCTAACAATTAG AGCATGTCGG	n/a	lethal	n/a
23	84063	<i>gfzf</i>	GTTTAGACGCCG TACGCGTAGGG	n/a	lethal	n/a

\* sgRNA1 and sgRNA2 sequences from the TRiP-OE-VPR collection available on Flybase (<http://flybase.org/reports/FBic0003381.html>). Where only one sgRNA sequence is mentioned, the line used is from an improved version, the TRiP-OE-flySAM collection, a set of transgenic constructs that each consist of a single sgRNA with MS2 loops (<http://flybase.org/reports/FBic0003482.html>).

## References

- Rybak-Wolf, A., Wyler, E., Legnini, I., Loewa, A., Glažar, P., Kim, S.J., Pentimalli, T.M., Martinez, A.O., Beyersdorf, B., Woehler, A., et al. (2021). Neurodegeneration in human brain organoids infected with herpes simplex virus type 1. *bioRxiv*. 10.1101/2021.03.05.434122.
- Ewen-Campen, B., Yang-Zhou, D., Fernandes, V.R., González, D.P., Liu, L.P., Tao, R., Ren, X., Sun, J., Hu, Y., Zirin, J., et al. (2017). Optimized strategy for in vivo Cas9-activation in *Drosophila*. *Proc Natl Acad Sci U S A* 114, 9409-9414. 10.1073/pnas.1707635114.
- Zirin, J., Hu, Y., Liu, L., Yang-Zhou, D., Colbeth, R., Yan, D., Ewen-Campen, B., Tao, R., Vogt, E., VanNest, S., et al. (2020). Large-Scale Transgenic *Drosophila* Resource Collections for Loss- and Gain-of-Function Studies. *Genetics* 214, 755-767. 10.1534/genetics.119.302964.