VIDEO - Visual Integration of Drosophila Enhancer Organization: A tool for integrating and visualizing chromatin accessibility, in vivo transcription factor binding and motif occurrence in tissue-specific differentially expressed genes

Running title: A Tool for Visualizing Drosophila Proximal Enhancers

Vidya Ajay[1,2], Nathaniel Laughner[1], Patrick Cahan[2] and Deborah J. Andrew[1]

[1]Department of Cell Biology and [2]Department of Biomedical Engineering

725 N. Wolfe St., Hunterian G1, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

Correspondence: patrick.cahan@jhmi.edu and dandrew@jhmi.edu

Key words: Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC seq), Chromatin Immunoprecipitation Sequencing (ChIP-Seq), DNA binding motifs, promoter proximal enhancers, single cell RNA-Seq (scRNA-Seq)

ABSTRACT

Dissecting gene regulation today relies on many genomic assays – including transcriptional output from RNA-seq, chromatin accessibility from ATAC-seq, and transcription factor (TF) binding from ChIP-seq. Whereas numerous tools exist for each modality and some integrate data across modalities, few allow researchers to interactively explore and visualize how TF binding motifs intersect with transcriptional activity and chromatin accessibility in a tissue-specific context. Here we introduce VIDEO (Visual Integration of Drosophila Enhancer Organization), a web-based analysis tool that enables visualization of conserved TF binding motifs within proximal promoters of genes differentially expressed in specific tissues. Starting with gene lists derived from in situ hybridization, microarray, and/or scRNA-seq studies of WT or mutant samples, one can identify the TFs expressed in each tissue and learn if and where the consensus binding motifs for those TFs are found within the proximal enhancers of a custom gene set. This pipeline also allows for coincident visualization of active chromatin, as determined from ATAC-seq data, and for the visualization of DNA binding data from ChIP-seq datasets for specific TFs. To demonstrate its utility, we apply VIDEO to the well-characterized regulatory system of CrebA and the secretory pathway in the *Drosophila* salivary gland. We also explore a lesser-known system in the embryonic hindgut to show how utilization of this tool can serve to generate hypotheses regarding regulatory interactions.

52 **INTRODUCTION**

53 The application of recent advances in high-throughput sequencing have revealed large sets of
54 differentially expressed (DE) genes from multiple *Drosophila* cell types at distinct developmental stages
55 and under a variety of mutant conditions (KARAISKOS *et al.* 2017; CALDERON *et al.* 2022; SEROKA *et al.* 2022;
56 SAKAGUCHI *et al.* 2023; PENG *et al.* 2024; JACKSON *et al.* 2025; WANG *et al.* 2025). Related technologies have
57 likewise generated extensive *Drosophila* chromatin accessibility and transcription factor (TF) binding
58 datasets with associated binding site consensus motif data (BERGMAN *et al.* 2005; ZHU *et al.* 2011;
59 SHAZMAN *et al.* 2014; NITTA *et al.* 2015; KUDRON *et al.* 2024; RAULUSEVICIUTE *et al.* 2024). These motifs, 6-12
60 bp long sequences recognized by specific transcription factors, are a critical component of gene
61 regulatory networks, and mapping their patterning across the genome is crucial to formulating
62 hypotheses about TF-target gene interactions. A challenge that remains is connecting these datasets to
63 identify which TFs may regulate a given set of DE genes and where their binding motifs occur in open
64 chromatin.
65
66 Existing workflows for motif analysis extract peak regions from ATAC-seq (open chromatin) and ChIP-seq
67 (TF occupancy) data, convert them to FASTA sequences, and run motif discovery tools such as MEME
68 (BAILEY AND ELKAN 1994), HOMER (HEINZ *et al.* 2010) and RSAT (THOMAS-CHOLLIER *et al.* 2012). Single-cell
69 RNA-seq further expands possible motif-analysis inputs by generating lists of genes of interest from
70 differential expression analysis. For all inputs, users must retrieve the peak or promoter/enhancer
71 sequences, wrangle multiple file formats, and merge motif hits with chromatin or binding tracks using
72 command-line tools (e.g. bedtools) and write custom scripts for visualization. This fragmented, multi-
73 step workflow and the lack of user-friendly visualization tools often leaves motif data under-utilized in
74 the context of gene regulatory network studies, and there is still a lack of an intuitive way to visualize
75 and explore TF-motif landscapes in their native genomic context.
76
77 Existing visualization tools address parts of the pipeline; MEME-Suite (BAILEY *et al.* 2015) allows for a
78 basic position display for motifs it discovers but does not accommodate externally defined motifs.
79 TFMotifView (LEPORCQ *et al.* 2020) allows browsing of a pre-computed motif library in user-defined
80 genomic regions, but it does not take as input user-generated motifs in standard formats (PFM, PWM,
81 PCM, IUPAC as described in **Supplemental Table S1**) and it does not support the latest *Drosophila*
82 genome assemblies. Additionally, the output of these existing tools fails to utilize the potential of
83 overlaying TF motifs with user-discovered gene lists and sequencing data.

To address these shortcomings, we have built VIDEO, a web-based platform designed specifically for use with *Drosophila,* with a streamlined, user-friendly graphical interface that requires no programming knowledge. VIDEO integrates processed ATAC-seq and ChIP-seq peak information into signal tracks with genomic input, integrates motif discovery while also accepting user-defined motifs in any common format, and displays motif hits alongside genomic signal overlays. VIDEO allows users to rapidly explore many transcriptional regulation hypotheses and questions such as: Are the motifs of the TF of interest enriched in the user-provided list? Do they occur more densely in the list compared to a control group? Which isoform/transcription start site (TSS) is the richest in the TF motif? Do motif hits occur in regions of accessible chromatin? Which of these hits are supported by DNA binding data? Currently, there is no single public resource that can be mined to address all these questions. VIDEO also serves as a data bank for users who want to extract processed scRNA sequencing data for *Drosophila* embryos (stage 10-16, (PENG *et al.* 2024) and find TF motifs of interest, enabling both an end-to-end pipeline for initial hypothesis generation, and a much more detailed and focused pipeline for exploring motif occurrences in user-generated gene lists with supporting ChIP-seq and ATAC-seq data. VIDEO can be accessed through this link: VIDEO-motif-tool.

MATERIALS AND METHODS

Input data and preprocessing

User inputs consist of either a gene list of FlyBase IDs or official gene symbols or a BED file with genomic peak regions. In the "compare" mode users can input two gene lists with distinct labels and perform the same downstream analysis. For gene-list inputs, transcript annotations for all isoforms and ±500 bp (or a user-defined window size up to 2 kb) flanking sequences around each TSS were indexed from a copy of the *Drosophila melanogaster* reference genome (dm6) FASTA using pyfaidx (v0.8.1.4). For BED-file inputs, peak coordinates were used directly to extract the corresponding FASTA sequences.

Motif scanning

User-defined motifs (PFM, PWM, PCM, or IUPAC formats) were scanned against the extracted sequences using FIMO (v4.12.0; (GRANT *et al.* 2011)) with a default significance threshold of $p < 1e-4$. A custom Python wrapper launched FIMO as a subprocess, parsed its output, and collated motif hits into a unified table recording genomic coordinates, strand, p-values, q-values, and matched sequences.

Visualization of motif occurrences

Motif occurrences were plotted with Plotly (v6.2.0) in python. For each gene or peak region, hits were positioned relative to the TSS (zero), spanning ±500 bp by default. The opacity of each hit reflects its FIMO match score, with darker hits indicating a stronger match. Hits with negative log-likelihood ratio scores were excluded, as they are more likely to represent random matches and correspond to higher p-values. Motif hits are ordered either by the user's input list or by total motif count or mean match score, as selected in the interface. The plot is interactive and all plotting defaults are configurable in the web user interface (UI). Match stringency is an adjustable variable and users can further adjust the p-values for each motif specifically.

Data Integration

To filter motif hits by chromatin context, ATAC-seq and ChIP-seq peaks were integrated via Pybedtools (v0.12.0) calling bedtools intersect with a minimum overlap of 50% of the motif length.

BigWig Track Overlay

To visualize chromatin accessibility and transcription factor occupancy in the same genomic window as motif hits, BigWig tracks from ATAC-seq or ChIP-seq experiments were uploaded. For each selected peak

or gene, coverage values are extracted from the corresponding chromosome interval using pyBigWig (v0.3.22). Coverage is plotted in Plotly alongside motif positions, with the x-axis centered on the peak midpoint or TSS and consistent y-axis scaling across all uploaded tracks to facilitate comparison. Motif hits from the motif occurrence plot are overlaid in the top panel and strand direction is accounted for.

VIDEO implementation

VIDEO was implemented using FastAPI (v0.95.2) backend and a React.js (v18.2.0) frontend. Core dependencies include FIMO (v4.12.0), pyfaidx (v0.5.9), Pybedtools (v0.12.0), and Plotly (v6.2.0). The application was deployed on an EC2 Instance with Redis as broker. Reference genome FASTA and GTF annotation (dm6, release 6.45) were downloaded from FlyBase; motif libraries were bulk downloaded from JASPAR (RAULUSEVICIUTE *et al.* 2024).

Data and code availability

All the code used for developing the tool is available at https://github.com/vidyaajay1/VIDEO-motif-tool.

Embryo immunostaining

Embryos were fixed and immunostained as described (REUTER *et al.* 1990). Rat αRetn (HABERMAN *et al.* 2003) was used at a 1/20,000 dilution. Biotin-conjugated Donkey αRat secondary was used at a dilution of 1:2000.

157 **RESULTS**

158 VIDEO presents an interactive, browser-based overview designed to guide users from raw inputs to
159 interpretable motif visualizations without the need for multi-tool command-line operations. Here, we
160 provide a flowchart that can be used as a guide for new users (**Fig 1**). Upon starting the application, the
161 landing page has a collapsible sidebar navigator that has two functional pages "Motif Viewer" and "TF
162 Finder" and a tutorial page with instructions (**Fig 2A**). The Motif Viewer has the pipeline for generating
163 the motif occurrence plot, while TF Finder is a helper page with functions intended to provide users with
164 a starting gene list and allow discovery of TFs and motifs of interest.

165

166 **1. Using TF Finder to acquire gene and motif data**

167 *Identification of gene sets of interest*

168 There are multiple ways a user can acquire a gene list of interest. These include, but are not limited to,
169 mining transcriptomic databases such as BDGP (TOMANCAK *et al.* 2007) for genes differentially expressed
170 (DE) in a specific tissue, extracting tissue-specific DE genes from scRNA-seq data, and identifying TF-
171 dependent gene sets from comparisons of WT and mutant scRNA-seq, microarray, and/or in situ data.
172 Users are encouraged to rank their gene lists in order of most important to least if there are available
173 metrics such as p-values or log fold change expression, because VIDEO can retain the input order while
174 plotting. To facilitate very straightforward explorations on DE gene sets, TF Finder contains a curated set
175 of DE genes for two stages of development of the *Drosophila* embryo from Peng et. al 2024 single cell
176 RNA sequencing data (**Fig 2B**). Users can choose the stage of development and the tissue, and download
177 a list of the top DE genes in that tissue.

178

179 *Identify all TFs in a gene list*

180 TF Finder has a helper function that filters any gene list for transcription factors (as reported in FlyBase
181 (OZTURK-COLAK *et al.* 2024)). Users can identify TFs differentially expressed in their tissue of interest by
182 choosing the "Use Top 250 DE Genes for Selected Tissue" option or upload their own gene list to extract
183 the TFs in it (**Fig 2C**). Once the gene list is filtered for TFs, VIDEO also displays the consensus motifs for
184 each TF (if available) in the JASPAR database with their motif ID, PWM and logo.

185

186 *Identification of TF motifs or DE gene conserved motifs*

187 If the user does not have a TF motif of interest already, or if the TF of interest does not have any
188 reported motifs shown in VIDEO, they can make use of the in-built "Motif Discovery for Genes" function

189 (**Fig 2D**). Again, the user can select the "Top 250 DE Genes for Selected Tissue" option or upload their
190 own gene list. Once the gene list is uploaded, the user can set a promoter region flanking the TSS and
191 the minimum and maximum motif width. If nothing is specified, the program will default to its preset
192 values (6 bp to 10 bp width). On clicking the "Run STREME" button, the motif discovery program will
193 output a link to view and download the motifs that are enriched in the list of genes. Once the motifs of
194 interest are decided, it is recommended that the user keep the motifs in PWM format (MEME format). If
195 the motifs were chosen from VIDEO, the displayed PWM can directly be copied to a clipboard and
196 pasted as motif input.
197
198 **2. Using Motif Viewer to create motif hit plots**
199 *Setting up genomic input*
200 This is the landing page of VIDEO, and in the regular mode users can upload their gene list or a BED file
201 with genomic regions of interest (**Fig 3A**). In the compare mode, users are required to upload two gene
202 lists and give each list a unique name, as elaborated further in the case study later. Once the gene lists
203 are uploaded, the next step is to choose a promoter window size flanking the TSS of each gene. If a gene
204 has multiple TSSs, all unique TSSs will be considered and annotated with the transcript ID. If no choice is
205 made, the program will default to the preset 500 bp upstream and downstream. The last step is to click
206 on the "Process Genomic Input" button, which validates the gene list and fetches all the transcript
207 annotations and sequences for downstream processing.
208
209 *Setting up motif input*
210 The motif input page requires users to drop in their motifs in any of three formats (IUPAC consensus
211 sequence, PWM, PFM, or PCM), give the motif a name, and choose a color for it for the motif
212 occurrence plot (**Fig 3B**). If users are dropping in PWM, PFM or PCM, they can either upload it as a .txt
213 file or copy and paste the matrix and then click "Parse Pasted Matrix". Once the matrix is parsed and all
214 the motifs are input, clicking the "Process Motifs" button will both validate the motifs and process it for
215 the plot.
216
217 *Scanning for motif occurrences*
218 After the gene list and motif data are given to the program, it is ready for scanning the gene list for motif
219 occurrences. Users can choose the overall p-value threshold for the motif scanning algorithm (default is

220 0.001) (**Fig 3C**). Once the p-value is set, the user should click on the "Scan FIMO" button to scan for
221 motif hits.
222
223 *Integrating ATAC-seq and ChIP-seq data*
224 If the user has supporting chromatin accessibility and/or in vivo TF binding data for their gene lists, they
225 can upload the BED files with the respective peaks into VIDEO. This is the fourth window in VIDEO and is
226 completely optional – click on the "Next" button to proceed to the plot if you don't have this data (**Fig**
227 **3D**). VIDEO will use the ATAC and ChIP data as an option to filter the motif occurrences for hits that
228 occur only in "open chromatin" or hits in "ChIP peak" regions and the filters can be turned on and off
229 during plotting. For this demonstration, we mined scATAC data from (WANG *et al.* 2025) for the hindgut
230 tissue in stages 10-12. Once the data are uploaded, the user should press the "Filter Motif Hits" button
231 and then proceed to the next page.
232
233 *Adjusting the plot settings*
234 This window is an array of configurable plot settings and a motif occurrence plot below it (**Fig 4**). If the
235 users chose to upload ATAC/ChIP data they can toggle the filters on and off by checking the respective
236 options under the "Show only:" heading. The motif occurrence plot will default to the order of the input
237 gene list and will show all the transcripts of each gene, but the user has the option to customize this.
238 First, go to the "Sort by motif:" heading and select the motif you want to prioritize. This will display
239 three options, the first two of which are metrics you need to provide for the sorting. Checking "Number
240 of Hits" will sort the gene list by the number of hits for the selected motif, and checking "Motif Match
241 Score" will sort the gene list by the motif match score of the best scoring motif hit for each gene. By
242 default, all the transcripts of a gene will be clustered together with the top scoring transcript, and the
243 user has the option to choose "Show only best transcripts for selected motif hit", which will show only
244 the top ranking TSS for each gene based on the metrics the user selected for ranking.
245 The last tunable setting is the individual p-value threshold for each motif. Some motifs are more
246 permissive than others and therefore might need further tuning for match stringency, so users have the
247 option to make the p-values for each motif independently tighter. For this, click on the "Motif p-value
248 thresholds" dropdown and drag the slider down the desired p-value. Once all the settings are decided,
249 click on the "Generate Plot" button and the motif plot will be updated accordingly. All the displayed
250 motif hits can be downloaded from the button "Download merged hits". For the compare mode, users

251 can also individually download hits for both genes list as a zip file from the "Download individual hits"
252 button.
253
254 *Interpreting the motif occurrence plot*
255 The motif occurrence plot contains all the input genes stacked and aligned along the TSS (**Fig 4**). VIDEO
256 accounts for transcription direction and reverses the strand so that all the genes have upstream to the
257 left and downstream to the right of the center line (zero). Each row in the plot corresponds to a gene's
258 promoter region, and is named in the GeneSymbol_TranscriptID convention, where the transcript ID is
259 unique to a TSS. The plot is interactive, so users are requested to scroll outside the plot region to go
260 further down the list. Double clicking on the zoomed-in plot can revert it to the original layout, and
261 there are also helper buttons to download the image, zoom in, and reset axes at the top right corner of
262 the plot. Hover over any motif hit to see details about the matched sequence, exact coordinates, strand,
263 p-value and match score. A darker motif hit corresponds to a stronger motif match score, and a more
264 transparent hit reflects a weaker match score.
265
266 *Visualizing ATAC-seq and ChIP-seq tracks along with motif hits*
267 The last window of VIDEO is optional and intended for users who wish to visualize ATAC/ChIP-seq signal
268 tracks in the context of the motif hits. To use this tool, users must first upload their BigWig signal tracks
269 of interest (up to three tracks for a plot), give it a name, and pick a color for the track and click Process
270 Tracks. To view the tracks overlaid with the motif hits, select a gene from "Select a peak" dropdown and
271 select the tracks to visualize, and then click on "Generate Overlay". This will generate a plot with the
272 motif hits on top and the signal tracks aligned below it, along the TSS with the strand accounted for
273 (**Supplemental Fig S1**). The y-axis values are all auto-scaled as a group, and users can visualize where the
274 peaks in the signal are with respect to the motif hits. This plot is also interactive, and hovering over the
275 tracks will display the signal values and chromosomal coordinates.
276
277 ***3. Exploring TFs of interest in hindgut tissue***
278 We first sought to use VIDEO explore the transcriptional landscape of the hindgut tissue and identify
279 candidate TFs that make up the regulatory network of hindgut development. Using VIDEO on scRNA-seq
280 data from stage 10-12 hindgut cells, we identified *orthopedia* (*otp*)*, brachyenteron* (*byn*)*, and retained*
281 (*retn*) (aka *dead ringer*) as the top three DE TFs, followed by other hindgut-expressed TFs including *fork*
282 *head* (*fkh*)*, luna, drum* (*drm*)*,* and *brother of odd with entrails limited* (*bowl*). (**Fig 2C**). We plotted the

occurrences of the binding motifs of the top three TFs in the top 50 DE genes (three of which are the TFs themselves) in the hindgut to identify candidate early targets of these TFs.

otp encodes a conserved homeodomain TF required for proper hindgut development (HILDEBRANDT *et al.* 2020), and is strongly expressed in the hindgut as well as in subsets of the CNS (**Supplemental Fig S2**). Its expression in the hindgut is activated by *byn,* where *otp* regulation depends on the number and arrangement of multiple high and low-affinity *byn* motifs (KUSCH *et al.* 2002). In the *otp* promoter, VIDEO identified three *byn* motifs downstream of the TSS, along with one *otp* motif (**Fig 4**).

byn (brachyenteron) encodes a T-box TF expressed in cells that internalize to form the hindgut, where it acts as a broad regulator of hindgut identity (SINGER *et al.* 1996). In the *byn* promoter, VIDEO revealed two *otp* motifs adjacent to the TSS and one *byn* motif downstream. While this arrangement suggests potential autoregulation, and Brachyury orthologs in other species do autoregulate, prior work indicates that *byn* does not autoregulate in *Drosophila* (SINGER *et al.* 1996). The observed *byn* motif may therefore represent a conserved but non-functional site, and experimental validation, such as deleting this motif and assaying *byn* expression would help clarify its role.

The third TF recovered in our analysis was *retn*, which encodes an ARID-domain protein that recognizes AT-rich DNA sequences. *Retn* is expressed in the hindgut epithelium and other tissues such as the amnioserosa, ring gland, glia, etc.; within the hindgut epithelium, it is expressed strongly in the boundary cell rows of the large intestine (IWAKI *et al.* 2001; SHANDALA *et al.* 2002) (**Supplemental Fig S2B**). While its precise role in hindgut development remains unknown, *retn* is required for early embryonic patterning (SHANDALA *et al.* 1999) and for CNS development (SHANDALA *et al.* 2003), where it seems to act in combination with other upstream regulators to modulate gene expression.

Not all motif occurrences reflect true binding events, as some may fall outside accessible chromatin regions or arise by chance. To identify motif hits that are more likely to correspond to regulatory activity in the hindgut, we filtered the motif hits using published scATAC data (WANG *et al.* 2025) focusing on hindgut tissue peaks from stage 10-12 embryos. We broadly restricted peaks to those present in at least 50% of hindgut cells (**Supplemental Fig S3**). Under this cutoff, no motifs were detected in the *retn* promoter, suggesting that the filter was too stringent. Because *retn* expression is limited to a subset of hindgut cells, we relaxed the cutoff to include peaks present in at least 25% of the hindgut tissue. This adjustment restored the expected motif hits in the *retn* promoter, confirming that the restricted expression of *retn* correlates with localized chromatin accessibility. Notably, motifs for *byn, otp,* and *retn* were observed in the *retn* promoter, suggesting possible regulatory interactions that merit further

314 testing in mutant or motif-deletion backgrounds. Further reducing the threshold to 10% did not reveal
315 any additional motifs.
316 Although all the genes in the list are the top differentially expressed in the hindgut based on scRNA
317 annotation (further confirmed by their high expression in the hindgut from in situ hybridization data in
318 BDGP), we found that the promoter regions of 12 genes lacked detectable open chromatin even at the
319 lowest threshold of 10%. This finding seems counterintuitive: highly expressed genes are generally
320 expected to display promoter accessibility, since transcription requires open chromatin. The discrepancy
321 may stem from the technical limitations of scATAC-seq (e.g., the small number of fragments per cell,
322 high sparsity of signal and locus-level dropout), which, although widely used to assay open chromatin,
323 can fail to capture regions where the promoter accessibility is weak, hindered, or transient (CHEN *et al.*
324 2019; LI *et al.* 2021). That said, we were able to identify candidate regulatory events in 75% of the top
325 DE genes in the hindgut, providing a set of hypotheses that can be further tested experimentally.
326
327 ***4. Case study: CrebA motif occurrences in SPCGs***
328 To assess whether VIDEO can recover known transcription factor-target gene relationships, we used
329 CrebA as a test case, asking whether its binding motifs are preferentially enriched and stronger in the
330 promoters of secretory pathway component genes (SPCGs) compared to ribosomal protein genes
331 (RpGs). The SPCGs represent a well-established transcriptional target set for the bZIP TF CrebA in the
332 *Drosophila* salivary gland and other secretory tissues (ABRAMS AND ANDREW 2005; FOX *et al.* 2010; JOHNSON
333 *et al.* 2020; JACKSON *et al.* 2025). CrebA directly regulates genes encoding secretory machinery,
334 enhancing the secretory capacity of embryonic secretory tissues such as the salivary gland and
335 epidermis. It has also been implicated in dendrite development through upregulation of CopII
336 components (IYER *et al.* 2013) and in infection tolerance, where its expression in the fat body is induced
337 by the Toll and Imd pathways to activate secretory pathway genes (TROHA *et al.* 2018).
338
339 For our analysis, we compared 90 SPCGs (JOHNSON *et al.* 2020) to 82 cytoplasmic ribosomal protein
340 genes (RpGs), which serve as a size-matched gene set that, like the SPCGs, is expected to be found in
341 open chromatin based on the basal levels of these factors being required in all cells. This analysis also
342 addresses the question of whether CrebA also increases translation by boosting expression of RPGs, a
343 regulatory role that has been suggested for the mammalian CrebA orthologue Creb3L2 in the
344 mammalian pituitary gland. The VIDEO workflow for this experiment is described in **Fig 5**. Motif scanning
345 revealed that CrebA motifs were substantially more abundant in the proximal promoters of SPCGs than

346 in RpGs (**Fig 6A**). We tuned the p-value thresholds for both motifs so that the core ACGTG motif was
347 retained in all hits. This adjustment revealed at least one CrebA motif within 500 bp of the TSS in 84%
348 (76/90) of SPCGs, whereas only 61% (51/82) of RpGs had a motif hit in the corresponding promoter
349 proximal regions (**Fig 6B**). Further, among the genes that had at least one motif hit, we examined the
350 density of motif occurrences within 500 bp of the TSS (**Fig 6C**). In SPCGs, nearly three-quarters of motif-
351 positive genes carried multiple hits, with 22% containing three and 33% harboring four or more motifs.
352 By contrast, motif-positive RpGs were skewed toward single occurrences, with almost half (47%)
353 carrying only one motif and a much smaller fraction (18%) with four or more motifs.
354 These results indicate that CrebA binding sites are not only more prevalent in SPCGs than RpGs, but also
355 tend to occur in higher local densities near SPCG promoters. This enrichment of clustered motifs
356 suggests that SPCG regulation may involve stronger or more cooperative CrebA binding compared to
357 RpGs. To test if the CrebA motif occurrences showed any particular positional preference, we plotted
358 the positional distribution of the best hit for each motif in each gene, for both gene lists (**Fig 6D**). The
359 IUPAC consensus sequence (CrebA_1, derived from Jackson et al., 2025) displayed a sharp enrichment
360 immediately upstream of the TSS in SPCGs, suggesting that high-confidence CrebA motifs preferentially
361 cluster at promoter-proximal sites in this gene group, whereas in RpGs they lacked a pronounced
362 positional peak. The more permissive PWM-based motif (CrebA_2, from the JASPAR 2024 database)
363 showed relatively flat distributions in both SPCGs and RpGs, consistent with its broader degeneracy and
364 lower information content. Although these trends qualitatively support the idea that SPCGs may be
365 subject to more direct promoter-proximal regulation by CrebA, quantification using Kolmogorov-
366 Smirnov tests did not detect significant positional biases for either motif ($p > 0.05$). Thus, while the
367 observed peak near the TSS in SPCGs is suggestive, we cannot rule out the possibility that it arises from
368 stochastic variation rather than actual positional preference.
369 In addition to motif frequency, motif match quality also differed markedly between the two genes sets.
370 The log-likelihood ratio scores for the motif matches were significantly higher in SPCGs than RpGs,
371 indicating stronger consensus matches among SPCGs (**Fig 7A**). Likewise, log-transformed p-values
372 showed the same trend with SPCGs enriched for significant motif hits compared to RpGs (**Fig 7B**).
373 Interestingly, analysis of motif positions relative to the transcription start sites did not reveal a
374 significant difference between SPCGs and RpGs (**Fig 7C**). This finding is consistent with the positional
375 distribution analysis (Fig. 6D), where CrebA_1 showed a visible promoter-proximal enrichment in SPCGs,
376 but the effect did not reach statistical significance. Together, these data suggest while SPCGs harbor

377 more frequent and higher-quality CrebA motifs, their relative positioning within the 500 bp promoter
378 window is not systematically different from RpGs.
379 Although the precise positional distribution of motifs remains statistically unresolved, likely due to
380 limited sample size (fewer than 90 genes), the overall enrichment and strength of CrebA motifs in SPCGs
381 align with their established regulatory role. These results confirm that VIDEO accurately captures known
382 biology of CrebA regulation, showing abundance, enrichment and stronger motif matches in SPCGs
383 compared to control. This case study illustrates how VIDEO integrates motif scanning, visualization, and
384 tailored data for statistical comparison into a streamlined user-friendly workflow that validates known
385 transcription factor-gene relationships.
386

387 **DISCUSSION**

388
389 Motif analysis remains a cornerstone for connecting TF binding to regulatory function, but practical
390 barriers continue to limit its accessibility. Existing workflows typically require multiple command-line
391 tools, file conversions and custom scripts, making motif analysis challenging for researchers without
392 strong computational backgrounds. Several visualization tools have emerged, but they often impose
393 restrictions on input formats, motif databases, or integration with genomic datasets.
394 VIDEO was developed to address these limitations by offering an integrated, web-based platform
395 tailored to *Drosophila* data. It is the first of its kind to facilitate an intuitive side-by-side comparison of
396 user-defined gene lists, incorporates and chooses the best annotated TSSs for each motif based on user-
397 defined metrics such , and allows for ATAC-seq and ChIP-seq data integration (**Table 1**). It also accepts
398 custom motifs and gene list/BED file inputs and presents all the plots in a graphical, interactive format,
399 with downloadable, organized motif hit datasets. These features significantly lower the barrier for motif
400 exploration and visualization compared to existing tools.

401

402 ***Table 1: Comparison of VIDEO features to other known motif search and visualization tools***

	VIDEO	**TFMotifView**	**MEME-Suite**	**HOMER**
Online web server	Yes	Yes	Yes	No (command-line)
Input format	Gene lists, BED regions, typed genes	BED regions	FASTA sequences	Gene lists, BED regions
Compare gene lists	Yes	No	No	No
Genome	fly	Human, mouse	Available locally	Available locally
Motif database	Custom motifs, JASPAR	JASPAR	Several	Custom
View motif occurrences in-app	Yes	Yes	Limited	No
ATAC/ChIP data integration	Yes	No	No	No
Includes all TSS	Yes	No	No	No
Scanning algorithm	FIMO	Motif Information Content	FIMO	HOMER

403

404 The CrebA case study highlights VIDEO's ability to recapitulate known biology. SPCGs, a well-
405 characterized set of CrebA targets, displayed both higher motif frequencies and stronger consensus
406 matches to CrebA than control RpGs. While the motif positioning relative to the TSS did not differ
407 significantly between the two sets, this could be insightful about CrebA possibly not having a significant
408 preference for a certain position for binding to its target genes provided the binding motif lies within the
409 regulatory window.
410 Beyond this validation, VIDEO provides a flexible framework that is also suited for hypothesis
411 generation, with helper functions in TF Finder designed to guide users in exploring the transcriptional
412 landscape of the early *Drosophila* embryo. The application of VIDEO to stage 10-12 hindgut data
413 illustrates how visualizing motif occurrences with chromatin accessibility in the right context can be
414 useful for generating hypotheses about gene regulatory networks. The presence of *otp* motifs adjacent
415 to the *byn* TSS, and *byn* motifs downstream of *byn* itself, suggests possible autoregulatory or cross-

416 regulatory interactions. While autoregulation by Brachyury homologs has been demonstrated in other
417 species, it has been excluded for *Drosophila byn* (SINGER *et al.* 1996), raising the possibility that these
418 sites are conserved but vestigial or that autoregulation may occur in different developmental contexts.
419 VIDEO thus highlights candidate motifs that can be directly tested by mutation to clarify their
420 functionality. Motif filtering by ATAC-seq data revealed the importance of thresholding stringency when
421 studying genes expressed in rarer cell types, and demonstrates how overly restrictive filters can obscure
422 possibly relevant motifs in genes in specialized subpopulations.
423 By combining motif scanning, visualization, and data integration into a single user-friendly tool, VIDEO
424 enables broader adoption of motif analysis approaches and expands the potential for researchers to link
425 sequence motifs with regulatory function.
426

ACKNOWLEDGEMENTS

We thank Flybase for the genomic data that was essential for the development and utilization of this tool. We gratefully acknowledge support from the National Institutes of Health (NIH) for support for this research: NIH RO1 DE013899.

FIGURE LEGENDS

Figure 1. Flowchart of many possible pipelines using VIDEO. TF Finder and Motif Viewer are the two pages in VIDEO. Each blue and green box represents a module in the page In Motif Viewer (green), these modules are presented as navigable cards, and in TF Finder (blue) they are shown one below another in the same page. Gray boxes represent user-provided data (not native to VIDEO). Yellow boxes represent final outputs from VIDEO. Arrows represent a sequential order of execution of the modules. Arrows start from a module that gives an output and go towards a module that can use this output as its input. Dashed lines represent a hierarchical connection between the pages and their modules.

Figure 2. Steps for Using TF finder. (A) The collapsible sidebar navigator in VIDEO with two functional pages and a tutorial page. **(B)** Getting DE gene lists from TF Finder. Top: before choosing a developmental stage and tissue type. Bottom: after making the selections and pressing Get DE genes. **(C)** Filtering gene lists for TFs. Top: After pressing Filter TFs, Bottom: Motif details after pressing View PWM **(D)** Discovering motifs from gene lists. Top: Choose parameters and upload gene list before running STREME. Bottom: After pressing Run STREME, with link to outputs and a button to download the discovered motifs.

Figure 3. Steps for Using Motif Viewer. **(A)** Set up genomic input. Left: Before uploading gene list, Right: After uploading data, choosing window size and pressing Process Genomic Input. **(B)** Set up motif input. Top: Initial window, Middle: Window after selecting the PWM/PFM format. Paste the motif matrix, press Parse Pasted Matrix and give the motif a name and color. Bottom: After three motifs are added and Process Motifs is pressed. **(C)** Scan for motif hits. User sets the p-value threshold and presses Scan

458 Motifs (FIMO). **(D)** Add ATAC/ChIP-seq data if available. Top: Before uploading data, Bottom: After
459 uploading ATAC data (Wang et al. 2025) and clicking on the Filter and Score Motif Hits button.
460
461 **Figure 4. The motif occurrence plot**. Motif occurrences of *otp, byn* and *retn* the top 50 DE genes in the
462 hindgut. Plot settings configured to sort the genes by *otp* hits and match score, and showing only the
463 best scoring transcript per gene. The configured filtered hits can be downloaded using the Download
464 hits (CSV) button. Each plot corresponds to the plot settings above it. **(A)** Genes ranked by *otp*
465 occurrences, scored by number of hits and motif match score. **(B)** Plot further filtered for motif hits only
466 in open chromatin regions. Note that motifs not in open chromatin disappear, and the genes are sorted
467 by the remaining motif occurrences.
468
469 **Figure 5. Workflow of CrebA motif occurrence case study** (**A)** In the compare mode, input a gene list of
470 interest (SPCGs) and a control (RpGs) as gene symbols, choose a window size of 500bp, then press
471 process genomic input. **(B)** Input the motifs of interest two slightly different forms of the CrebA motif –
472 the IUPAC consensus sequence from Jackson et al. 2025 (CrebA_1, red) and the PWM from JASPAR 2024
473 (CrebA_2, yellow) and choose colors for each. **(C)** Adjust plot settings and tune p-value thresholds for
474 individual motifs and then click on Generate Plot.
475
476 **Figure 6. CrebA motifs are more abundant in SPCGs than RpGs. (A)** Plot shows the occurrence of CrebA
477 motifs in promoter regions of (left) 90 SPCGs and (right) 82 RpGs. Genes are ordered by the number of
478 hits and motif match scores of CrebA_1 motifs. **(B)** The majority of secretory pathway component genes
479 (SPCGs) contain at least one CrebA motif (85%), compared to only 62% of ribosomal protein genes
480 (RpGs). Fisher's exact test confirms a significant enrichment of CrebA motifs in SPCGs (odds ratio = 3.4, p
481 < 0.001). Motif p-value thresholds are CrebA_1: $p<5.25e-4$ CrebA_2: $p<3.24e-4$. **(C)** Of the genes that
482 have at least 1 CrebA motif hit, a larger fraction of the SPCGs had more than 4 hits as compared to RpGs.
483 **(D)** The IUPAC sequence version of the CrebA motif (from Jackson et al., 2025, CrebA_1) shows a peak
484 near at the TSS for SPCGs but not for the RpGs, although on quantification, the preference is not
485 significant (KS test $p = 0.23$). The consensus PWM version of CrebA (from JASPAR 2020, CrebA_2) does
486 not show obvious positional preference for either group.
487
488 **Fig 7: CrebA motif matches are stronger and more significant in SPCGs compared to RpGs. (A)** Motif
489 match scores (log-likelihood ratio scores) are significantly higher in SPCGs (median = 10.84) than RpGs

490 (median = 6.25; *p* = 1.84e-12). **(B)** Motif match log transformed p-values show the same trend, with
491 SPCGs (median = 3.82) enriched for more significant motif hits (*p* = 3.49e-9) as compared to RpGs
492 (median = 3.58). **(C)** Absolute distance to TSS for all motif hits does not vary significantly between SPCGs
493 (median = 202.0) and RpGs (194.5, p = 0.36). p-values are for all three analyses are from Mann Whitney
494 U-test.
495
496
497
498 **References**

499 Abrams, E. W., and D. J. Andrew, 2005 CrebA regulates secretory activity in the Drosophila salivary gland
500 and epidermis. Development 132**:** 2743-2758.
501 Bailey, T. L., and C. Elkan, 1994 Fitting a mixture model by expectation maximization to discover motifs
502 in biopolymers. Proc Int Conf Intell Syst Mol Biol 2**:** 28-36.
503 Bailey, T. L., J. Johnson, C. E. Grant and W. S. Noble, 2015 The MEME Suite. Nucleic Acids Res 43**:** W39-
504 49.
505 Bergman, C. M., J. W. Carlson and S. E. Celniker, 2005 Drosophila DNase I footprint database: a
506 systematic genome annotation of transcription factor binding sites in the fruitfly, Drosophila
507 melanogaster. Bioinformatics 21**:** 1747-1749.
508 Calderon, D., R. Blecher-Gonen, X. Huang, S. Secchia, J. Kentro *et al.*, 2022 The continuum of Drosophila
509 embryonic development at single-cell resolution. Science 377**:** eabn5800.
510 Chen, H., C. Lareau, T. Andreani, M. E. Vinyard, S. P. Garcia *et al.*, 2019 Assessment of computational
511 methods for the analysis of single-cell ATAC-seq data. Genome Biol 20**:** 241.
512 Fox, R. M., C. D. Hanlon and D. J. Andrew, 2010 The CrebA/Creb3-like transcription factors are major and
513 direct regulators of secretory capacity. J Cell Biol 191**:** 479-492.
514 Grant, C. E., T. L. Bailey and W. S. Noble, 2011 FIMO: scanning for occurrences of a given motif.
515 Bioinformatics 27**:** 1017-1018.
516 Haberman, A. S., D. D. Isaac and D. J. Andrew, 2003 Specification of cell fates within the salivary gland
517 primordium. Dev Biol 258**:** 443-453.
518 Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin *et al.*, 2010 Simple combinations of lineage-
519 determining transcription factors prime cis-regulatory elements required for macrophage and B
520 cell identities. Mol Cell 38**:** 576-589.
521 Hildebrandt, K., N. Bach, D. Kolb and U. Walldorf, 2020 The homeodomain transcription factor
522 Orthopedia is involved in development of the Drosophila hindgut. Hereditas 157**:** 46.
523 Iwaki, D. D., K. A. Johansen, J. B. Singer and J. A. Lengyel, 2001 drumstick, bowl, and lines are required
524 for patterning and cell rearrangement in the Drosophila embryonic hindgut. Dev Biol 240**:** 611-
525 626.
526 Iyer, S. C., E. P. Ramachandran Iyer, R. Meduri, M. Rubaharan, A. Kuntimaddi *et al.*, 2013 Cut, via CrebA,
527 transcriptionally regulates the COPII secretory pathway to direct dendrite development in
528 Drosophila. J Cell Sci 126**:** 4732-4745.
529 Jackson, D. J., D. Peng, S. A. Shinde, A. Holenarasipura, V. Ajay *et al.*, 2025 CrebA regulation of secretory
530 capacity: Genome-wide transcription profiling coupled with in vivo DNA binding studies.
531 Genetics.

532 Johnson, D. M., M. B. Wells, R. Fox, J. S. Lee, R. Loganathan *et al.*, 2020 CrebA increases secretory
533 capacity through direct transcriptional regulation of the secretory machinery, a subset of
534 secretory cargo, and other key regulators. Traffic 21**:** 560-577.
535 Karaiskos, N., P. Wahle, J. Alles, A. Boltengagen, S. Ayoub *et al.*, 2017 The Drosophila embryo at single-
536 cell transcriptome resolution. Science 358**:** 194-199.
537 Kudron, M., L. Gevirtzman, A. Victorsen, B. C. Lear, J. Gao *et al.*, 2024 Binding profiles for 961 Drosophila
538 and C. elegans transcription factors reveal tissue-specific regulatory relationships. Genome Res
539 34**:** 2319-2334.
540 Kusch, T., T. Storck, U. Walldorf and R. Reuter, 2002 Brachyury proteins regulate target genes through
541 modular binding sites in a cooperative fashion. Genes Dev 16**:** 518-529.
542 Leporcq, C., Y. Spill, D. Balaramane, C. Toussaint, M. Weber *et al.*, 2020 TFmotifView: a webserver for
543 the visualization of transcription factor motifs in genomic regions. Nucleic Acids Res 48**:** W208-
544 W217.
545 Li, Z., C. Kuppe, S. Ziegler, M. Cheng, N. Kabgani *et al.*, 2021 Chromatin-accessibility estimation from
546 single-cell ATAC-seq data with scOpen. Nat Commun 12**:** 6386.
547 Nitta, K. R., A. Jolma, Y. Yin, E. Morgunova, T. Kivioja *et al.*, 2015 Conservation of transcription factor
548 binding specificities across 600 million years of bilateria evolution. Elife 4.
549 Ozturk-Colak, A., S. J. Marygold, G. Antonazzo, H. Attrill, D. Goutte-Gattat *et al.*, 2024 FlyBase: updates
550 to the Drosophila genes and genomes database. Genetics 227.
551 Peng, D., D. Jackson, B. Palicha, E. Kernfeld, N. Laughner *et al.*, 2024 Organogenetic transcriptomes of
552 the Drosophila embryo at single cell resolution. Development 151.
553 Rauluseviciute, I., R. Riudavets-Puig, R. Blanc-Mathieu, J. A. Castro-Mondragon, K. Ferenc *et al.*, 2024
554 JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding
555 profiles. Nucleic Acids Res 52**:** D174-D182.
556 Reuter, R., G. E. Panganiban, F. M. Hoffmann and M. P. Scott, 1990 Homeotic genes regulate the spatial
557 expression of putative growth factors in the visceral mesoderm of Drosophila embryos.
558 Development 110**:** 1031-1040.
559 Sakaguchi, S., S. Mizuno, Y. Okochi, C. Tanegashima, O. Nishimura *et al.*, 2023 Single-cell transcriptome
560 atlas of Drosophila gastrula 2.0. Cell Rep 42**:** 112707.
561 Seroka, A., S. L. Lai and C. Q. Doe, 2022 Transcriptional profiling from whole embryos to single
562 neuroblast lineages in Drosophila. Dev Biol 489**:** 21-33.
563 Shandala, T., R. D. Kortschak, S. Gregory and R. Saint, 1999 The Drosophila dead ringer gene is required
564 for early embryonic patterning through regulation of argos and buttonhead expression.
565 Development 126**:** 4341-4349.
566 Shandala, T., R. D. Kortschak and R. Saint, 2002 The Drosophila retained/dead ringer gene and ARID gene
567 family function during development. Int J Dev Biol 46**:** 423-430.
568 Shandala, T., K. Takizawa and R. Saint, 2003 The dead ringer/retained transcriptional regulatory gene is
569 required for positioning of the longitudinal glia in the Drosophila embryonic CNS. Development
570 130**:** 1505-1513.
571 Shazman, S., H. Lee, Y. Socol, R. S. Mann and B. Honig, 2014 OnTheFly: a database of Drosophila
572 melanogaster transcription factors and their binding sites. Nucleic Acids Res 42**:** D167-171.
573 Singer, J. B., R. Harbecke, T. Kusch, R. Reuter and J. A. Lengyel, 1996 Drosophila brachyenteron regulates
574 gene activity and morphogenesis in the gut. Development 122**:** 3707-3718.
575 Thomas-Chollier, M., C. Herrmann, M. Defrance, O. Sand, D. Thieffry *et al.*, 2012 RSAT peak-motifs: motif
576 analysis in full-size ChIP-seq datasets. Nucleic Acids Res 40**:** e31.
577 Tomancak, P., B. P. Berman, A. Beaton, R. Weiszmann, E. Kwan *et al.*, 2007 Global analysis of patterns of
578 gene expression during Drosophila embryogenesis. Genome Biol 8**:** R145.

579 Troha, K., J. H. Im, J. Revah, B. P. Lazzaro and N. Buchon, 2018 Comparative transcriptomics reveals
580 CrebA as a novel regulator of infection tolerance in D. melanogaster. PLoS Pathog 14**:** e1006847.
581 Wang, M., Q. Hu, Z. Tu, L. Kong, T. Yu *et al.*, 2025 A Drosophila single-cell 3D spatiotemporal multi-omics
582 atlas unveils panoramic key regulators of cell-type differentiation. Cell 188**:** 4734-4753 e4731.
583 Zhu, L. J., R. G. Christensen, M. Kazemian, C. J. Hull, M. S. Enuameh *et al.*, 2011 FlyFactorSurvey: a
584 database of Drosophila transcription factor binding specificities determined using the bacterial
585 one-hybrid system. Nucleic Acids Res 39**:** D111-117.
586

Figure 1

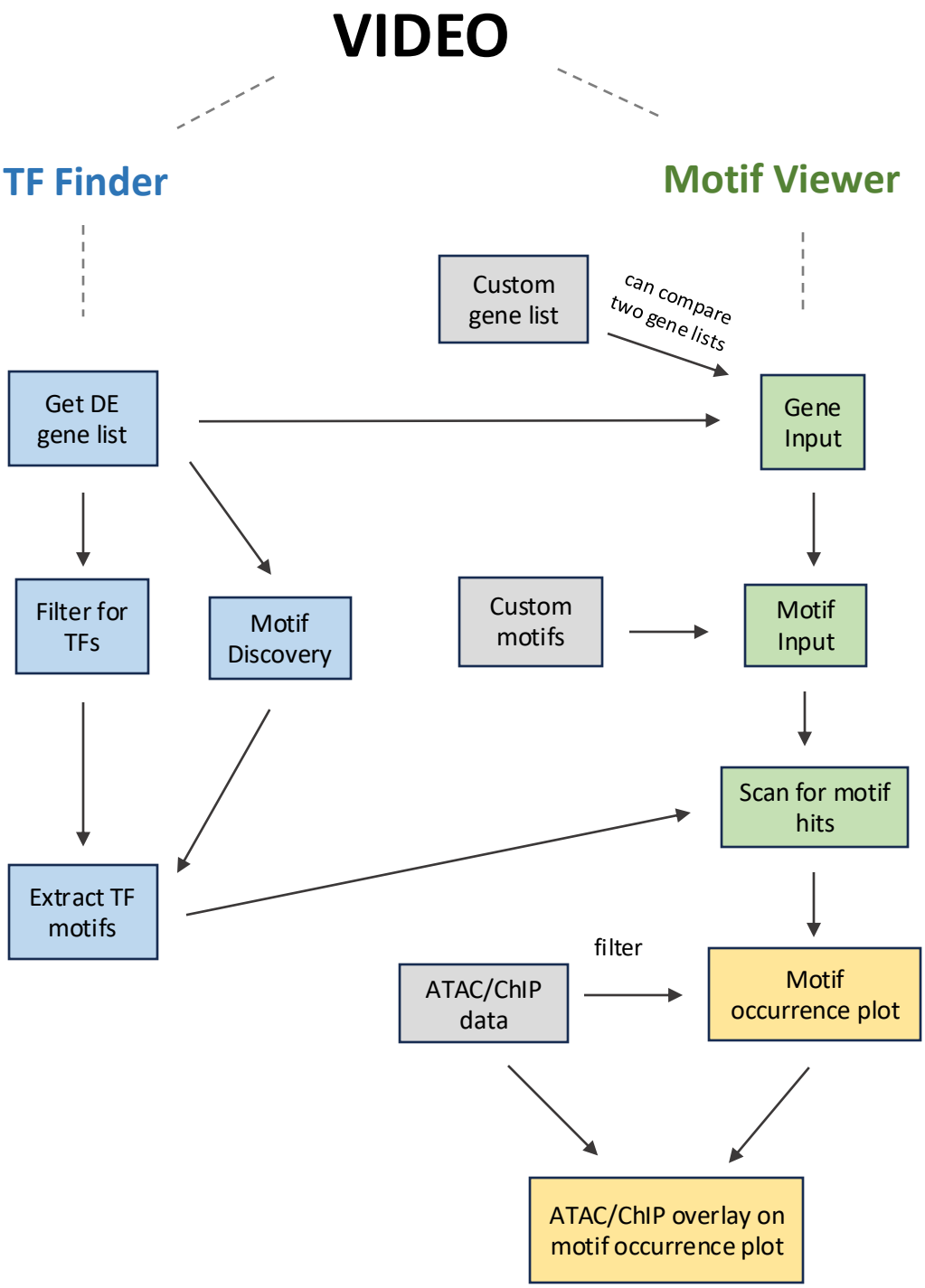
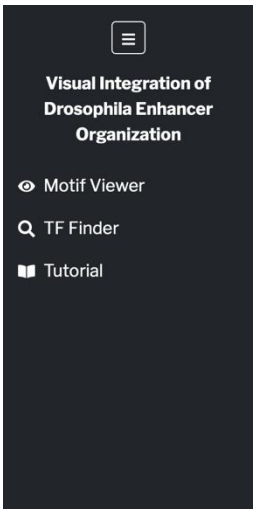
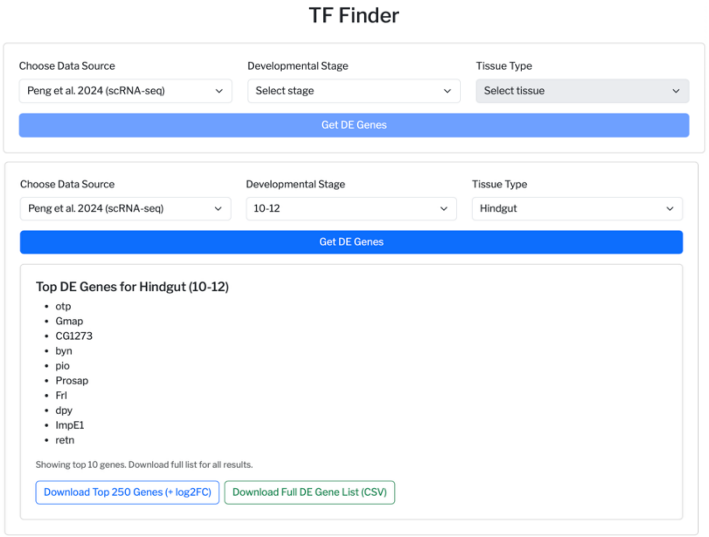

VIDEO
TF Finder
Motif Viewer
Custom gene list
can compare two gene lists
Get DE gene list
Gene Input
Filter for TFs
Motif Discovery
Custom motifs
Motif Input
Scan for motif hits
Extract TF motifs
filter
ATAC/ChIP data
Motif occurrence plot
ATAC/ChIP overlay on motif occurrence plot

Figure 2

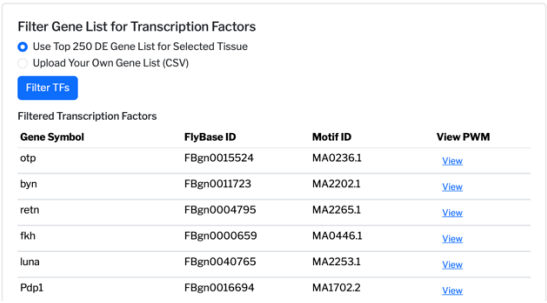
A

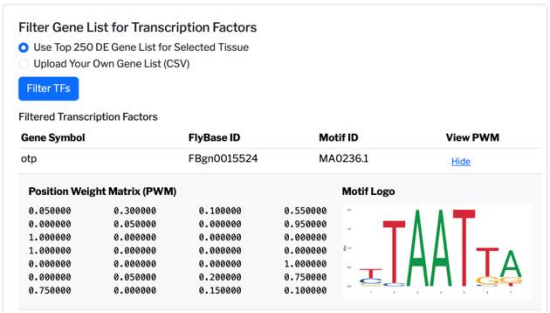


B

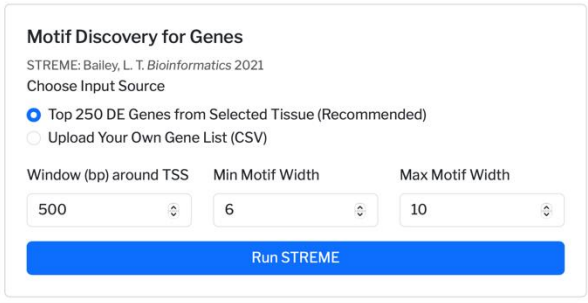


C





D



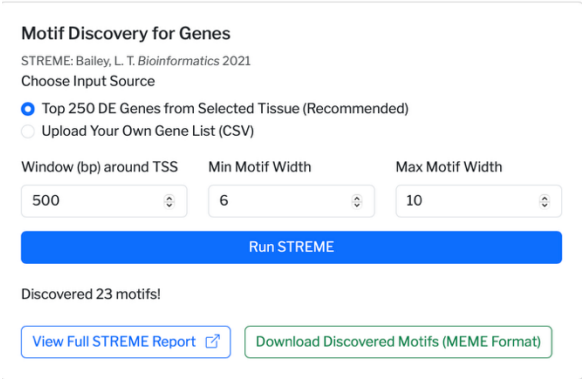
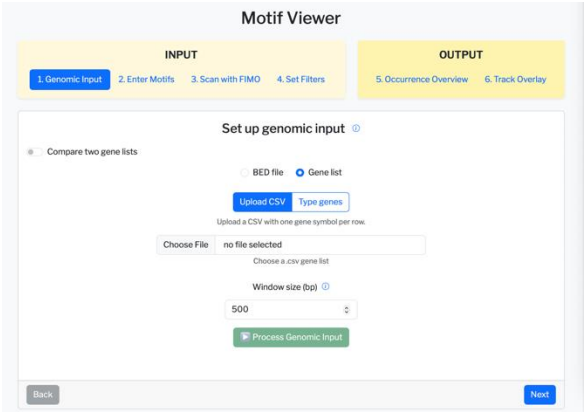
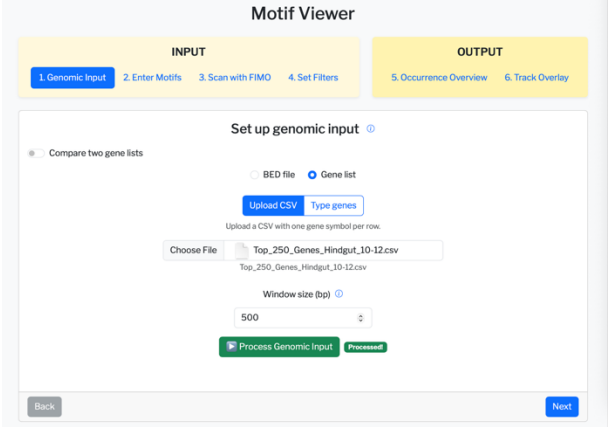


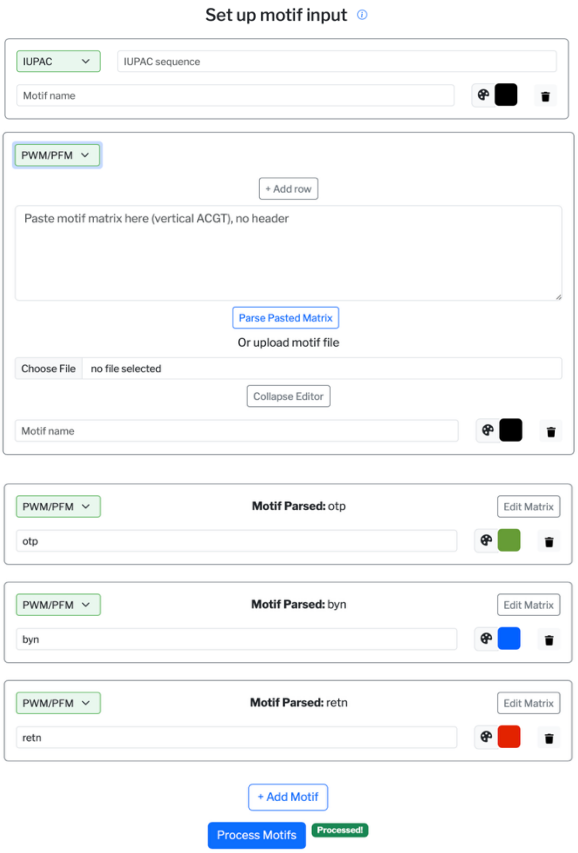
Figure 3

A

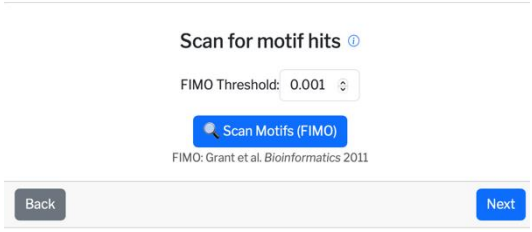




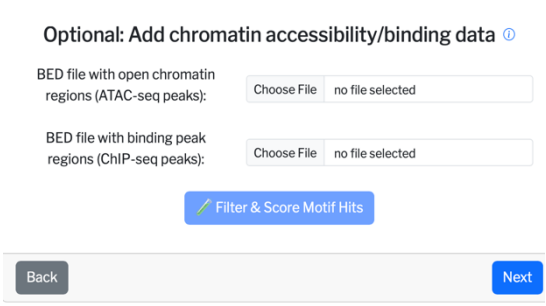
B



C



D



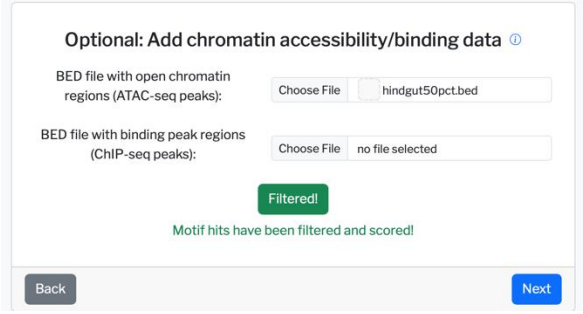
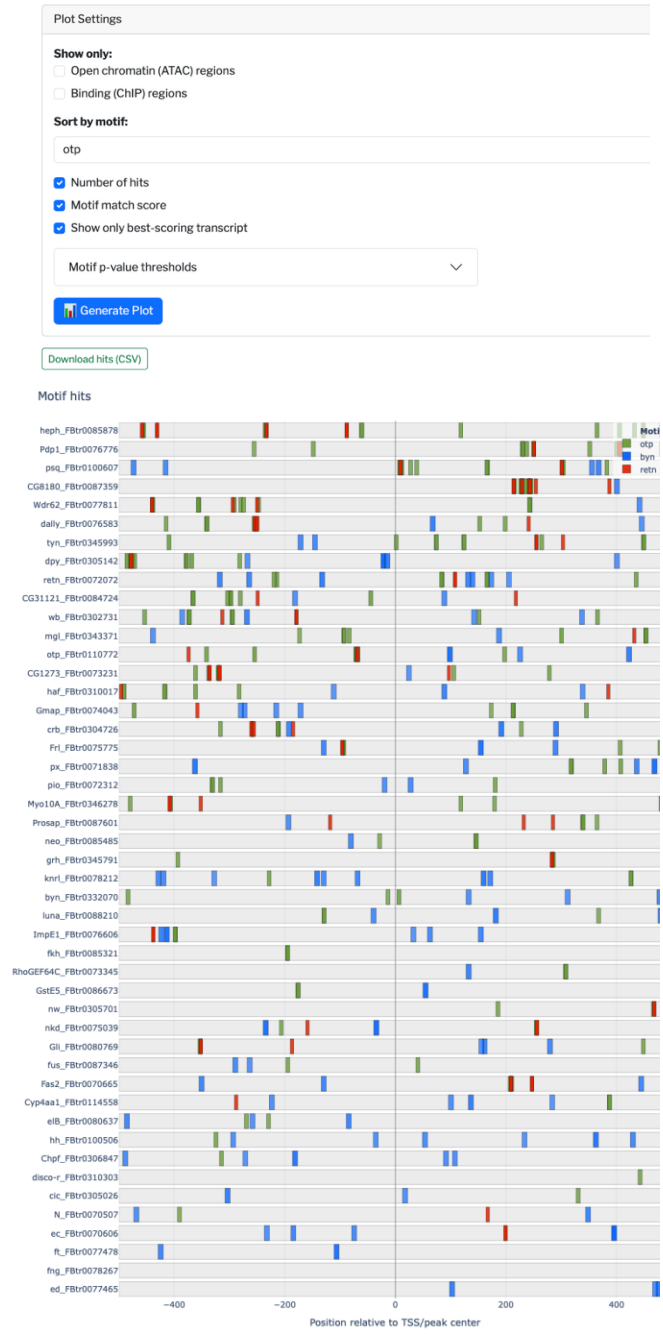


Figure 4

A


Plot Settings
Show only:
Open chromatin (ATAC) regions
Binding (ChIP) regions
Sort by motif:
otp
Number of hits
Motif match score
Show only best-scoring transcript
Motif p-value thresholds
Generate Plot
Download hits (CSV)
Motif hits
otp
byn
retn
−400
−200
0
200
400
Position relative to TSS/peak center

B

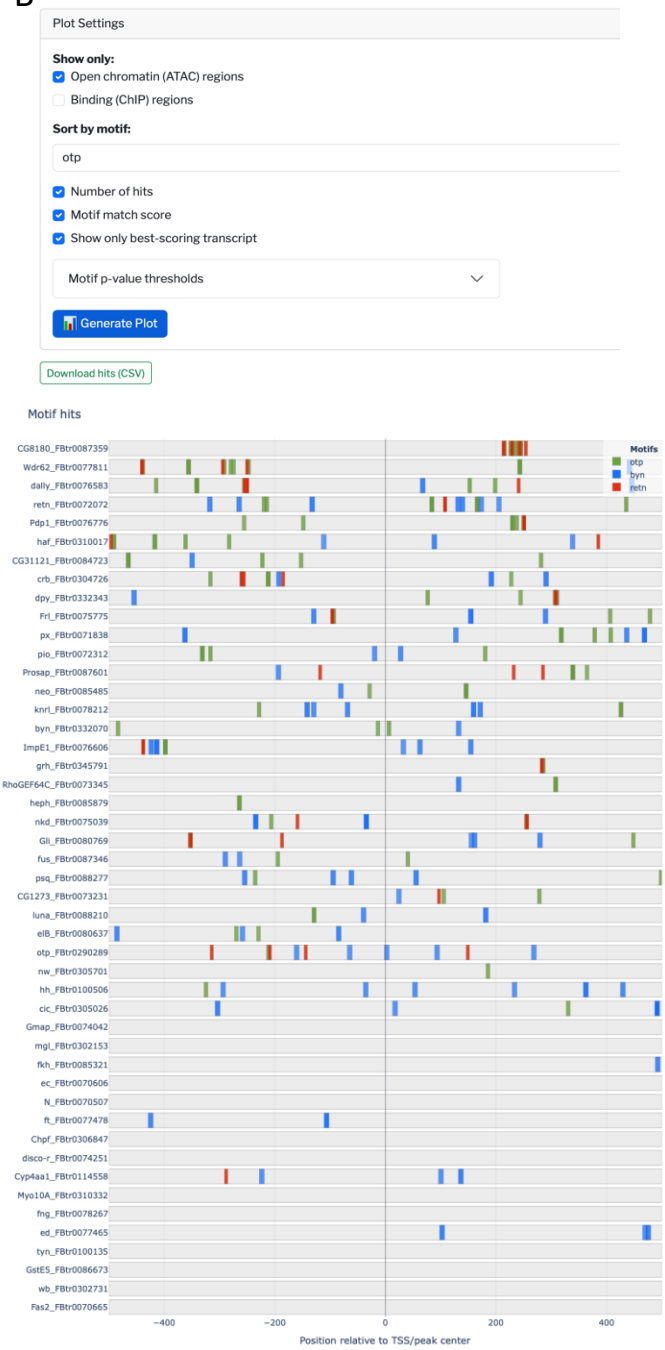
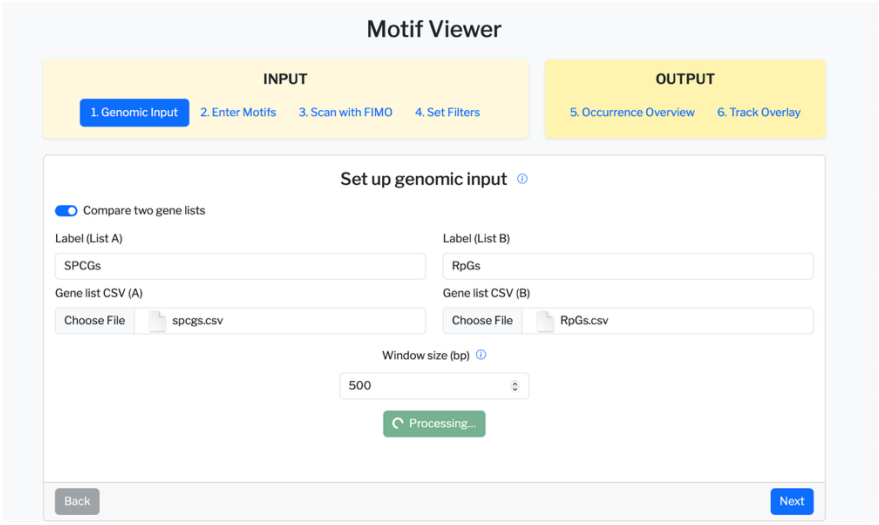
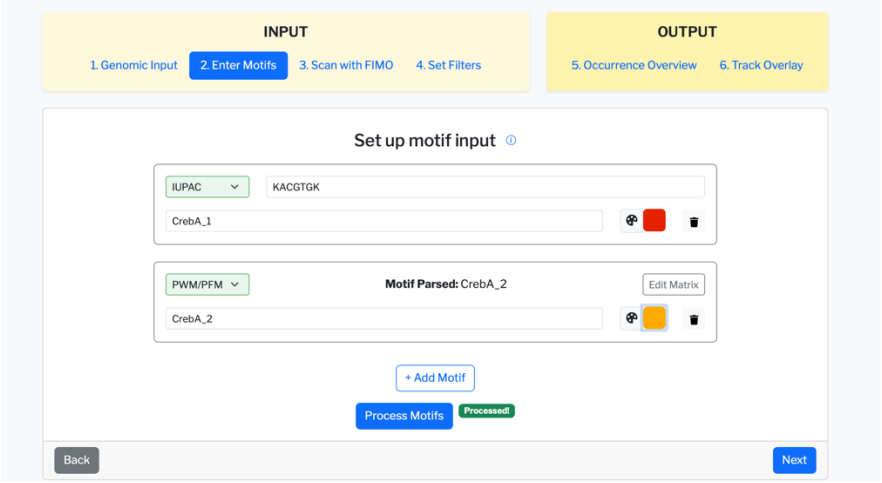

Plot Settings
Show only:
Open chromatin (ATAC) regions
Binding (ChIP) regions
Sort by motif:
otp
Number of hits
Motif match score
Show only best-scoring transcript
Motif p-value thresholds
Generate Plot
Download hits (CSV)
Motif hits
Motifs
otp
byn
retn
−400
−200
0
200
400
Position relative to TSS/peak center

Figure 5

A


Motif Viewer
INPUT
1. Genomic Input
2. Enter Motifs
3. Scan with FIMO
4. Set Filters
OUTPUT
5. Occurrence Overview
6. Track Overlay
Set up genomic input
Compare two gene lists
Label (List A)
SPCGs
Label (List B)
RpGs
Gene list CSV (A)
Choose File
spcgs.csv
Gene list CSV (B)
Choose File
RpGs.csv
Window size (bp)
500
Processing...
Back
Next

B


INPUT
1. Genomic Input
2. Enter Motifs
3. Scan with FIMO
4. Set Filters
OUTPUT
5. Occurrence Overview
6. Track Overlay
Set up motif input
IUPAC
KACGTGK
CrebA_1
PWM/PFM
Motif Parsed: CrebA_2
Edit Matrix
CrebA_2
+ Add Motif
Process Motifs
Processed!
Back
Next

C

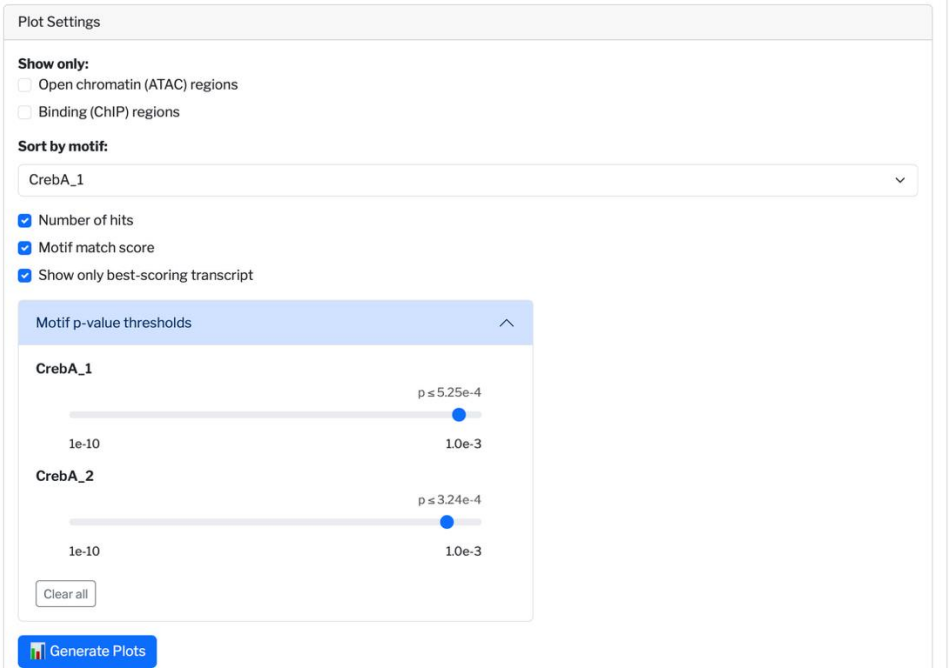

Plot Settings
Show only:
Open chromatin (ATAC) regions
Binding (ChIP) regions
Sort by motif:
CrebA_1
Number of hits
Motif match score
Show only best-scoring transcript
Motif p-value thresholds
CrebA_1
p ≤ 5.25e-4
1e-10
1.0e-3
CrebA_2
p ≤ 3.24e-4
1e-10
1.0e-3
Clear all
Generate Plots

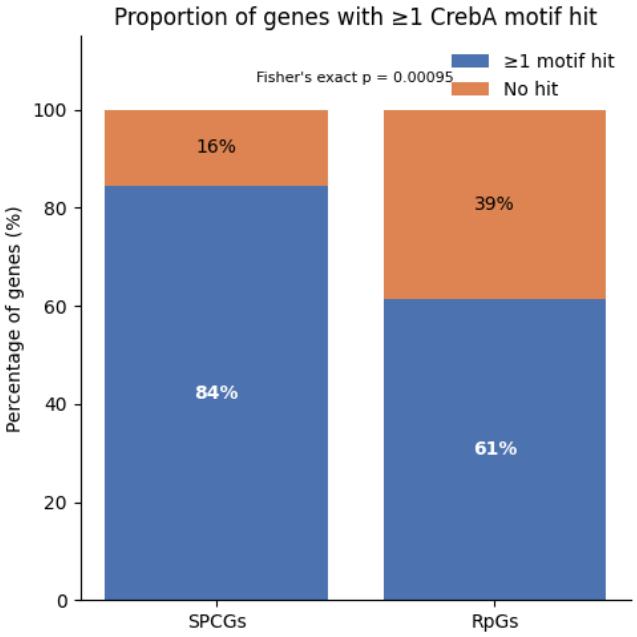
Figure 6 (A)

Motif hits - SPCGs
Motifs
CrebA_1
CrebA_2
Hsc70-3_FBtr0073611
p24-1_FBtr0073604
beta'COP_FBtr0080469
Arf1_FBtr0078575
sel_FBtr0088433
Golgin84_FBtr0084629
Sec61beta_FBtr0087498
Trp1_FBtr0079932
deltaCOP_FBtr0070371
ArfGAP3_FBtr0078539
Sec61gamma_FBtr0074745
Sec23_FBtr0309091
KdelR_FBtr0080021
Srp54k_FBtr0077169
ERp60_FBtr0088058
CHOp24_FBtr0070675
Sec24AB_FBtr0088451
Pdi_FBtr0075648
Sar1_FBtr0084206
Xbp1_FBtr0310044
Sec63_FBtr0076847
Grp170_FBtr0070411
CaBP1_FBtr0080837
alphaCOP_FBtr0072811
CG5021_FBtr0076513
Sil1_FBtr0084870
CG9175_FBtr0079211
zetaCOP_FBtr0075352
ReepB_FBtr0302451
Srp68_FBtr0076517
CG33774_FBtr0091775
p24-2_FBtr0082093
gammaCOP_FBtr0085793
sage_FBtr0081982
Spase25_FBtr0073506
Ostgamma_FBtr0079613
prtp_FBtr0073551
PH4alphaSG1_FBtr0085686
Ufc1_FBtr0087264
Sec13_FBtr0084409
Membrin_FBtr0077145
GCC88_FBtr0082366
Rer1_FBtr0084814
Cog6_FBtr0088595
Sec61alpha_FBtr0079266
TRAM_FBtr0070169
Spase22-23_FBtr0084564
Calr_FBtr0082103
wbl_FBtr0086510
Osbp_FBtr0084801
Smu1_FBtr0304689
Sec16_FBtr0113445
Slh_FBtr0077782
Grasp65_FBtr0074895
tw_FBtr0335403
betaCOP_FBtr0074576
Cog4_FBtr0080038
Cog7_FBtr0085509
Rtnl1_FBtr0079041
twr_FBtr0081687
Der-1_FBtr0077813
SrpRbeta_FBtr0076598
Spase12_FBtr0085435
CG7872_FBtr0074046
CG8230_FBtr0088674
Srp14_FBtr0273271
Tapdelta_FBtr0088153
Sec31_FBtr0088642
rt_FBtr0076146
Cog1_FBtr0082624
Cog2_FBtr0083114
Cog3_FBtr0077573
Srp9_FBtr0076750
Srp72_FBtr0083911
Srp19_FBtr0076879
Sec71_FBtr0080537
CG9356_FBtr0082047
spas_FBtr0084533
atl_FBtr0084656
PH4alphaSG2_FBtr0085653
Uggt_FBtr0075064
lqfR_FBtr0299515
epsilonCOP_FBtr0079261
ArfGAP1_FBtr0075981
fws_FBtr0080999
Cog8_FBtr0080248
TBC1D23_FBtr0077956
Tudor-SN_FBtr0072519
−400
−200
0
200
400
Position relative to TSS/peak center

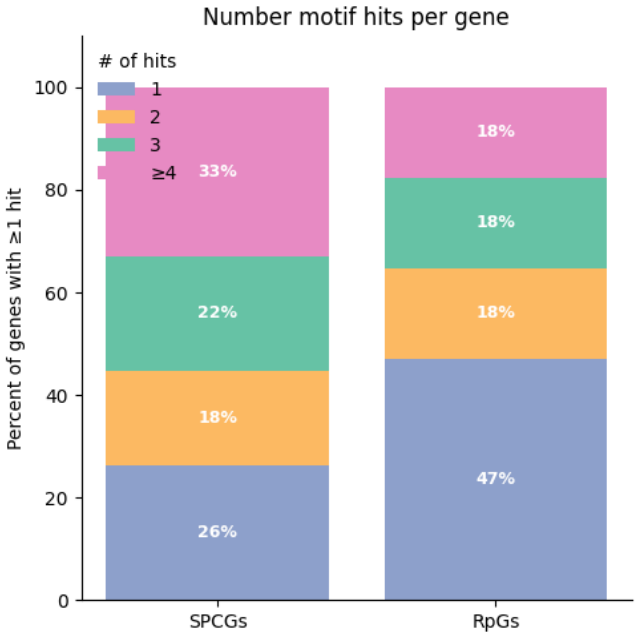
Motif hits - RpGs
Motifs
CrebA_1
CrebA_2
RpL23_FBtr0071897
RpL17_FBtr0345292
RpS20_FBtr0083964
RpL18_FBtr0076892
RpS2_FBtr0079905
RpL3_FBtr0082347
RpL24_FBtr0080524
RpL24-like_FBtr0082403
RpS24_FBtr0071935
RpS28b_FBtr0071360
RpS5a_FBtr0074406
RpS15Aa_FBtr0073792
RpS19a_FBtr0340149
RpS29_FBtr0082136
RpL13_FBtr0079888
RpL28_FBtr0073095
RpL30_FBtr0302491
RpL35_FBtr0070800
RpS17_FBtr0076479
RpL5_FBtr0111129
RpL8_FBtr0072924
RpL10Ab_FBtr0076032
RpL11_FBtr0086533
RpL14_FBtr0076633
RpL18A_FBtr0086947
RpL22-like_FBtr0072049
RpL23A_FBtr0072805
RpL27A_FBtr0077452
RpL29_FBtr0071592
RpL32_FBtr0085592
RpL35A_FBtr0078769
RpL37A_FBtr0302570
RpL40_FBtr0077470
RpS3_FBtr0084410
RpS4_FBtr0075884
RpS5b_FBtr0083006
RpS7_FBtr0339183
RpS9_FBtr0076425
RpS10a_FBtr0085194
RpS10b_FBtr0344339
RpS14a_FBtr0071094
RpS18_FBtr0086273
RpS21_FBtr0335131
RpL6_FBtr0085804
RpL7_FBtr0079946
RpL9_FBtr0080163
RpL10_FBtr0299869
RpL12_FBtr0072175
RpL13A_FBtr0078705
RpL15_FBtr0113742
RpL19_FBtr0072405
RpL21_FBtr0085961
RpL22_FBtr0070148
RpL26_FBtr0075066
RpL27_FBtr0084932
RpL31_FBtr0088525
RpL36_FBtr0070155
RpL36A_FBtr0079546
RpL38_FBtr0334985
RpL39_FBtr0072185
RpL41_FBtr0100231
RpS3A_FBtr0089175
RpS6_FBtr0071135
RpS8_FBtr0085489
RpS11_FBtr0088006
RpS12_FBtr0075878
RpS13_FBtr0079724
RpS14b_FBtr0071096
RpS15Ab_FBtr0088226
RpS15_FBtr0087123
RpS16_FBtr0071855
RpS19b_FBtr0084519
RpS23_FBtr0339970
RpS25_FBtr0082369
RpS26_FBtr0081090
RpS27_FBtr0084868
RpS27A_FBtr0080016
RpS28a_FBtr0085601
RpS30_FBtr0083969
−400
−200
0
200
400
Position relative to TSS/peak center

Figure 6 (continued)

B



C



D

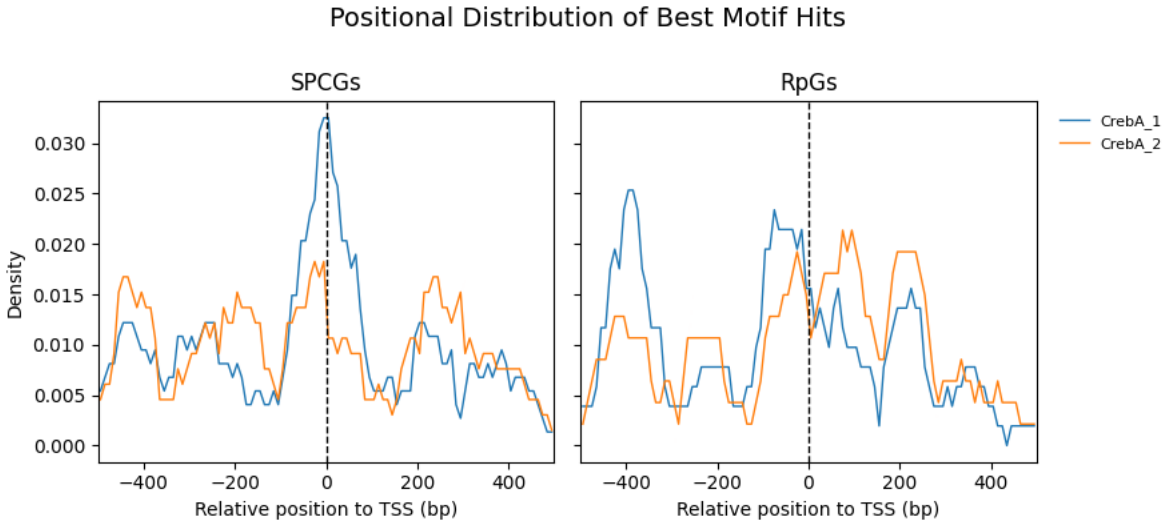
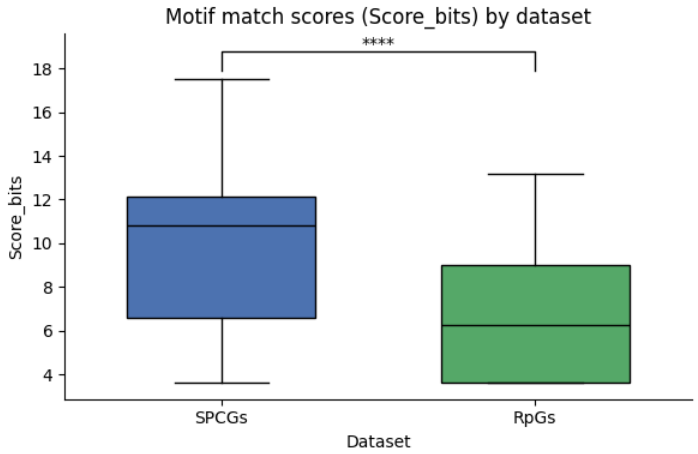
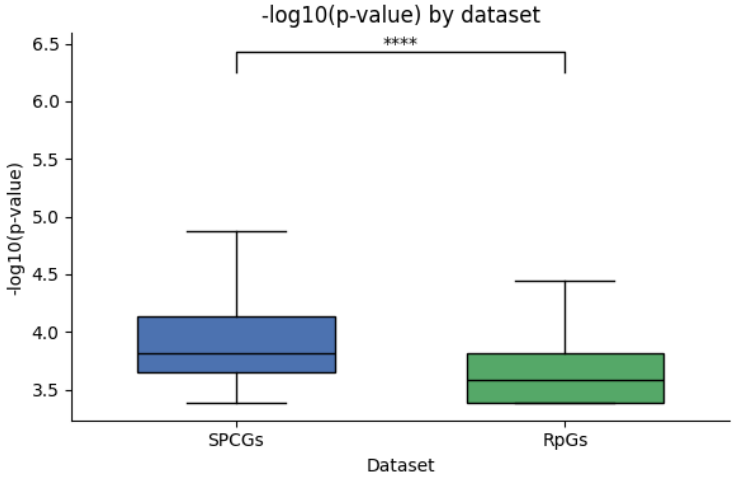


Figure 7

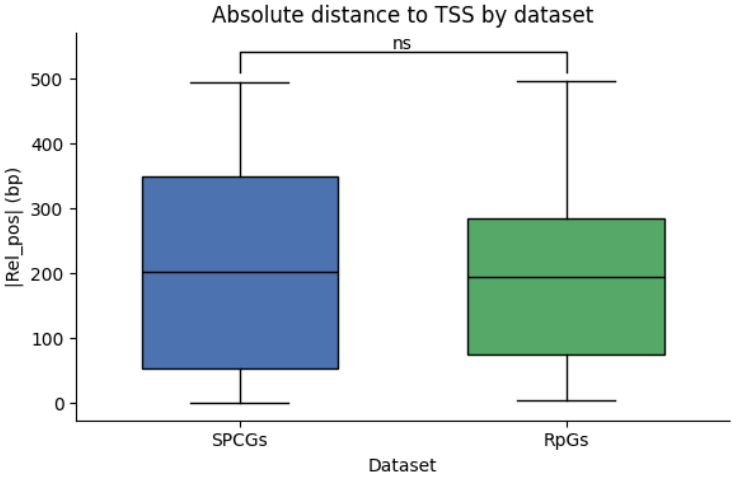
A


Motif match scores (Score_bits) by dataset
****
Score_bits
18
16
14
12
10
8
6
4
SPCGs
RpGs
Dataset

B


-log10(p-value) by dataset
****
-log10(p-value)
6.5
6.0
5.5
5.0
4.5
4.0
3.5
SPCGs
RpGs
Dataset

C


Absolute distance to TSS by dataset
ns
|Rel_pos| (bp)
500
400
300
200
100
0
SPCGs
RpGs
Dataset