

Identification of putative enhancer-like elements predicts regulatory networks active in planarian adult stem cells

Jakke Neuro^{1*}, Divya Sridhar¹, Anish Dattani², Aziz Aboobaker^{1*}

¹Department of Zoology, University of Oxford, Oxford, United Kingdom; ²Living Systems Institute, University of Exeter, Exeter, United Kingdom

Abstract Planarians have become an established model system to study regeneration and stem cells, but the regulatory elements in the genome remain almost entirely undescribed. Here, by integrating epigenetic and expression data we use multiple sources of evidence to predict enhancer elements active in the adult stem cell populations that drive regeneration. We have used ChIP-seq data to identify genomic regions with histone modifications consistent with enhancer activity, and ATAC-seq data to identify accessible chromatin. Overlapping these signals allowed for the identification of a set of high-confidence candidate enhancers predicted to be active in planarian adult stem cells. These enhancers are enriched for predicted transcription factor (TF) binding sites for TFs and TF families expressed in planarian adult stem cells. Footprinting analyses provided further evidence that these potential TF binding sites are likely to be occupied in adult stem cells. We integrated these analyses to build testable hypotheses for the regulatory function of TFs in stem cells, both with respect to how pluripotency might be regulated, and to how lineage differentiation programs are controlled. We found that our predicted GRNs were independently supported by existing TF RNAi/RNA-seq datasets, providing further evidence that our work predicts active enhancers that regulate adult stem cells and regenerative mechanisms.

***For correspondence:**

jakke.neiro@queens.ox.ac.uk
(JN);

aziz.aboobaker@zoo.ox.ac.uk
(AA)

Competing interest: The authors declare that no competing interests exist.

Funding: See page 24

Preprinted: 04 February 2022

Received: 05 May 2022

Accepted: 23 August 2022

Published: 23 August 2022

Reviewing Editor: Marianne E Bronner, California Institute of Technology, United States

© Copyright Neuro *et al.* This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

Editor's evaluation

The authors have generated a comprehensive list of transcription factors in the planarian *Schmidtea mediterranea*, providing insight into which factors are highly expressed in the stem cell compartment. The computational identification of transcription factors and putative enhancers will be helpful to researchers studying stem cell and regenerative biology using planarians and provides a large dataset that informs upon the evolution of regulatory sequences and transcription factor function.

Introduction

The molecular and evolutionary mechanisms of regeneration remain underexplained compared to animal development. This can be attributed to the greater technical difficulty historically associated with studying the molecular mechanisms of adult biology compared with development. However, an ever-broadening repertoire of model organisms for regeneration, progress in understanding the variety of cellular and molecular mechanisms used across taxa, and advances in experimental tools are serving to close this gap. One area of slower progress has been our knowledge of the regulatory logic of regenerative mechanisms, with relatively few studies in highly regenerative models providing precise insight into the epigenetic regulation of regeneration. While transcription factors (TFs) have been assigned functions, and in some cases a list of likely targets through transcriptome analysis, regulatory logic has been studied in only a few cases (*Pascual-Carreras et al., 2020; Wang et al.,*

2020). This is particularly true in the highly regenerative planarian model system, where only a few studies credibly address epigenetic and regulatory mechanisms (Duncan et al., 2015; Dattani et al., 2018; Mihaylova et al., 2018; Dattani et al., 2019; Pascual-Carreras et al., 2020).

Enhancers are distal-acting elements that regulate transcription initiation when TFs attach to TF binding sites, containing conserved sequences. In a regenerative context, active enhancers have been identified at the tissue level in zebrafish, the African killifish *Nothobranchius furzeri*, and the fruit fly *Drosophila melanogaster*, and at the whole-body level in the acoele worm *Hofstenia miamia* and the cnidarian *Hydra vulgaris* (Harris et al., 2016; Kang et al., 2016; Goldman et al., 2017; Gehrke et al., 2019; Murad et al., 2019, Yang and Kang, 2019; Wang et al., 2020). Despite their importance as a model for understanding regeneration, there are only a few putative enhancers identified or implicated in planarians (Pascual-Carreras et al., 2020), partly due to the lack of transgenic approaches allowing for direct functional testing of potential regulatory elements. Hence, other analytical methods must be used to discover enhancer regions de novo across the genome. A set of putative planarian TFs have been identified based on available transcriptomic data, but genome-wide TF interaction with potential enhancers has not been elucidated (Swapna et al., 2018; Pascual-Carreras et al., 2020). Ultimately, TF interactions with enhancers can be used to construct gene regulatory networks (GRNs) (Duren et al., 2017; Lowe et al., 2019; Miraldi et al., 2019, Duren et al., 2020; Janssens et al., 2022).

In planarians, regeneration is based on a population of somatic pluripotent stem cells called neoblasts that can differentiate into all cell types of the adult body plan and are the only cycling cells in the adult worm (Aboobaker, 2011; Rink, 2013, Zhu and Pearson, 2016). Nonetheless, how neoblasts dynamically retain pluripotency, specify fate, and differentiate remains unknown from a regulatory network perspective, while a number of different models at the cell biological level have been hypothesized (Adler and Sánchez Alvarado, 2015, Swapna et al., 2018; Raz et al., 2021). Fluorescence-activated cell sorting (FACS) separates cells into three compartments: S/G2/M cell-cycle stage neoblasts (known as X1), G1 stage neoblasts and postmitotic progeny (known as X2), and differentiated cells (known as Xins, as the insensitive to radiation treatment) (Hayashi et al., 2006; Labbé et al., 2012; Onal et al., 2012; Romero et al., 2012; Solana et al., 2012). Analysis of the X1 cells has been used to catalog genes expressed in neoblasts, and these studies have revealed that neoblasts are heterogeneous and can be subdivided into multiple classes based on their gene expression profiles (van Wolfswinkel et al., 2014; Fincher et al., 2018; Plass et al., 2018). Some classes of fate-specified neoblasts are termed specialized neoblasts, which include precursors to eyes, protonephridia, epidermis, intestine, pharynx, neurons, and muscle (van Wolfswinkel et al., 2014; Raz et al., 2021). Fate-specific transcription factors (FSTFs) are expressed in S/G2/M phase and are thought to direct fate specification into different cell types (Raz et al., 2021). Evidence from current experimental data supports a model where specialized neoblasts can divide asymmetrically giving rise to one fate-specified, postmitotic daughter and a proliferative neoblast that may still specify different fates (Raz et al., 2021). Genome-wide identification of active enhancers in the X1 compartment would shed light on the GRNs regulating the dynamic behavior of neoblasts in planarian regeneration (Labbé et al., 2012; Onal et al., 2012; Solana et al., 2012). Currently, no such predictions of enhancers or GRNs of TFs exist.

Identifying enhancers is challenging and especially so in non-model organisms, but various genome-wide high-throughput sequencing techniques have revealed signatures indicative of enhancers (Shlyueva et al., 2014, Tomoyasu and Halfon, 2020). Chromatin accessibility has proven to be a universal attribute of active enhancers and other regulatory regions in eukaryotes (Thomas et al., 2011; West et al., 2014; Zhu et al., 2015; Daugherty et al., 2017; Klemm et al., 2019). Assay for transposase-accessible chromatin using sequencing (ATAC-seq) has become the standard method of mapping open chromatin regions in various taxa (Buenrostro et al., 2013; Li et al., 2019; Yan et al., 2020). However, chromatin accessibility is not unique and specific to enhancers, and thus other complementary methods are used to discriminate enhancers from other open chromatin regions, particularly at the point of initial identification.

Enhancers are known to be flanked by nucleosomes with specific histone modifications (Shlyueva et al., 2014). In both mammals and *D. melanogaster*, acetylated lysine 27 of histone H3 (H3K27ac) is known to mark active enhancers together with mono-methylation of lysine 4 on histone H3 (H3K4me1), whereas H3K4me1 alone marks predetermined or poised enhancers (Heintzman et al.,

2007; Creyghton et al., 2010; Ernst et al., 2011; Rada-Iglesias et al., 2011; Bonn et al., 2012; Arnold et al., 2013; Calo and Wysocka, 2013). However, poised enhancers are not always activated, most enhancers are activated without a prior poised state, and some poised enhancers are later actively repressed, indicating that the poised enhancer state is not necessarily indicative of a pre-activation state (Bonn et al., 2012; Rada-Iglesias et al., 2011; Koenecke et al., 2017). Nonetheless, the bivalent epigenetic signature of H3K27ac and H3K4me1 seems to be a conserved indicator of active enhancers in metazoans and has been successfully used for enhancer detection in non-model organisms (Gaiti et al., 2017a, Jänes et al., 2018). In planarian neoblasts, the same histone modifications mark the active, suppressed, and bivalent state of promoters as in vertebrates (Dattani et al., 2018). Taken together, these data suggest that epigenetic marks may have a conserved association with regulatory elements across bilaterians (Schwaiger et al., 2014; Sebé-Pedrós et al., 2016; Gaiti et al., 2017b, Dattani et al., 2018).

Here, we significantly improve upon the annotation of the planarian genome (Grohme et al., 2018), take a computational approach to identify all TFs in this annotation, identify putative enhancers in the planarian genome supported by multiple lines of evidence, and then construct hypothetical GRNs active in neoblasts. We find that multiple enhancers have evidence of FSTF-mediated regulation, supporting the view that fate specification occurs in the S/G2/M phases. The FSTFs of different cell types seem to cross-regulate each other, revealing the potential for a dynamic GRN governing neoblast fate specification. We identified enhancers linked to several unstudied TFs implicating them in potentially central roles in neoblast TF GRNs. Enhancers linked to well-known planarian positional genes suggest regulatory mechanisms for some of the known links between these genes implicated by phenotypic studies. Finally, we show that our GRNs predict TF regulatory interactions implicated by available data measuring gene expression changes after TF RNAi. Overall, this work provides a foundation for future work on the regulatory logic of planarian stem cell biology and identifies many candidate TFs with predicted roles in regulating adult stem cells.

Results

Annotation with a full range of transcriptome samples identifies more than 3000 new protein coding genes in the *Schmidtea mediterranea* genome

We refined and extended the current annotation of the *S. mediterranea* genome (SMESG.1 genome and SMESG high-confidence annotation at Planmine; Brandl et al., 2016). We performed a genome-directed annotation based on the genome sequence and 183 independent RNA-seq datasets, including data both from whole worms and cell compartments (Figure 1A and B; this new annotation has also been used in García-Castro et al., 2021, but is described in more detail here). By including a diverse and large set of RNA-seq data, we sought to characterize transcripts undetected in individual studies and annotation attempts (Hoff and Stanke, 2013; Hoff and Stanke, 2019). Furthermore, we calculated proportional expression values for each cell population defined by FACS using approaches established previously (Figure 1A; Dattani et al., 2018).

In total, our expression-driven annotation process identified 91,068 transcripts at 28,097 genomic loci (Figure 1B). The annotation process validated all gene models in the SMESG high-confidence annotation, as all loci, transcripts, and exons were also found in our new annotation. In total, 50,213 transcripts were identified as putative new isoforms of previously identified loci (Figure 1C). Furthermore, 7412 new loci with 10,636 transcripts were found (Figure 1B, C). The protein coding potential of these new transcripts was assessed by defining putative open-reading frames (ORFs) and scanning for protein structures (Figure 1B). In total, 3121 new loci with 4752 transcripts were predicted to be coding, while 4291 loci with 5884 transcripts were predicted to be non-coding (Figure 1B, C).

The newly described isoforms across the genome were slightly shorter than known transcripts (median length 1618 bp vs. 1656 bp), while the new coding and non-coding transcripts were much shorter (median length 583 and 388 bp, respectively) (Figure 1D). However, the mean transcripts per kilobase million (TPM) value measured across the RNA-seq samples for new isoforms did not differ much from the levels of previously known transcripts, while the mean TPM values for new coding and non-coding transcripts were slightly higher (Figure 1—figure supplement 1A). This suggests that the main advantage of our approach was to discover shorter transcripts (and encoded proteins) not found

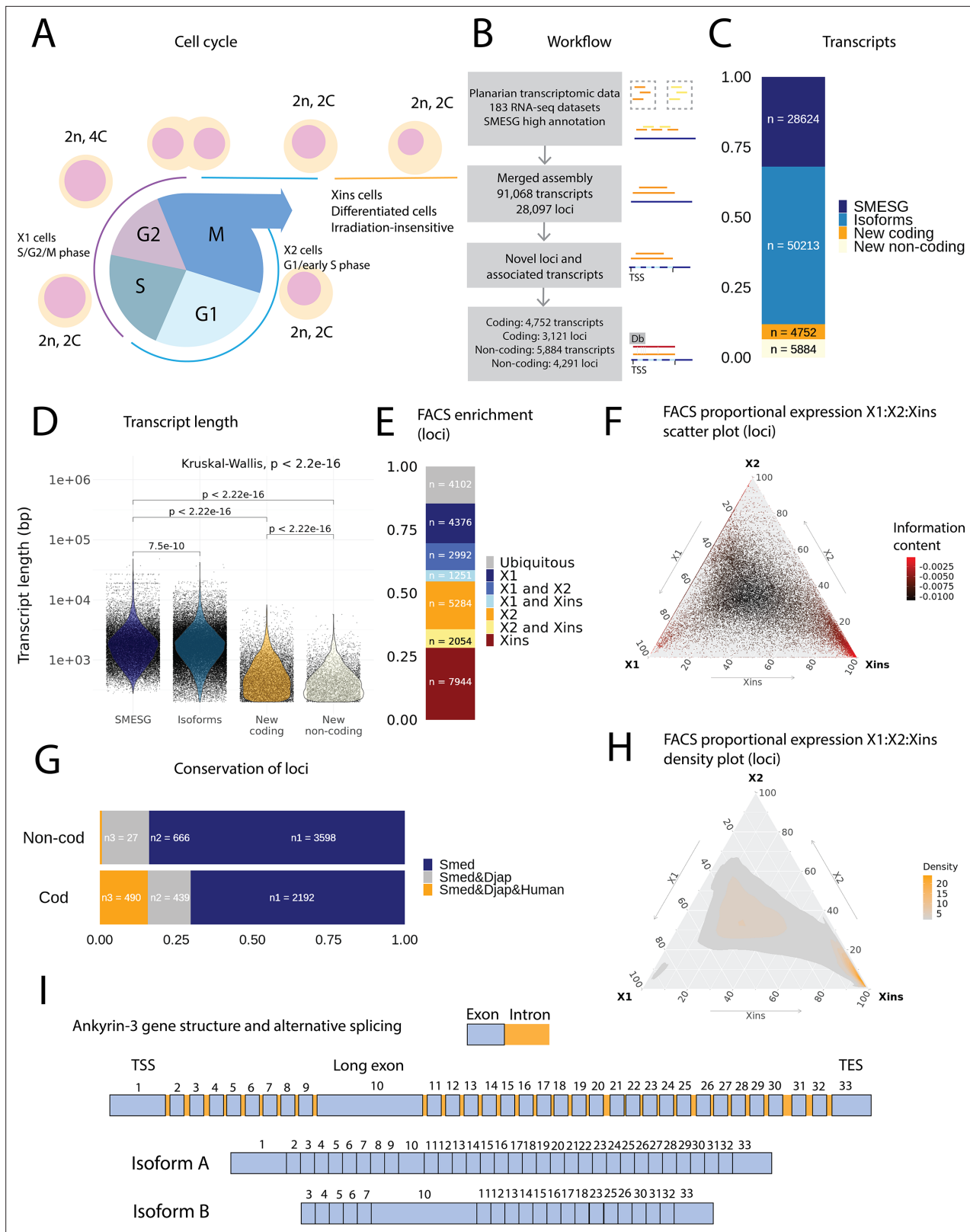


Figure 1. An expression-driven genome annotation of *Schmidtea mediterranea*. (A) Diagram of the planarian neoblast cell cycle illustrating how the fluorescence-activated cell sorting (FACS) cell compartments X1, X2, and Xins relate to different cell-cycle phases. (B) Overview of methodology for annotating the *S. mediterranea* genome and identifying novel coding and non-coding transcripts. In total, 183 RNA-seq datasets were aligned to the SMESG.1 genome (Planmine). Novel transcripts were assembled by using the SMESG-high-confidence annotation as a reference (Planmine).

Figure 1 continued on next page

Figure 1 continued

(C) The proportion of transcript identities in the new annotation. 'SMESG' are transcripts present in the known SMESG high-confidence annotation, 'Isoforms' are transcripts that are isoforms or splicing variants of known transcripts in the SMESG high-confidence annotation, and 'New coding' and 'New non-coding' are previously unknown transcripts deemed to have coding and non-coding potential, respectively. (D) The transcript lengths of known transcripts (SMESG, $n = 28624$), isoforms ($n = 50213$), new coding transcripts ($n = 4752$), and new non-coding transcripts ($n = 5884$). Kruskal-Wallis test is used to compare all groups and Wilcoxon rank-sum test is used for pairwise comparisons. (E) FACS categorization of 28,003 total loci into enrichment groups (Dattani et al., 2018). The enrichment groups are X1 (X1 proportional expression $\geq 50\%$), X2 (X2 proportional expression $\geq 50\%$), Xins (Xins proportional expression $\geq 50\%$), X1 and X2 (X1 + X2 proportional expression $\geq 75\%$, neither enriched in X1 nor X2), X1 and Xins (X1 + Xins proportional expression $\geq 75\%$, neither enriched in X1 nor Xins), X2 and Xins (X2 + Xins proportional expression $\geq 75\%$, neither enriched in X2 nor Xins), and ubiquitous (loci not categorized into the enrichment groups above and with roughly equal proportion in all three groups). (F) Ternary plot of proportional expression values in the X1, X2, and Xins cell compartments. Each dot represents one single locus. The information content metric represents the enrichment of a locus to one of the three compartments, meaning that gray loci are unspecific while red loci are specific to a single compartment (see 'Materials and methods' for details). (G) Conservation of new coding and non-coding loci to loci in *Dugesia japonica* and humans. 'Smed' (n1) are loci exclusive to *S. mediterranea*, 'Smed&Djap' (n2) are loci with a homolog in *D. japonica* but not in humans, and 'Smed&Djap&Human' (n3) are loci with homologs both in *D. japonica* and humans. (H) Ternary plot of the density distribution of proportional expression values in the X1, X2, and Xins cell compartments. The plot gives a visual condensation of ternary plot F. The density distribution is estimated by two-dimensional kernel density estimation. (I) Schematic of the gene structure of ankyrin-3 and two example isoforms that are alternatively spliced (SMEST010564001.1.p1 and MSTRG.3975.46.p1). The detailed genomic track of all the isoforms is given in Figure 1—figure supplement 1F. TSS, transcriptional start site; TES, transcript end site.

The online version of this article includes the following figure supplement(s) for figure 1:

Figure supplement 1. Information on the genome annotation of *Schmidtea mediterranea*.

by previous annotation approaches (Figure 1D, Figure 1—figure supplement 1). We sorted all annotated genes to FACS enrichment groups (Figure 1H), using previously described methods (Dattani et al., 2018). The number of genes with enriched expression was highest in the Xins compartment and lowest in X1 cells (Figure 1E). As expected by overall lineage relationships, the shared enrichment between X1 and Xins is less common than between X1 and X2 and between X2 and Xins cells, congruent with the fact that X1 cells are all neoblasts, X2 is an amalgam of G1 neoblasts and differentiating postmitotic progeny, and Xins cells are their collective differentiation products (Figure 1E). By inspecting the distribution of proportional expression values, we also see that the distribution is shifted towards the Xins compartment (Figure 1F, H), thus overall expression of genes is enriched in differentiated cell types.

The majority of the new coding loci were unique to *S. mediterranea* (70%), but 30% had a homolog in the *Dugesia japonica* genome (E-value cutoff 10^{-10} , Figure 1G; An et al., 2018). Only 27 non-coding loci (0.6%) had a potential homolog in the human genome, but 666 loci (16%) had a homolog in the *D. japonica* genome (Figure 1G). The Gene Ontology (GO) analysis of new coding loci revealed that the new genes were enriched for a function in RNA biosynthesis, metabolic processes, dephosphorylation, deacetylation, transmembrane transport, and mitochondrial functions (Figure 1—figure supplement 1C–E). These data suggest that many of our newly annotated genes are conserved amongst the group of planarians used for regeneration and stem cell research.

Table 1. Genes with more than 100 transcript isoforms.

Number of isoforms	Gene	Description
571	TMEM25	Transmembrane protein 25
291	CLEC18B	C-type lectin domain family 18 member B
206	ANK2	Ankyrin 2
143	GABPB2	GA binding protein transcription factor subunit
138	PACRG	Parkin coregulated
112	ANK3	Ankyrin 3
106	DST	Dystonin
104	ANK3	Ankyrin 3

Some genes in the new annotation displayed high levels of alternative splicing, including homologs of lectin, ankyrins, and dystonin (**Figure 1I, Figure 1—figure supplement 1B, Table 1**). In mammals, ankyrin-3 (ankyrin-G) is a structural protein localized to the axon initial segment (AIS) and the nodes of Ranvier, and alternative splicing is known to underlie its functional diversity (**Figure 1I, Figure 1—figure supplement 1F; Hopitzan et al., 2005; Lopez et al., 2017; Nelson and Jenkins, 2017**). A planarian homolog of ankyrin-3 has 112 isoforms, including isoforms with one long exon, supporting the finding that giant ankyrin-based cytoskeleton of the AIS may have been present in the last common ancestor of bilaterians (**Figure 1I, Figure 1—figure supplement 1F; Jegla et al., 2016**).

Taken together, these initial example analyses of our new annotation, particularly the discovery of many hundreds of new loci and thousands of putative alternative isoforms, suggest that it will have an important utility for the research community studying all aspects of planarian biology.

Comprehensive annotation of planarian transcription factors highlights a diversity of unknown zinc fingers

We screened for TFs in the new annotation using the same approach as **Swapna et al., 2018 (Figure 2A)**. We validated the TF potential through a systematic protein domain annotation, assessed for homology to known TFs, and manually reviewed the list of TFs to assign them to planarian TFs present in the literature and databases (**Figure 2A**). Altogether, we found predicted 551 TFs in *S. mediterranea*, of which we found 248 to be described in the planarian literature as a named molecular sequence (i.e., the sequence was assigned a proper TF name; for details, see 'Materials and methods'). The naming of planarian TFs in the literature was mostly consistent, but some inconsistencies were found (**Supplementary file 1**). We classified the TFs into four structural categories: basic domains, zinc domains, helix-turn-helix domains, and other domains (**Stegmaier et al., 2004**). Most basic domains had been described by **Cowles et al., 2013**, but we still identified new homologs of *Atf*, *Batf*, *Creb*, *Htf*, *Mad*, *Matf*, *MyoD*, *Npas*, and *Pdp* family members, each with broad established roles in metazoan biology (**Figure 2B, Supplementary file 1**).

In contrast, we identified multiple uncharacterized zinc finger domain TFs (ZNFs), many of which have not received much consideration yet in planarian regeneration research. While some of these unstudied ZNFs could be assigned to well-known ZNF families such as GATA, KLF, EGR, and PRDM, many could not be assigned to well-described families. The nomenclature of ZNFs was based on the naming of human proteins to which they have the highest identity, and hence many appear in the ZNF and ZNP categories (**Figure 2C**). Interestingly, we also find ZNFs related to SCAN-domain containing zinc fingers (ZSCAN) and pogo transposable elements with KRAB domain (POGK) previously only described in vertebrates (**Emerson and Thomas, 2011; Gao and Qian, 2020a**). However, these planarian ZNFs do not contain the SCAN or KRAB domains, and the similarity arises from the DNA-binding domains. While members of the pogo transposable element superfamily are found throughout the metazoans, the KRAB subfamily is specific to vertebrates (**Gao et al., 2020b**). For this reason, we have provisionally named these ZNFs Zscan-like and POGK-like (**Figure 2C**).

For helix-turn-helix domains, new homologs were found for several described TF families, including *Nkx* and *Six*, and a few TFs belonging to families not previously described in planarians were also discovered, such as *Lmx* and *Shox* (**Figure 2D**). Some TFs were also newly annotated for other domain families (**Figure 2E**).

We proceeded to allocate proportional expression values to the TFs with respect to the X1, X2, and Xins cell compartments (**Dattani et al., 2018**). The distribution of the proportional expression values was even, although a slight skew towards the X1 compartment was evident (**Figure 2F, G**). Most TFs were assigned to the X1 and Xins compartments and the least to the X2 compartment (**Figure 2F, G**). In other words, few TFs were specific to the X2 compartment, while most TFs were specifically expressed in stem cells and their nascent progeny (X1 and X2) or in progeny and differentiated cells (X2 and Xins). This would be expected by overall lineage relationships. There was no compartment-specific enrichment for different TF domains (χ^2 test, $p=0.2$).

We then moved to assigning predicted target binding motifs to the annotated set of planarian TFs. Most studies in non-model organisms have tended to use motifs directly from only a single-model organism or solely rely on de novo motif enrichment without reference to the TFs actually present in the studied organism (**Gaiti et al., 2017a, Gehrke et al., 2019; Murad et al., 2019**). However, we used the same approach as **Siebert et al., 2019** and searched the JASPAR database for TFs with the

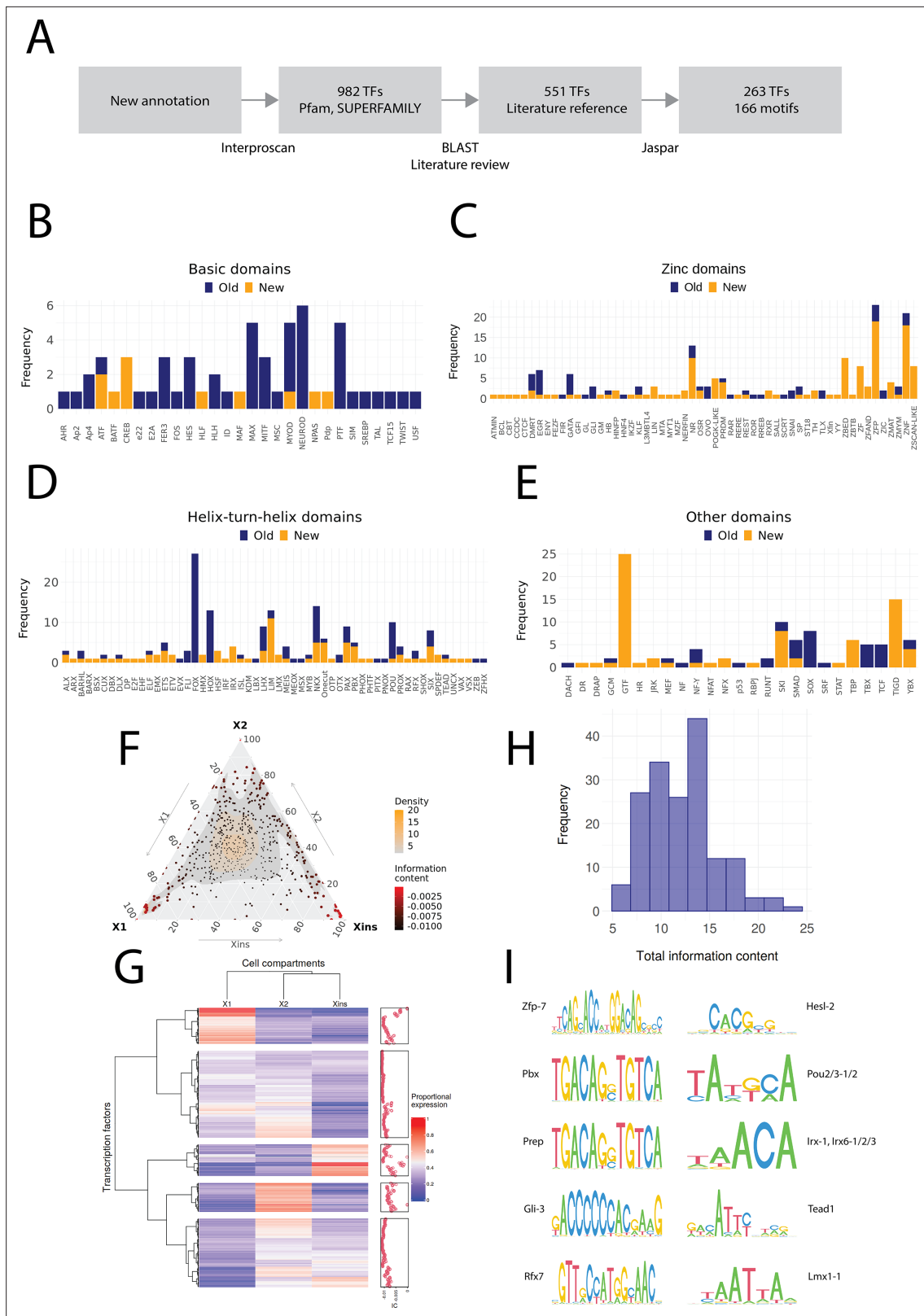


Figure 2. Annotation of planarian transcription factors. **(A)** Diagram of the transcription factor (TF) annotation process. The coding transcripts identified in the new genome annotation were screened for conserved TF structures by using Interproscan and the Pfam and SUPERFAMILY databases. Homologs were then searched for the TFs with BLAST, and the planarian literature was reviewed to assign screened TFs to previously described. The manually reviewed TF proteins were then used to predict motifs. **(B–E)** Identified TF families with basic domains **(B)**, **(C–E)**. ‘Old’ refers to TFs present in the *Figure 2 continued on next page*

Figure 2 continued

literature, while 'New' are TFs without a reference in the literature or new additional homologs. (F) Ternary plot of proportional expression values in the X1, X2, and Xins cell compartments. Each dot represents one single TF. The information content metric represents the enrichment of a locus to one of the three compartments (see 'Materials and methods' for details). (G) Heatmap and hierarchical clustering of proportional fluorescence-activated cell sorting (FACS) expression profiles for all TFs. (H) Histogram of the total information content of the identified motifs. (I) Example sequence logos for some of the best characterized motifs. The motifs of *pbx* (MA0782.1) and *prep* (MA0783.1) have almost identical PWMs, but they are listed as separate motif entries in JASPAR.

highest similarity to predict motifs for planarian TFs (Figure 2A). In total, we found 166 motifs that were assigned to 263 TFs with a normal distribution of motif information value (Figure 2H, Supplementary file 1). The most informative motifs were found for *zfp-7*, *pbx*, *prep*, *gli-3*, and *rfx7*, while the least informative motifs were for *hesl-2*, *pou2/3*, *irx*, *tead1*, and *lmx1* (Figure 2I).

Histone modifications and chromosome accessibility mark enhancer-like regions

To identify potential enhancer regions, we analyzed previously generated ChIP-seq data with respect to the enhancer-associated histone modification H3K4me1 in X1 cells (Mihaylova et al., 2018). Furthermore, we sequenced the epigenome of X1 cells with respect to the histone modification H3K27ac to identify genomic regions indicative of an active enhancer state. H3K27ac was enriched at promoter regions, suggesting that this epigenomic feature for this histone modification is conserved in planarians (Figure 3—figure supplement 1A, B; Gaiti et al., 2017b). We identified 37,345 H3K27ac peaks and 13,868 H3K4me1 peaks that were generally less than 200 bp wide (Figure 3A, B). At H3K4me1 peaks, the H3K27ac signal was strongest at the peak center at almost all peaks, while at H3K27ac peaks the H3K4me1 displayed a bimodal peak at the peak center of most peaks (Figure 3C, D). This pattern of H3K4me1 flanking H3K27ac peaks at enhancers has been previously described in mammals, and our data suggest that this may be evolutionarily conserved within metazoans (Gorkin et al., 2012; Spicuglia and Vanhille, 2012; Pundhir et al., 2016).

As our ChIP-seq data followed well-established enhancer-like patterns, we used the ChIP peaks to select putative enhancer-like regions. We calculated a mean peak value at all peaks with respect to the H3K27ac and H3K4me1 signal (Figure 3E) and selected all H3K27ac peaks that were at most 500 bp from H3K4me1 peaks and determined these 5529 peaks to be an initial set of putative active enhancer-like regions in cycling adult stem cells. The H3K27ac and H3K4me1 signals correlated at these enhancer-like regions more than at the other peaks in the genome (Figure 3F).

Furthermore, we calculated the mean peak value with respect to the change in H3K4me1 signal upon RNAi-mediated *Smed-lpt* knockdown (corresponding to the N-terminus of mammalian *ml3/4* or *kmt2c/d*, see Mihaylova et al., 2018). In mammals, *Mll3* and *Mll4* are two paralogous methyltransferases of the COMPASS family (SET1/MLL) that regulate enhancer activity by mono-methylating H3K4 (Wang et al., 2021). In addition, the *ml3/4* methyltransferase complex associates with the histone acetyltransferase *p300/CBP* that mediates H3K27 acetylation at enhancers and thus gives rise to the active enhancer landscape (Wang et al., 2021). The enhancer-like regions were clearly more responsive to the knockdown of *Smed-lpt* as compared to random points within the genome, and the response was most evident at the center of putative enhancer-like regions (Figure 3G, H and K). Thus, the H3K4me1 reduction after *Smed-lpt* knockdown provides further functional support for the identification of active enhancers in planarian stem cells as these are targets for active histone methylase activity associated with enhancers.

We optimized and performed ATAC-seq on X1 cells to measure high-resolution chromatin accessibility in conjunction with the histone modifications (Buenrostro et al., 2013) as a potential further source of evidence. Enhancer-like regions identified by ChIP-seq analysis had a more accessible chromatin configuration than random points in the genome, and the peaks of open regions were positioned at the center of the predicted enhancer-like regions implicated by ChIP-seq data (Figure 3I, J and L). Thus, ATAC-seq provides independent evidence for these regions being putative active enhancers in planarian stem cells.

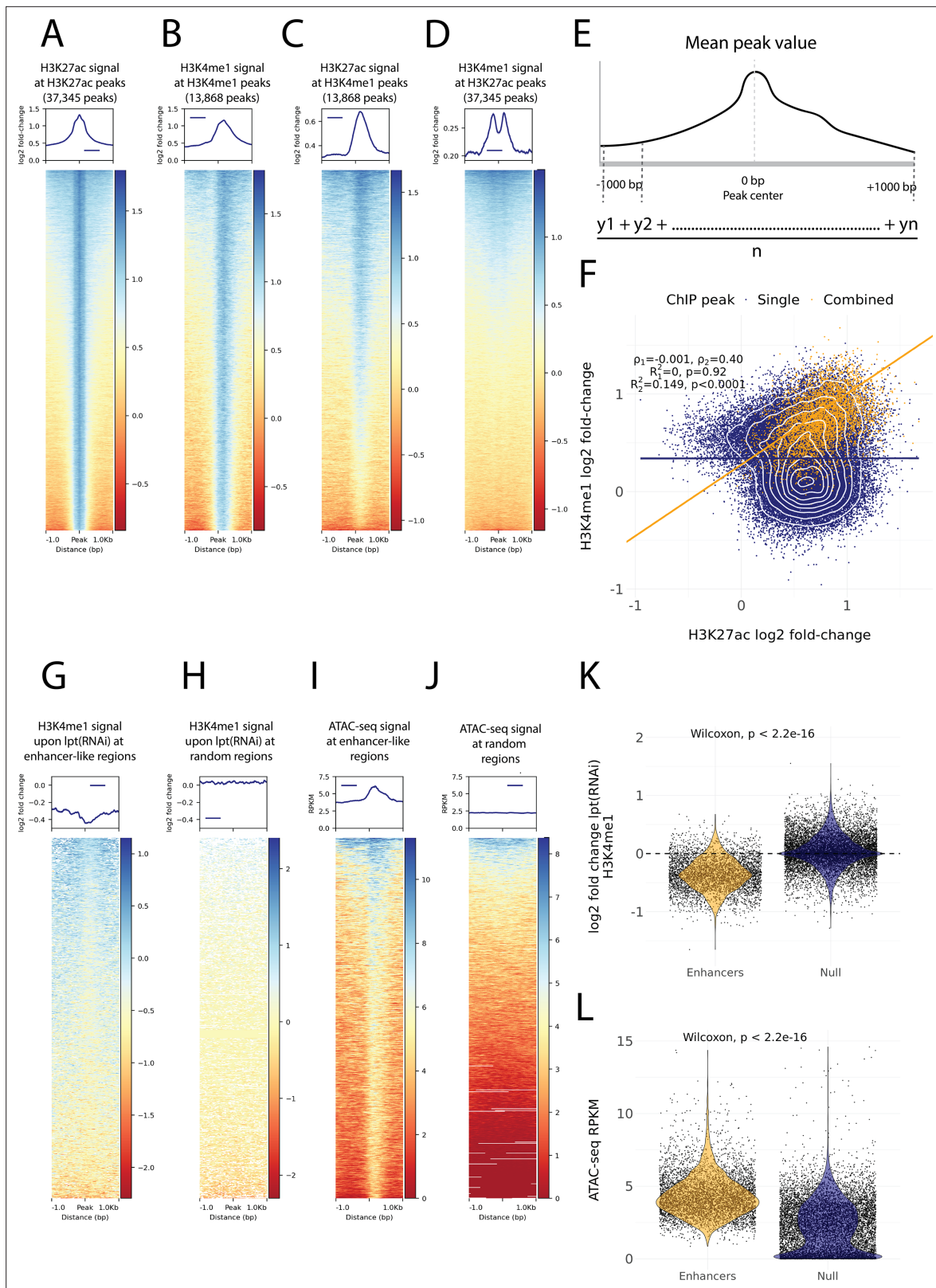


Figure 3. Histone modifications and chromatin accessibility of enhancer-like regions. **(A)** H3K27ac ChIP-seq signal around independently called H3K27ac peaks. **(B)** H3K4me1 ChIP-seq signal around independently called H3K4me1 peaks. **(C)** H3K27ac signal around the same H3K4me1 peaks as in **(B)**. The signal centers around the center region. **(D)** The H3K4me1 signal around the same H3K27ac peaks as in **(A)**. The signal displays a bimodal shape around the center region. The signal of the y-axis and the heatmap in **(A–D)** is given as the log₂ fold-change relative to the input and the peak width is

Figure 3 continued on next page

Figure 3 continued

2000 bp. (E) A visual diagram of the calculation of the mean peak value. Mean peak value is defined as the mean of signal values taken at an interval of 10 bp from 1000 bp upstream to 1000 bp downstream of the peak center. (F) Scatter plot of H3K27ac and H3K4me1 log₂ fold-change mean peak values at the ChIP-seq peaks. H3K4me1 peaks that are at most 500 bp from a H3K27ac peak are defined as combined peaks and putative enhancer-like regions (model 2, orange), while the non-overlapping H3K4me1 and H3K27ac peaks are referred to as single peaks (model 1, navyblue). (G) The change in the H3K4me1 signal upon *lpt* (mll3/4) RNAi at the combined peaks or putative enhancer-like regions. The change is strongest at the peak center. The log₂ fold-change is calculated as the RNAi value relative to the wildtype value. (H) The H3K4me1 *lpt*(RNAi) signal at 10,000 random regions in the genome. (I) The ATAC-seq signal at the combined peaks or putative enhancer-like regions. The signal is given as reads per kilo base per million mapped reads (RPKM) (J) The ATAC-seq signal at 10,000 random regions in the genome. (K) Violin plot of H3K4me1 *lpt*(RNAi) signal at putative enhancer-like regions (n = 5529) and random regions (n = 10,000) in the genome. The H3K4me1 signal is reduced at enhancers. Wilcoxon rank-sum test is used for pairwise comparisons. (L) Violin plot of the ATAC-seq signal at putative enhancer-like regions (n = 5529) and random regions (n = 10,000) in the genome. The ATAC-seq signal is higher at enhancers. Wilcoxon rank-sum test is used for pairwise comparisons.

The online version of this article includes the following figure supplement(s) for figure 3:

Figure supplement 1. Histone modifications at transcription start sites and footprinting scores in relation to enhancer-like regions.

Gene regulatory networks involving fate-specific transcription factors in neoblasts

Having defined enhancer-like regions (**Supplementary file 2**), integration with both TF and potential target gene expression data allowed us to begin constructing preliminary GRNs to demonstrate the utility of our neoblast enhancer predictions. First, we assigned enhancer-like regions to their closest gene, assuming a high probability that these will be the putative target genes of enhancer-like regions (**Supplementary file 2**). The distance from the enhancer-like regions to the target gene transcription start site (TSS) varied from being proximal to the promoter region to being as far as 89 kb away (**Figure 4A**). We found that TFs themselves were significantly enriched within the set of all predicted target genes (**Figure 4B**). Furthermore, these TFs were in turn enriched for the 43 FSTFs previously shown to be expressed in S/G2/M neoblasts (**Figure 4B; Raz et al., 2021**), again suggesting that our predicted set of enhancers are real. In addition to TF activity, the predicted target genes linked to enhancer-like regions were enriched for RNA metabolic and RNA biosynthetic processes and other biosynthetic processes, as well as transcription and regulation of gene expression (**Figure 4C**). These data suggest that our enhancer predictions include real enhancers involved in regulating key aspects of neoblast biology.

To establish more persuasive and direct regulatory links between TFs and enhancer-like regions, we used ATAC-seq footprinting to detect potentially bound motifs in the planarian stem cell genome (**Bentsen et al., 2020; Figure 4D, Supplementary file 3**). ATAC-seq footprints are short inaccessible or less accessible regions within an otherwise accessible region, indicative of DNA binding by a TF or another DNA-binding protein (**Bentsen et al., 2020; Figure 4D**). As for the raw ATAC-seq signal, footprint scores were higher in enhancer-like regions than in random regions of the genome (**Figure 3—figure supplement 1C, D**). Overall footprinting analysis found 22,489 putatively bound TF motifs in the enhancer-like regions but no bound motifs in random regions, providing further support that these regions are potential enhancers. The TFs with the most binding sites predicted by footprint analysis in the genome overall were *irx-1/irx6-1/2/3*, *egr4-1/2*, *smad1/9*, *sp5*, and *nf-ya1/2*, while the TFs with the least bound motifs were *mef2-1*, *mef2-2*, *phox*, *pou6-2*, and *lmx1-2* (**Figure 4ESupplementary file 3 and Supplementary file 4**). Overall, these data provide a resource of studying GRNs and specific regulatory interactions in planarians neoblasts. We found no obvious difference in the distance of predicted TF footprints from TSSs for different TFs (**Figure 4—figure supplement 1**).

To test the utility of these data for building and testing GRNs, we focused on the previously defined FSTFs, established the regulatory links between them, and constructed a putative stem cell GRN based on our datasets (**Figure 4F, Supplementary file 5**). In total, we could include 35/43 FSTFs into a GRN prediction and found evidence for multiple cross-regulatory links that may serve to allow stem cells to decide between fates. For example, neural FSTFs regulate enhancers of muscle FSTFs and vice versa (**Figure 4F**). The interactions in this GRN allow specific hypotheses to be formed and tested in the future. As examples, we present in detail the genomic region of the epidermal FSTF *p53*, the muscle FSTF *myoD*, and the neural FSTFs *nkx6-1* and *soxB1-1* (**Figure 5A–D**). The epidermal FSTF *p53* has multiple TFs predicted to bind at the promoter region, including a muscle FSTF (*myoD*) (**Scimone et al., 2014; Scimone et al., 2017**), and TFs relating to position (*zic-1*) (**Vásquez-Doorman**

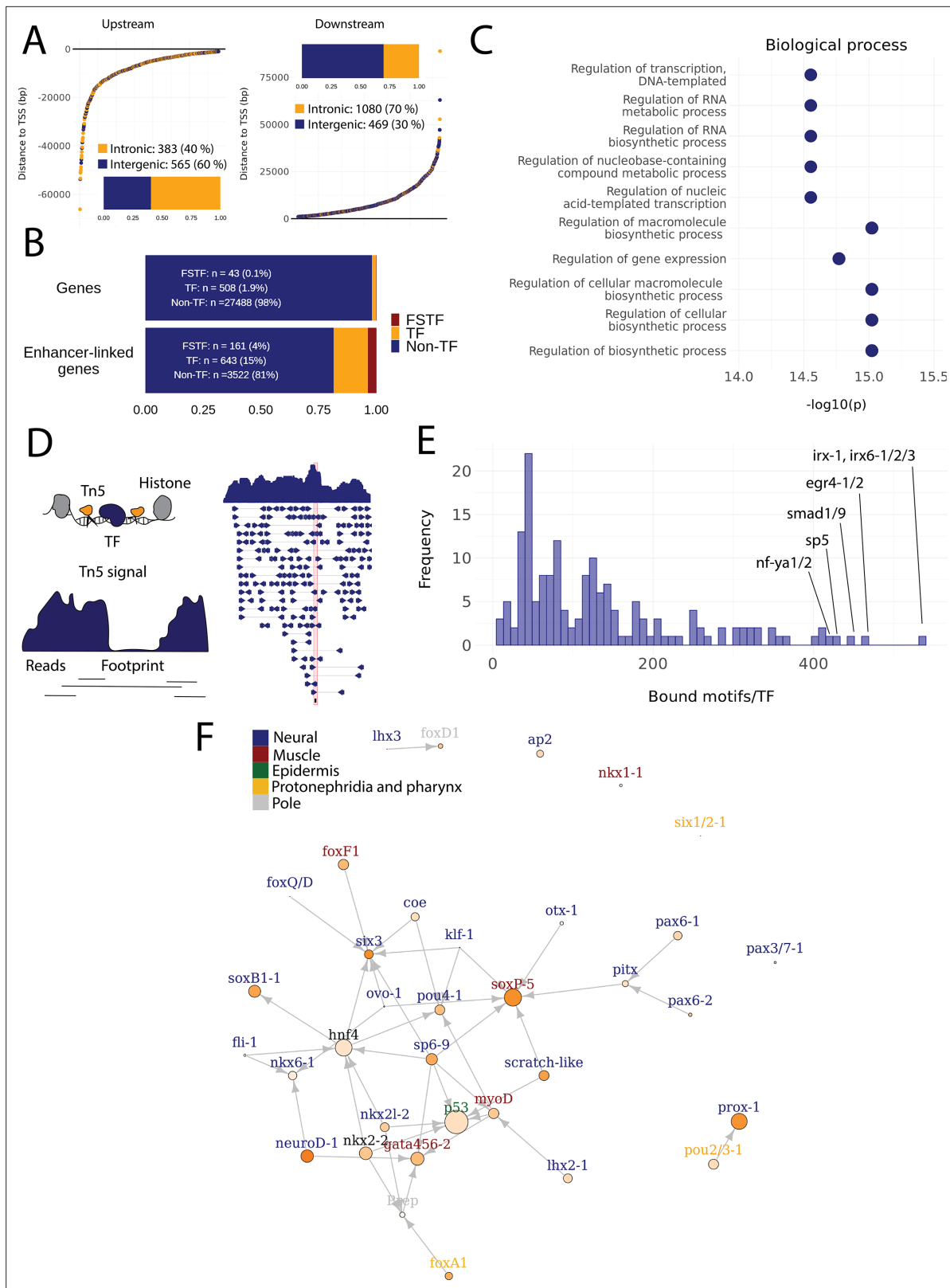


Figure 4. Predicted gene regulatory networks involving fate-specific transcription factors in neoblasts. **(A)** Distance of enhancer-like regions to the transcription start site (TSS) of nearest genes. The nearest genes are assumed to be the putative target genes of the enhancer-like regions. **(B)** The proportion of fate-specific transcription factors (FSTFs), transcription factors (TFs), and non-TF genes of all planarian genes and of the target genes linked to enhancer-like regions. FSTFs and TFs are enriched in the set of target genes ($\chi^2 = 2660, p < 0.001$). **(C)** Gene Ontology (GO) analysis of

Figure 4 continued on next page

Figure 4 continued

biological processes of enhancer-linked target genes compared to all genes in the planarian genome. (D) Schematic overview of ATAC-seq footprinting (Bentsen et al., 2020). Both the tagging enzyme Tn5 and TF bind to accessible regions of the genome. Footprints are short and sharp inaccessible segments in otherwise accessible regions. The genomic track figure represents an example of ATAC-seq reads and footprint score overlapping with an *sp5* motif in the genome. (E) Histogram of the total number of bound motifs in the genome for each TF. The TFs with most bound motifs in the genome are marked. (F) Putative gene regulatory network (GRN) of FSTFs in neoblasts. The epidermal FSTFs are labeled in green, the pole-associated FSTFs are labeled in gray, the muscle FSTFs are labeled in red, the intestinal FSTFs are labeled in black, and the neural FSTFs are labeled in blue. The size of the nodes reflects the absolute expression in X1 cells (transcripts per million reads [TPM]), and the color of the node reflects the proportional expression in X1 cells (the more orange, the higher proportional expression). The arrows represent a regulatory link but do not discriminate between positive and negative interactions.

The online version of this article includes the following figure supplement(s) for figure 4:

Figure supplement 1. Distance of bound transcription factor footprints to transcriptional start sites.

and Petersen, 2014) and cell migration (*zeb-1* and *snail-1/2*) (Abnave et al., 2017; Figure 5A). An enhancer-like region within an intron of the upstream gene *ythdc23* was predicted to be associated with the downstream muscle FSTF *myoD* as the TSS of *myoD* is closer of the two and *ythdc23* is lowly expressed in X1 and X2 cells. The enhancer-like region includes bound TF motifs such as the neoblast-enriched zinc finger *fir-1*, the muscle segment homeobox gene *msx*, the neural *pou-p1*, *sp6-9*, *lhx2-1*, and position-related *foxA(P)* (Figure 5B). A distal enhancer-like region was found downstream of *nkx6-1*, including potentially bound TF motifs such as the neoblast-enriched *egr-1*, the neural *ovo-1* and *fli-1*, the muscle-related *twist*, the position-related *smad4-1*, and the pigmentation-related *ets-1* (Figure 5C). A distal enhancer-like region was found upstream of *soxB1-1*, including bound TF motifs such as the position-related *smad1* and *foxA(P)* (Molina et al., 2007; Reddien et al., 2007; Pascual-Carreras et al., 2020), neural *da* (Cowles et al., 2013), intestinal *hnf4* (Wagner et al., 2011, Scimone et al., 2014; van Wolfswinkel et al., 2014), and *zeb-1* and *snail-1/2* relating to cell migration (Abnave et al., 2017; Figure 5D). In the future, these putative regulatory links can be verified and studied by experiments using functional genomics involving the RNAi knockdown of individual genes and subsequent RNA-seq and ATAC-seq analysis.

Multiple enhancers are linked to unstudied transcription factors

In addition to the relatively well-known planarian FSTFs (Raz et al., 2021), we investigated GRNs relating to less-well-studied TFs with multiple bound motifs in associated enhancer-like regions as these may also have a potentially central role in the neoblast regulation. We defined the number of enhancer-like regions, number of putative bound motifs, and number of putative unbound motifs for each TF (Figure 6A, Supplementary file 6). In addition, we selected TFs expressed in the X1 compartment (proportional expression of X1 > 1/3) and constructed the putative GRN summarizing all the 502 regulatory links of these 109 TFs (Figure 6B). In the X1 compartment (X1% > 1/3), the least-well-studied TFs with the most bound motifs in enhancer-like regions were *znf596*, *tbx-20*, *hesl-1*, *atoh8-1*, and *ikzf1* (Figure 6A, B, Supplementary file 2 and Supplementary file 6). Outside the X1 compartment, *msx*, *zf-6*, *vsx*, and *pdp-1* had the most bound motifs (Figure 6A).

The TF with most regulatory interactions, *znf596*, has been characterized to be expressed in neoblasts and more specifically in the neoblasts committed to the neural fate, but otherwise its function is unknown (Figure 6A–C, Supplementary file 6; Fincher et al., 2018). We found that numerous FSTFs were predicted to bind to putative intronic enhancers within *znf596*, including neural FSTFs *sp6-9*, *neuroD-1*, and *ovo-1*, the intestinal FSTF *gata456-1* and/or the muscle FSTF *gata456-1*, and the position control genes (PCGs) *smad1/9* (BMP signaling), *isl-1*, and *prep* (Figure 6C). In single-cell transcriptomics data, the gene is clearly expressed in a subset of neoblasts, calling for further mechanistic studies of its regulatory function. The binding motif of *znf596* could not be predicted, so its target genes cannot be implicated within the current datasets and approaches.

We identified the same five planarian *tbx* genes that have been previously reported, namely, *tbx-1/10*, *tbx-2/3a*, *tbx-2/3b*, *tbx-2/3c*, and *tbx-20* (Tewari et al., 2019). In mammals, the *tbx* genes have multiple roles in development, and *tbx3* is involved in regulating pluripotency of embryonic stem cells (Baldini et al., 2017; Khan et al., 2020). Nonetheless, the function of planarian *tbx* genes has not been clarified. Here, we found several predicted bound motifs of *tbx* genes in the genome (*tbx-1/10* 91 motifs, *tbx-2/3a/c* 81 motifs, *tbx-20* 62 motifs) and multiple bound motifs at enhancer-like regions

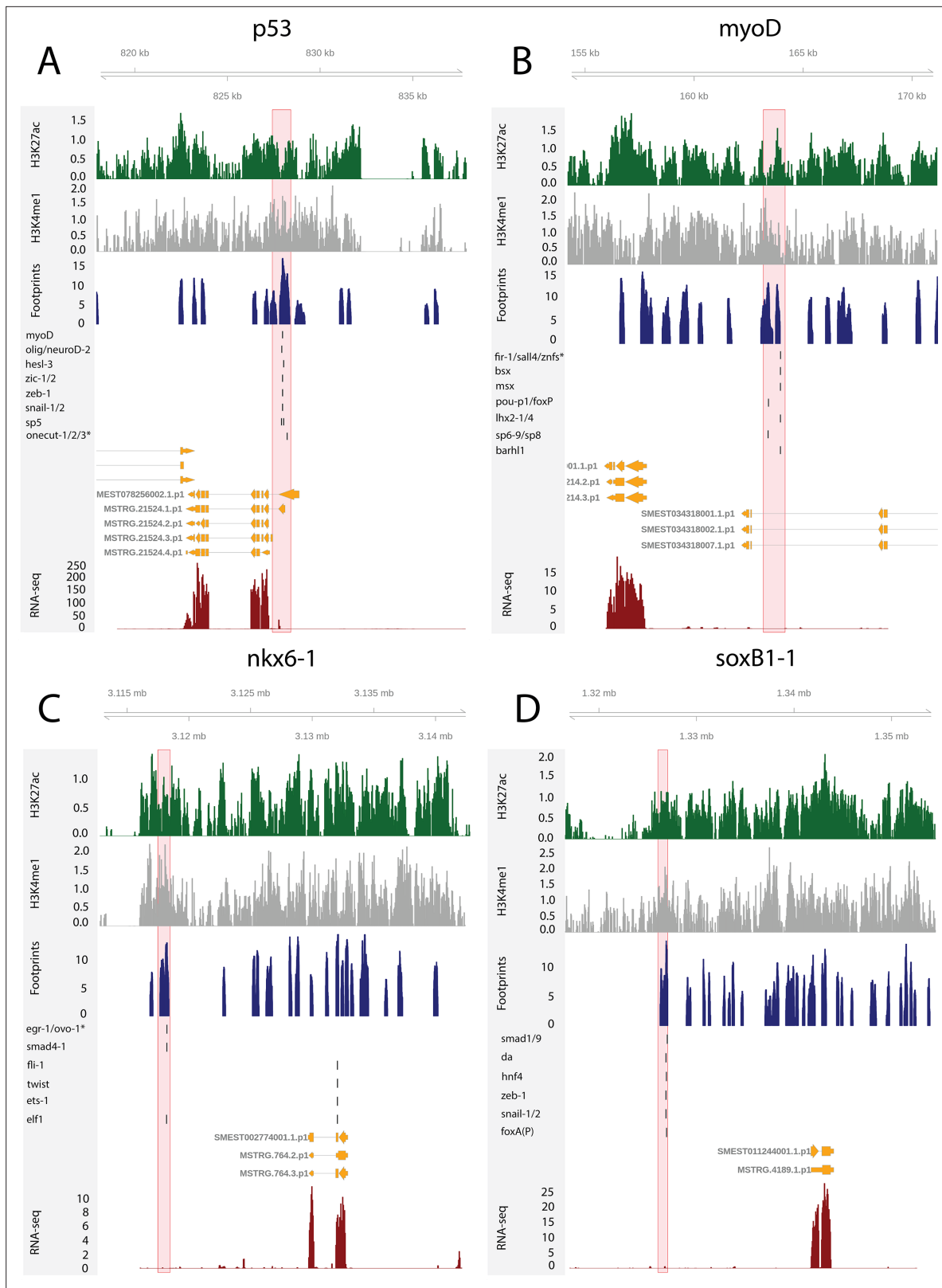


Figure 5. The genomic tracks of four fate-specific transcription factors (FSTFs) (A–D). The first track is the H3K27ac signal (log2 scale compared to the input sample), the second track is H3K4me1 signal (log2 scale compared to the input sample), the third track is the ATAC-seq footprint score (TOBIAS footprint score), the fourth track represents bound motifs in an enhancer-like region (as predicted by TOBIAS), the fifth track represents the transcripts of the gene in the new annotation, and the sixth track represents the gene expression level of the gene in X1 cells (RNA-seq fragments per kilobase

Figure 5 continued on next page

Figure 5 continued

of transcript per million [FPKM] mapped reads). **(A)** Genomic track of the epidermal FSTF *p53*. The promoter region marked in red (width 1000 bp) has multiple bound motifs. **(B)** Genomic track of the muscle FSTF *myoD*. An enhancer-like region marked in red (width 1000 bp) is located in the intron of the gene YTH domain containing 2 (*ythdc23*). The enhancer-like region is closer to the *myoD* promoter than the *ythdc23* promoter and *ythdc23* is not expressed almost at all in neoblasts, and hence the enhancer-like region is likely linked to *myoD* rather than *ythdc23*. **(C)** The neural FSTF *nkx6-1* has a distal enhancer-like region marked in red (width 1000 bp) downstream of the gene body. **(D)** The neural FSTF *soxB1-1* has a distal enhancer-like region marked in red (width 1000 bp) upstream of the gene body.

linked to *tbx* genes (*tbx-20* 142 motifs, *tbx-2/3c* 34 motifs, *tbx-2/3a* 11 motifs) (**Figure 6A, B and D, Supplementary file 2** and). *tbx-20* had a clear intronic enhancer-like region and a distal enhancer-like region, both containing motifs for TFs implicated as positional control genes (PCGs) (*hox1*, *hox4a*, *prep*, *smad1*, *smad9*, *zic-1*, *sp5*) (**Molina et al., 2007, Reddien et al., 2007; Felix and Aboobaker, 2010; Vásquez-Doorman and Petersen, 2014; Tewari et al., 2019**), TFs involved in neoblast migration (*snail-1* and *snail-2*) (**Abnave et al., 2017**), and various other TFs (**Figure 6D**).

Elements of the planarian positional gene regulatory network are active in neoblasts

Next, we studied enhancer-like regions and potential regulatory links associated with the well-known planarian PCGs. Planarians are a primary model system to understand how positional information guides and directs stem cell function during regeneration (**Reddien, 2018**), and therefore some functional genomics data exists with regard to PCGs, enabling the comparison with our findings (**Tewari et al., 2019**). A constitutive positional information system is established by the regional expression pattern of PCGs that pattern the anterior-posterior (AP), dorsal-ventral (DV), and medial-lateral (ML) axes (**Reddien, 2018**).

The AP axis is patterned by the Wnt signaling pathway: high Wnt activity specifies posterior identity, while low activity specifies anterior identity (**Reddien, 2018**). Upon β -catenin knockdown, posteriorly expressed *sp5* and the Hox genes *post-2c*, *post-2d*, *lox5a*, and *hox4b* are rapidly downregulated (**Tewari et al., 2019**). In multiple vertebrate species, *sp5* is known to be a direct target of Wnt signaling (**Weidinger et al., 2005; Fujimura et al., 2007**), and this regulatory link seems to be conserved in planarians (**Tewari et al., 2019**). Here, we found a putative bound *sp5* footprint in the vicinity of the *post-2c* and *lox5a* promoters, suggesting and further supporting that *sp5* regulates the expression of these posterior PCGs (**Figure 7A**). *sp5* had 431 putative bound motifs and was the fourth most bound motif in X1 cells overall, further suggesting that *sp5* mediates the broad positional information provided by Wnt signaling in planarians (**Supplementary file 2**).

Through ATAC-seq and ChIPmentation techniques, **Pascual-Carreras et al., 2020** screened for *cis*-regulatory elements in planarian tissues in *notum* and *wnt1* (RNAi) animals. Upon *wnt1* knockdown, posterior Hox genes *hox4b*, *post-2c*, *post-2b*, *lox5a*, *lox5b* and *wnt11-1*, *wnt11-2*, *fzd4*, and *sp5* were downregulated, replicating the results of β -catenin knockdown (**Pascual-Carreras et al., 2020**). In addition, two *foxG* binding sites were found in the first intron of *wnt1*, and *foxG* knockdown was found to phenocopy *wnt1* knockdown, supporting the hypothesis that *foxG* is an upstream regulator of *wnt1* (**Pascual-Carreras et al., 2020**). Here, we found one enhancer-like region in the first intron of *wnt1* with a high level of H3K27ac, H3K4me1, and ATAC-seq footprinting scores for Fox family TFs (**Figure 7B**). This motif implicated by footprinting analysis is the same as one of the motifs described by previous work (**Pascual-Carreras et al., 2020**). We did not see evidence of binding at the second motif, and neither was this motif within one of our predicted enhancers (**Figure 7B**).

Although the planarian Hox genes are expressed in a regionalized manner along the AP axis, the knockdown of the genes apart from *post-2d* does not result in homeostatic or regeneration-associated phenotypic changes (**Currie et al., 2016; Arnold et al., 2021**). Instead, the five Hox genes *hox1*, *hox3a*, *hox3b*, *lox5b*, and *post2b* have been shown to be involved in asexual reproduction by regulating fission at potential cryptic segment boundaries (**Arnold et al., 2021**). Here, we did not find any predicted bound motifs implicated by footprinting associated with *hox1*, *hox3a*, and *hox3b* (**Supplementary file 2**), and only a few at the promoters of *lox5b* and one at *post2b* (**Supplementary file 2, Figure 7—figure supplement 1A and B**). The proportional expression values of these Hox genes in X1 cells are 4% (*hox1*), 3% (*hox3a*), 63% (*hox3b*), 8% (*lox5b*), and 31% (*post2b*). Altogether,

Figure 6 continued

absolute expression in X1 cells (transcripts per million reads [TPM]), and the color of the node reflects the proportional expression in X1 cells (the more orange, the higher proportional expression). The genomic tracks of *znf596* (C) and *tbx-20* (D). The first track is the H3K27ac signal (log₂ scale compared to the input sample), the second track is H3K4me1 signal (log₂ scale compared to the input sample), the third track is the log₂ fold-change upon *lpt(RNAi)*, the fourth track is ATAC-seq footprint score (TOBIAS footprint score), the fifth track represents bound motifs in an enhancer-like region (as predicted by TOBIAS), the sixth track represents the transcripts of the gene in the new annotation, and the seventh track represents the gene expression level of the gene in X1 cells (RNA-seq fragments per kilobase of transcript per million [FPKM] mapped reads). (C) Genomic track of *znf596*. Two intronic enhancer-like regions are marked in red (width 500 bp). (D) Genomic track of *tbx-20*. Two intronic (width 500 bp) and one distal enhancer-like region (width 1200 bp) are marked in red.

this suggests that these Hox genes are not driving fission through their regulatory activity in adult stem cells.

Prep, *zic-1*, *isl-1*, and *foxD1* are PCGs that are expressed in the anterior pole of both intact and regenerating planarians (Felix and Aboobaker, 2010; Vásquez-Doorman and Petersen, 2014; Vogg et al., 2014). Interestingly, we found that these TFs are bound to several motifs in enhancer-like regions of X1 cells (Supplementary file 1 enhancers, *prep* 132 motifs, *zic-1/2* 330 motifs, *isl-1* 42 motifs, *foxD1/2/3* 42 motifs) and a few motifs were found to bind to the enhancer-like regions linked to *prep*, *zic-1*, *isl-1*, and *foxD1* (Figure 7C, Supplementary file 2, Supplementary file 5, and Supplementary file 6). *Prep* has enhancer-like regions that have bound motifs of smad factors (Bmp signaling components) and *nkx2-2/3/4* (Figure 7C). Knockdown of *nkx2-2* (also known as *DTH-1* or *nkx2.2*) causes blastemal defects both at the anterior and posterior ends, while knockdown of *prep* leads to defects in the anterior, suggesting that *nkx2-2* might work upstream of *prep* at the anterior end (Felix and Aboobaker, 2010; Forsthoefel et al., 2020).

Lastly, we studied the Hedgehog signaling receptor *patched-1* (*ptch-1*) and found an intronic enhancer-like region containing multiple bound motifs, including the Hedgehog pathway TFs *gli-1/2* as would be expected, for the Wnt effector *sp5*, the Hox genes *lox5a/b*, BMP signaling components (*smad4-1*, *smad1/9*), the anterior pole TFs *prep* and *zic-1*, the cell migration factors *snail-1/2*, and the muscle TFs *myoD* and *twist* (Figure 7D). Hedgehog signaling in planarians is known to have pleiotropic functions, including interaction with the Wnt signaling pathway and glial cell function (Rink et al., 2009; Yazawa et al., 2009; Wang et al., 2016). Taken together, we found that elements of positional control gene regulatory network are active in planarian S/G2/M neoblasts and an integral part of the overall GRN determining neoblast behavior (Figure 7E).

Available transcriptomic data supports putative gene regulatory networks

As an independent test of the regulatory links between TFs predicted by our GRNs, we used transcriptomic data from published RNAi knockdown experiments for TFs that had predicted footprints. We analyzed transcriptomic data for *coe(RNAi)* (Cowles et al., 2014), *foxD1(RNAi)* (Vogg et al., 2014), *pax2/5/8-1* (Stower's Institute, 2016), *myoD(RNAi)* (Scimone et al., 2017), *nkx1-1* (Scimone et al., 2017), *hox1*, *hox3a*, and *lox5b* (Arnold et al., 2021). These available datasets had been generated from experiments using whole worms or blastemal tissues rather than the stem cell population. Nonetheless, in all cases both stem cells (X1 and X2), recent stem cell progeny (X2), and differentiated tissues (Xins) will have been present in these samples. We performed differential gene expression (DGE) analysis (Supplementary file 7) and obtained differentially expressed transcription factors (DETFs) as an independently derived set of functionally validated potential target TFs (Supplementary file 8).

To investigate whether the predicted regulatory links of our putative GRNs were supported by this collection of RNAi knockdown transcriptomic data, we compared the frequency of observed DETFs with predicted regulation by the RNAi-targeted TF in our GRNs to the expected frequency with random assignment (Figure 8A–C). The null hypothesis assumes that differential expression and enhancers are randomly assigned to genes and that footprints are randomly assigned to the enhancers (for details, see 'Materials and methods'). On average across these eight TFs with RNAi transcriptome data, we discovered 10.38 times more regulatory links between TFs than expected supported by transcriptome data with unbound footprints in a predicted enhancer. For TFs with predicted bound footprints, we found on average 21.09 times more regulatory links than expected (unbound p-value < 0.0001,

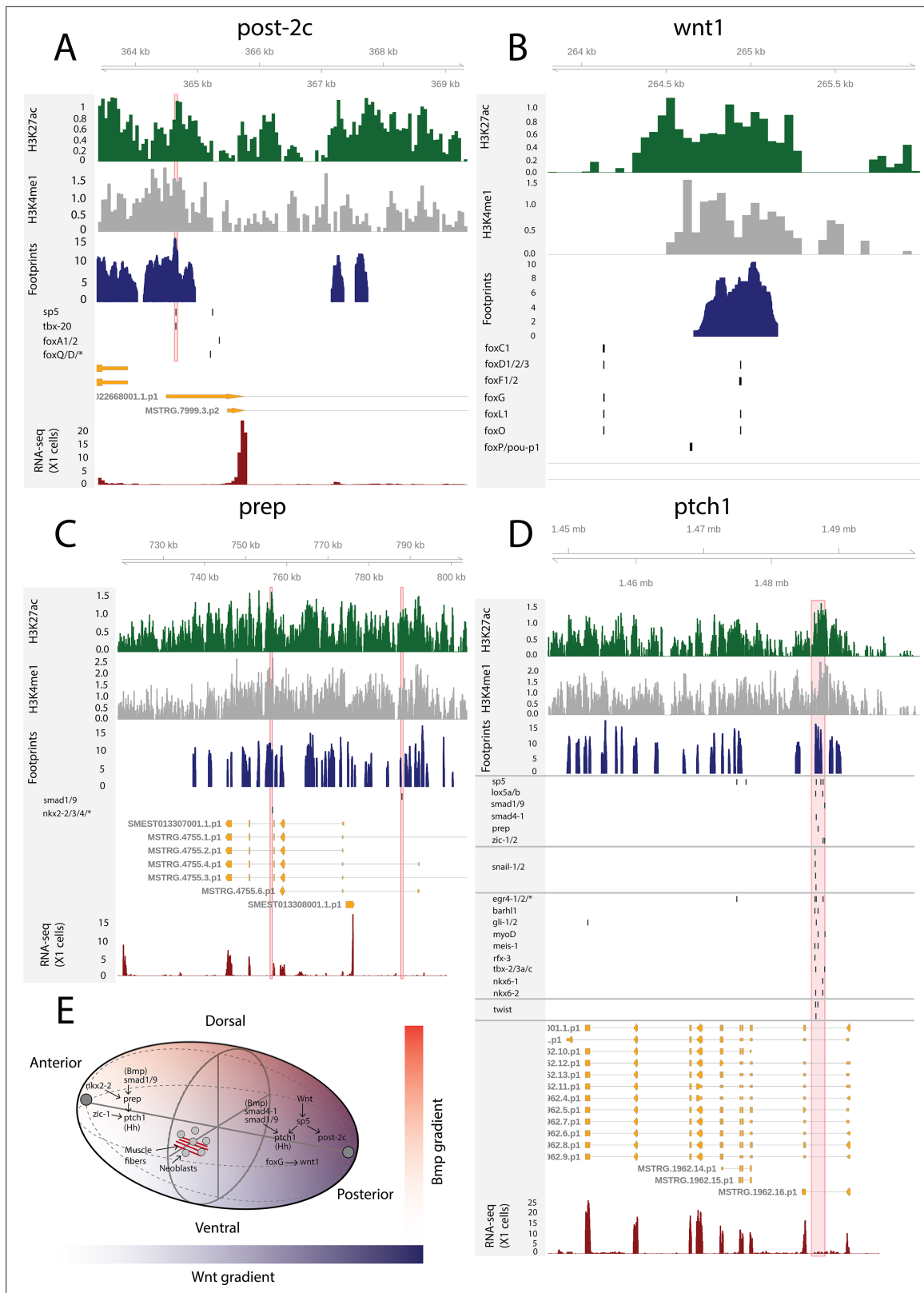


Figure 7. The genomic tracks of *post-2c* (A), *wnt1* (B), *prep* (C), and *ptch1* (D). The first track is the H3K27ac signal (log2 scale compared to the input sample), the second track is H3K4me1 signal (log2 scale compared to the input sample), the third track is ATAC-seq footprint score (TOBIAS footprint score), the fourth track represents bound motifs in an enhancer-like region (as predicted by TOBIAS), the fifth track represents the transcripts of the gene in the new annotation, and the sixth track represents the gene expression level of the gene in X1 cells (RNA-seq fragments per kilobase of transcript per *Figure 7 continued on next page*

Figure 7 continued

million [FPKM] mapped reads). (A) Genomic track of *post-2c*. One upstream *cis*-regulatory regions is marked in red (width 50 bp). The foxQ/D* footprint represents motif MA0851.1 corresponding to TFs foxQ/D, foxJ1-1, foxJ1-2, foxJ1-3, foxJ1-4, foxJ1-5, foxN2/3-1, foxN2/3-2, and foxN2/3-3. (B) Genomic track of *wnt1*. One intronic enhancer-like region with one potential Fox family TF binding motif with evidence of binding from footprinting analysis, other Fix TF motifs lie outside the enhancer predicted by ChIP-seq data (Pascual-Carreras et al., 2020). (C) Genomic track of *prep*. Two intronic enhancer-like regions are marked in red (width 500 bp). (D) Genomic track of *ptch1*. One intronic enhancer-like region (width 2000 bp) is marked in red. (E) Schematic of position control gene (PCG) regulatory links discussed.

The online version of this article includes the following figure supplement(s) for figure 7:

Figure supplement 1. Regulatory regions in relation to Hox genes.

bound p -value < 0.0001; **Figure 8A–C**). Results varied between knockdown experiments, indicating that the whole worm transcriptomic data reflects stem cell transcriptomics better for some TFs than others (**Figure 8A–D**). The predictive power of the GRNs developed in our study correlates with the proportional expression of the knockdown TF in the X1 compartment, which only contains stem cells, and inversely correlates with the level of expression in the Xins compartment, which only contains differentiated cells (**Figure 8D**, **Figure 8—figure supplement 1A and B**), suggesting that our GRNs are indeed reflective of activity in stem cells but not necessarily in postmitotic differentiating or differentiated cells.

For instance, we examined the potential regulatory links of *coe* in our GRN model and found eight putative target TFs associated with an enhancer containing a potentially bound footprint for *coe*. Upon *coe*(RNAi) of whole worms, we observed eight DETFs, and by collating these two datasets, we obtained *pou4-1* as the only DETF containing a putative bound *coe* footprint (**Figure 8D**). This putative footprint is located within an enhancer downstream of the gene body (**Figure 8D**). Indeed, Cowles et al., 2014 tested DETFs in their transcriptomic analysis and observed that both *coe*(RNAi) and *pou4-1*(RNAi) lead to neural defects, demonstrating that the regulatory link corroborated by our GRN could be functionally validated within the limits of the experimental toolkit of planarians. As for unstudied zinc fingers, we found that *zfhx3* is differentially expressed in *lox5b*(RNAi) whole worms and has a putative bound footprint of *lox5b* within an enhancer upstream of the gene body (**Figure 8E**). These analyses overall independently validate some of our predictions and provides a set of high-confidence predictions for further studies of TFs that were found to be differentially expressed after RNAi of TFs predicted to bind a nearby enhancers.

Discussion

In this study, we improved the current genome annotation of *S. mediterranea* by integrating available RNA-seq data and identified new coding and non-coding transcripts. We reviewed the literature on both computationally and experimentally derived planarian TFs and defined a high-confidence set of TFs. If possible, we also predicted binding motifs for the planarian TFs. We developed and performed ATAC-seq on the proliferating stem cell compartment and determined genomic regions of open chromatin. We analyzed genome-wide profiles of the histone modifications H3K27ac and H3K4me1, along with chromatin accessibility data, and used these epigenetic datasets to delineate putative active enhancers in planarian stem cells. Lastly, we identified TFs binding to the enhancers of potential target genes and constructed hypothetical GRNs active in the stem cells.

We found more than 50,000 new isoforms to known transcripts in our annotation. This information has already helped to improve the clustering of single-cell RNA-seq data (García-Castro et al., 2021). Our approach demonstrated that the principle of integrating available RNA-seq data into a comprehensive expression-driven annotation can significantly improve genome annotation. We were also able to annotate more than 7000 new loci, with over 3000 predicted to be coding, increasing the number of protein coding genes by more than 10%. These new proteins tend to be shorter, are more likely to be proteins specific to planarians, but are expressed at similar levels to the rest of the coding transcriptome. This, together with the extensive annotation of alternate splice forms, gives a more complete picture of the genome of the model planarian.

Altogether, we annotated 551 TFs that were distributed evenly with respect to enrichment across all three FACS cell compartments in planarians (X1, X2, and Xins). In planarians, most studied functionally have been involved in regulating differentiation. Here, we define many neoblast-enriched

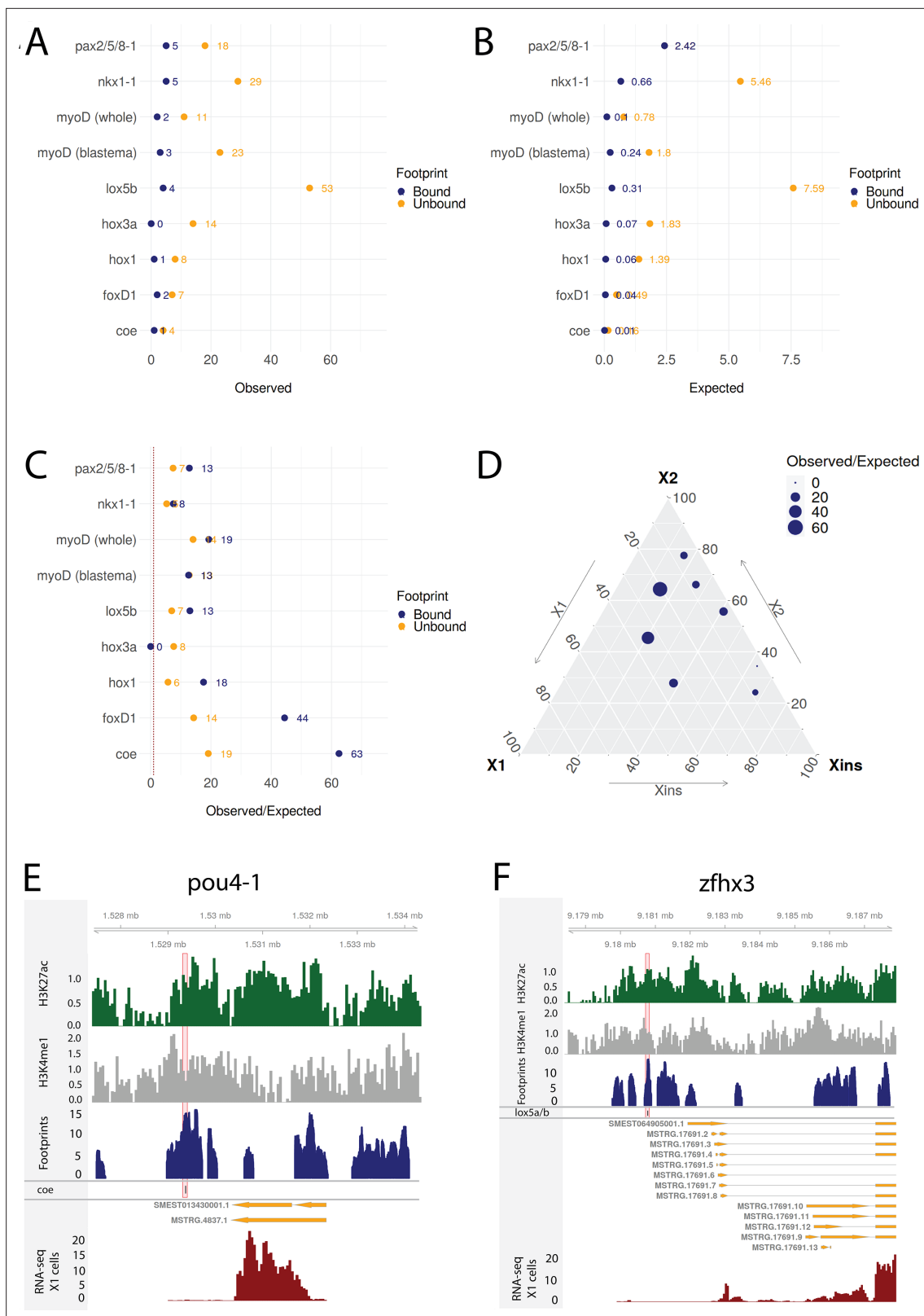


Figure 8. Transcriptomic data from knockdown experiments supports predicted regulatory links. **(A)** The number of observed regulatory links between the knocked down transcription factor (TF) (y-axis) and differentially expressed TFs in our gene regulatory network (GRN) model. The regulatory links with unbound and bound footprints are displayed separately. **(B)** The number of expected regulatory links between the knocked down TF (y-axis) and differentially expressed TFs. **(C)** The ratio of observed and expected regulatory links as obtained from transcriptomic data for knocked down TFs. The *Figure 8 continued on next page*

Figure 8 continued

regulatory links with unbound and bound footprints are displayed separately. (D) A ternary plot displaying the observed/expected ratio alongside the proportional expression values of the X1, X2, and Xins compartments. The size of the dot corresponds to the observed/expected ratio. The ratio is higher closer to the X1 and X2 compartment maxima, indicating that the GRN model predicts links better for TFs with higher expression in stem cells. (E) Genomic track of *pou4-1*. One coe bound footprint within a downstream enhancer-like region (width 100 bp) is marked in red. (F) Genomic track of *zfh3*. One *lox5a/b* bound footprint within an upstream enhancer-like region (width 100 bp) is marked in red.

The online version of this article includes the following figure supplement(s) for figure 8:

Figure supplement 1. Correlation of compartment-specific proportional expression and the enrichment of observed regulatory links.

TFs that have not been formally studied, with potential roles in maintaining neoblast pluripotency and potentially early lineage commitment. A future systematic functional screen of these may help to uncover the GRN network that maintains pluripotency analogous to that in mammals and vertebrates (Takahashi and Yamanaka, 2006; Takahashi et al., 2007). In particular, we found a large number of uncharacterized planarian zinc finger TFs, and this large and diverse family of TFs is still relatively poorly studied, perhaps because they evolve relatively rapidly in metazoans in general and often homology cannot be confidently assigned across phyla (Albà, 2017; Cassandri et al., 2017; Najafabadi et al., 2017). Currently little is known about transcriptional control of pluripotency in adult stem cells across animals, outside of vertebrates. In planarians, both the NuRD (Jaber-Hijazi et al., 2013) and BRG/Brahma complexes Onal et al., 2012 have been implicated in regulating pluripotency through their RNAi phenotypes that block differentiation without affecting stem cell maintenance/self-renewal. By analogy with vertebrates, these chromatin remodeling complexes may directly regulate the activity of pluripotency TFs, which may include some unstudied TFs in our predicted GRNs.

Overall, for the main goals of this study, our analysis and identification of TFs allowed us to confidently assign likely binding motifs to just under half of the annotated planarian TFs (Figure 2E, Supplementary file 1). In the future as the number of planarians studies using ATAC data increases, and by looking at the actual sequence motifs implicated by footprinting analyses, it should be possible to refine motifs for planarian TFs and even define motifs for some TFs to which motifs could not be assigned, for example, in the diverse zinc finger TF group.

Based on combining epigenomic experiments and data types, we could identify putative intergenic and intronic enhancers in the planarian the genome of proliferating stem cells. The combined use of ChIP-seq data, RNAi of a histone methyltransferase combined with ChIP-seq, ATAC-seq data, and footprinting analyses together provided strong evidence for the identification of bona fide planarians stem cell enhancers. In the future, genome-wide ChIP-seq against the transcriptional cofactor p300 could serve as a complementary high-throughput approach to further substantiate these enhancers (Visel et al., 2009; Schwaiger et al., 2014), if available antibodies recognize the planarian ortholog of this protein (Fraguas et al., 2021; Stelman et al., 2021).

Verification of the function and targets of planarian enhancers is currently not possible using traditional approaches as no transgenic reporter technologies enabling enhancer-reporter constructs to be assayed are available. Therefore, we are bound to rely on less direct evidence from genome-wide sequencing technologies. Here, we assigned enhancers to target genes based on distance and the expression of the target gene, similar to established protocols used by others (Duren et al., 2017). We find cases where regulatory interactions suggested by previous expression-based studies of neoblast fate control and RNAi-based studies of gene function are supported by the GRNs interactions uncovered by our analyses. We find evidence that regulators of one differentiation lineage bind the enhancers of genes that regulate another, presumably acting as repressors. Our predicted GRNs were statistically supported by differential expression analyses of RNA-seq sets collected after RNAi of specific TFs (Figure 8), suggesting that many of the predicted regulatory interactions may be real. Experiments based on the predictions from our data combining RNAi against TFs with RNA-seq and ATAC-seq approaches will allow these GRNs to be studied further to help us understand how stem cells drive regeneration and homeostasis in planarians. In the future, both promoter-capture HiC and the co-accessibility of putative enhancers and promoters in scATAC-seq data would offer further computational possibilities to study these promoter–enhancer interactions (Schoenfelder et al., 2015; Pliner et al., 2018).

Taken together, our definition of enhancers in stem cells genome wide creates a foundation for constructing detailed GRNs to help understand regenerative mechanisms driven by stem cells in planarians.

Materials and methods

Reference assembly and annotations

The sexual genome SMESG.1 genome, the SMESG high-confidence annotation (SMESG-high release 1), and the SMESG annotation filtered for repeats (SMESG-repeat release 2) were downloaded from PlanMine (Brandl et al., 2016; Grohme et al., 2018). Available RNA-seq datasets were aligned to the genome with HISAT2 (version 2.1.0) using default parameter settings and providing known splice sites from the SMESG-high annotation (Kim et al., 2015). Transcripts were assembled and merged from the alignments with StringTie using the SMESG-high annotation as a reference (Pertea et al., 2015; Pertea et al., 2016).

The new expression-driven annotation (García-Castro et al., 2021, https://github.com/jakke-neiro/Oxplatys/raw/gh-pages/Schmidtea_mediterranea_Oxford_v1.gtf.zip) was compared to the SMESG-high annotation with gffcompare (Pertea et al., 2015; Pertea et al., 2016). Transcripts labeled with the class code "=" were classified as full matches corresponding to the transcripts in the SMESG.1 genome, while transcripts labeled with the class codes "c", "k", "m", "n", "j", "e", and "o" were classified as isoforms. Lastly, transcripts labeled with class codes "u" (unknown or intergenic), "i" (fully contained within a reference intron), and "x" (exonic overlap on the opposite strand) were selected as potential candidates for new coding and non-coding transcripts (Pertea et al., 2016; Wu et al., 2016, Azlan et al., 2019). Transdecoder was used to identify ORFs in these new transcripts, and transcripts longer than 100 amino acids were classified as putative coding transcripts (Haas et al., 2013; Wu et al., 2016, Azlan et al., 2019). Subsequently, InteProScan and BlastX against UniProt were used to look for protein-like structures in the translated ORFs of the remaining transcripts, and transcripts without any hits were retained as putative non-coding transcripts Jones et al., 2014, Wu et al., 2016.

Expression values, homology, and GO

To obtain FACS-specific expression values, the same FACS RNA-seq datasets used by Dattani et al., 2018 were used. The RNA-seq datasets were pseudo-aligned with Salmon using selective alignment, k-mer size 31 and 50 bootstrap iterations (Patro et al., 2017). The abundance estimates were converted to the Kallisto compatible format with wasabi, and the transcript counts were normalized with Sleuth (Pimentel et al., 2017). The TPM values of individual transcripts were summed to calculate TPM values for each gene. The mean TPM value across all samples in a FACS cell compartment was calculated for each gene to get the final absolute TPM value. The proportional expression values ($X_k\%$) were calculated by dividing the compartment-specific TPM value by the sum of TPM values of all compartments:

$$X_k\% = \frac{X_k(TPM)}{X_1(TPM) + X_2(TPM) + X_{ins}(TPM)}, X_k \in \{X_1, X_2, X_{ins}\}$$

Transcripts were categorized into FACS enrichment categories as follows: X_1 when $X_1\% > 0.5$, X_2 when $X_2\% > 0.5$, X_{ins} when $X_{ins}\% > 0.5$, X_1 and X_2 when $X_1\% + X_2\% > 0.75$ and neither enriched in X_1 nor X_2 , X_1 , and X_{ins} when $X_1\% + X_{ins}\% > 0.75$ and neither enriched in X_1 nor X_{ins} , X_2 , and X_{ins} when $X_2\% + X_{ins}\% > 0.75$ and neither enriched in X_2 nor X_{ins} , and ubiquitous when not enriched in any of the above categories (Dattani et al., 2018). Transcript lengths and expression values were compared with Kruskal–Wallis test. Ternary plots and kernel density estimation of the proportional expression values were generated with the package ggtern (Hamilton and ggtern, 2018). Information content (IC) was defined to represent as a singular scalar the divergence of a combination of proportional expression values from the even distribution point ($X_1\% = X_2\% = X_{ins}\%$):

$$IC = 0.01 * p_1 * \log(p_1) + 0.01 * p_2 * \log(p_2) + 0.01 * p_3 * \log(p_3)$$

where $p_1 = X_1\%$, $p_2 = X_2\%$, $p_3 = X_{ins}\%$, and $\min(\{p_1, p_2, p_3\}) = 0.01$, meaning that proportional expression values below 0.01 were assigned the value 0.01.

The homology of the transcripts was investigated by using Blastn (*Altschul et al., 1990*). The non-coding and coding transcripts were aligned to the transcriptomes of humans (GRCh38.p13) and *D. japonica* (assembled in *García-Castro et al., 2021*). The threshold for non-coding transcripts was $e\text{-value} = 10^{-5}$ and for coding transcripts $e\text{-value} = 10^{-10}$. GO enrichment analysis of new coding transcripts with respect to known coding transcripts was performed using topGO with Fisher's exact test (*Alexa and Rahnenfuhrer, 2020*). The data for proportional expression and homology assigned by BLAST of all transcripts is available at <https://jakke-neiro.github.io/Oxplatys/>; *Neiro, 2020*.

Identification and characterization of transcription factors

We conducted a systematic domain annotation of all transcripts by using the InterProScan resource (*Jones et al., 2014*). Transcripts with an InterProScan description 'transcription factor', a Pfam hit to the Pfam families listed in the TF database DNA-binding domain (DBD) v2.0, or a SUPERFAMILY hit listed in the SUPERFAMILY families in DBD were classified as TFs (*Wilson et al., 2008*). We used Blastn to align the potential planarian TFs to TFs in humans and fruit flies, and potential TFs without hits were filtered out. The planarian literature was systematically reviewed (*Supplementary file 1*), and if a TF was mentioned, the GenBank accession number or primer information was retrieved to establish the exact sequence in the literature. This sequence from the literature was aligned to the potential TFs (default parameters for GenBank accession entries and word size 10 for primers) to correctly assign the TFs in the literature to our transcripts (*Supplementary file 1*). If a planarian TF had previously been described in the literature under a certain name, one name was chosen as the primary name, while all other alternative synonyms used for the same TF were listed as secondary names (*Supplementary file 1*). If a potential TF had not been described in the planarian literature, the name of the best human or fruit fly Blast hit was used (*Supplementary file 1*). The TFs were categorized into four main groups (Basic domain, Zinc fingers, Helix-turn-helix, Alpha-helical, Immunoglobulin-fold, Beta-hairpin, Beta-sheet, Beta-barrel, Others) according to the TRANSFAC database (*Stegmaier et al., 2004; Supplementary file 1*).

Ternary plots, kernel density estimation, and the information content of the proportional expression values of TFs were calculated as for other transcripts (see 'Expression values, homology, and GO'). Hierarchical clustering of FACS proportions was performed using Euclidean distance and Ward's method with the package hclust (*van den Boogaart and Tolosana-Delgado, 2008*). The frequencies of different TF domains among the cellular enrichment classes X1, X2, and Xins (for which FACS proportional expression is higher or equal to 50%) were compared with the chi-square test using Yates correction.

Motifs of transcription factors

The motifs of the TFs (*Supplementary file 1*) were predicted by using the JASPAR profile inference tool (*Fornes et al., 2020*). The protein sequences were retrieved with Transdecoder (see 'Reference assembly and annotations'). The information content for each motif was determined by calculating the mean of Shannon's entropy at each position in the position weight matrix (PWM). The motifs were visualized with seqLogo (*Bembom, 2019*).

ATAC-seq library preparation

A standard protocol was used for preparing the ATAC-seq library (*Buenrostro et al., 2013*). The X1 cell compartments were isolated by FACS (*Romero et al., 2012*), and two replicates with 120,000–250,000 cells were collected from each compartment. The cells were washed and centrifuged (1 XPBS, 1200 RPM). Lysis buffer (10 mM Tri-Cl [pH 7.5], 10 mM NaCl, 3 mM MgCl₂, 0.1% NP-40) was added, the cells were centrifuged (500 RPM, 10 min, 4°C), and the nuclei pellet was collected. Then, the transposase mix (25 µl 2X TD Buffer, 2.5 µl Tn5 Transposase, 22.5 µl of nuclease-free water) was added, and the cells were resuspended and incubated (37°C, 60 min). Finally, DNA was isolated using the Zymogen Clean & Concentrator Kit, and eluted in EB buffer.

Subsequently, the eluted DNA was used for PCR amplification and library purification. DNA was amplified using a standard reaction (10 µl of purified transposed DNA, 10 µl of nuclease-free water, 15 µl Nextera PCR Master Mix, 5 µl of PCR primer cocktail, 5 µl Index Primer 1, 5 µl Index Primer 2, 72°C, 3 min, 98°C, 30 s, 14 cycles of 98°C, 10 s, 63°C, 30 s, 72°C for 1 min). Finally, the libraries were cleaned with the AMPure bead purification kit. Samples were paired-end sequenced on the

Illumina NextSeq. The sample replicates can be found on Sequence Read Archive (SRR18923214 and SRR18923213).

ChIP-seq library preparation and sequencing

The H3K27ac ChIP-seq library was prepared and sequenced using established protocols (*Mihaylova et al., 2018; Dattani et al., 2018*). A total of 600,000–700,000 planarian X1 cells were isolated for each experimental replicate (actual sample and input control). Two sample replicates (Sequence Read Archive SRR18925505 and SRR18925504) and two input replicates (Sequence Read Archive SRR18925503 and SRR18925502) were prepared. The H3K27ac Abcam ab4729 antibody was used for immunoprecipitation.

ChIP-seq data analysis

The ChIP-seq data with respect to H3K4me1 and H3K4me1 after *lpt(RNAi)* was reanalyzed (*Mihaylova et al., 2018*, PRJNA338116) alongside the H3K27ac data prepared here. The reads were quality-checked with FastQC and trimmed with Trimmomatic (*Andrews et al., 2010; Bolger et al., 2014*). The reads were aligned to the SMESG.1 genome with Bowtie2 (*Langmead et al., 2009*). Only uniquely mapped reads and reads with a quality score greater than 10 were retained with Samtools (*Li et al., 2009*). Peaks were called with MACS2 with default parameters for H3K4me1 and the broad peak parameter for H3K27ac (*Zhang et al., 2008*). Coverage tracks were generated with deepTools bamCoverage, and the mean coverage track was determined with wiggletools (*Zerbino et al., 2014; Ramírez et al., 2016*). Heatmaps of ChIP-seq profiles were generated with deepTools computeMatrix and plotHeatmap 1000 bp upstream and downstream of the peak center (*Ramírez et al., 2016*). The mean peak value was calculated as the mean of the ChIP-seq signal 1000 bp upstream and downstream of the peak center. The H3K27ac peaks that were at most 500 bp from the nearest H3K4me1 peak were selected as putative enhancers. Random genomic regions were generated with bedtools random (*Quinlan, 2014*). The difference of mean peak value at enhancers and random regions was tested with Wilcoxon's test. Detailed analysis and code are available at <https://jakke-neiro.github.io/Oxplatys/>; *Neiro, 2020*.

ATAC-seq data analysis and motif footprinting

The quality of paired-end reads was assessed with FastQC and adapter sequences were removed with Trimmomatic (*Andrews et al., 2010; Bolger et al., 2014*). The reads were aligned to the SMESG.1 genome with Bowtie2 (*Langmead et al., 2009*). Only uniquely mapped reads and reads with a quality score greater than 10 were retained with Samtools (*Li et al., 2009*). Duplicates were removed with Samtools (*Li et al., 2009*). Insert sizes were calculated with Picard CollectInsertSizeMetrics (<http://broadinstitute.github.io/picard/>; *Magner, 2022*).

For coverage track generation and peak calling, samples were downsampled to a standard of ~20M reads using Picard DownsampleSam (<http://broadinstitute.github.io/picard/>; *Magner, 2022*). DeepTools2 bamCoverage was used to generate coverage tracks for each sample in bigwig and bedgraph formats by normalizing with respect to reads million mapped reads (RPKM) (*Ramírez et al., 2016*). TOBIAS was used to look for footprints in the ATAC-seq by using the motifs predicted by JASPAR (see 'Motifs of transcription factors'; *Bentsen et al., 2020*). BAM files of mapped reads are available at <https://jakke-neiro.github.io/Oxplatys/>; *Neiro, 2020*.

Enhancer targets and gene regulatory networks

The putative enhancers were assigned to the nearest TSS with ChIPSeeker (*Yu et al., 2015*). The footprints in enhancers were used to create links between TFs and target genes for network construction using a Python script (available at <https://jakke-neiro.github.io/Oxplatys/>; *Neiro, 2020*). Networks were visualized with the R package Igraph (<https://github.com/igraph/igraph>; *Nepusz, 2022*) while the genomic regions were visualized with Gviz (*Hahne and Ivanek, 2016*).

Functional transcriptomics and validation of gene regulatory network predictions

To calculate the odds ratio of our approach, we defined the expected value of the number regulatory links for a given TF. A regulatory link for a given TF assumes that the target gene is linked to an

enhancer and that the enhancer contains a footprint for the given TF. In addition, if we assume that all differentially expressed genes are real direct target genes for the given TF, then the target gene of the regulatory link is also differentially expressed upon perturbation of TF activity. First, we define A as the event that a TF is differentially expressed. The probability of this event is

$$P(A) = \frac{n_{dge}}{n_{ge}}$$

where n_{ge} is the total number of TFs in the genome, and n_{dge} is the number of differentially expressed TFs. Second, we define B as the event that a given TF is linked to an enhancer. The probability of this event is

$$P(B) = \frac{n_{cis}}{n_{ge}}$$

where n_{cis} is the total number of TFs linked to an enhancer. Third, we define C as the event that an enhancer contains a footprint (bound motif) of the given TF. The probability of this event is

$$P(C) = \frac{n_{cisTF}}{n_{cis}}$$

where n_{cisTF} is the number of enhancer-linked genes containing the footprint of the given TF in their enhancers.

We assume that the events A , B , and C are independent. The expected number of differentially expressed genes linked to enhancers with the footprint of a given TF is equal to

$$P(A \cap B \cap C) n_{ge} = P(ABC) n_{ge} = P(A) P(B) P(C) n_{ge}$$

Code availability

All codes are available in a notebook provided as a zip file (**Source code 1**).

Additional information

Funding

Funder	Grant reference number	Author
Medical Research Council	MR/T028165/1	Aziz Aboobaker
Biotechnology and Biological Sciences Research Council	BB/J014427/1	Anish Dattani
Biotechnology and Biological Sciences Research Council	BB/M011224/1	Jakke Neuro

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Author contributions

Jakke Neuro, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing - original draft, Writing - review and editing; Divya Sridhar, Investigation, Methodology, Writing - review and editing; Anish Dattani, Formal analysis, Investigation, Visualization, Methodology, Writing - review and editing; Aziz Aboobaker, Conceptualization, Supervision, Funding acquisition, Investigation, Methodology, Writing - original draft, Project administration, Writing - review and editing

Author ORCIDs

Jakke Neuro  <http://orcid.org/0000-0002-5077-4958>
 Anish Dattani  <http://orcid.org/0000-0003-2211-1521>
 Aziz Aboobaker  <http://orcid.org/0000-0003-4902-5797>

Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.79675.sa1>

Author response <https://doi.org/10.7554/eLife.79675.sa2>

Additional files

Supplementary files

- Supplementary file 1. Full list of transcription factors (TFs) annotated in the *S. mediterranea* genome. TFs are named according to existing names in the literature or according to the currently agreed convention. The absolute (transcripts per million) or proportional expression in the fluorescence-activated cell sorting (FACS)-associated cell compartments X1, X2, and Xins is also recorded.
- Supplementary file 2. Full list of putative enhancers in adult stem cells defined by ChIP-seq and ATAC-seq analyses. Includes the position in the genome and the identify of the predicted target gene based on proximity to the predicted transcriptional start site.
- Supplementary file 3. Full list of predicted ATAC-seq footprints in the genome, including predicted transcription factor (TF)/TF family, position of binding site in the genome, and position of the peak/enhancer in the genome.
- Supplementary file 4. Full list of motifs with and without footprints organized by transcription factor (TF)/TF family in predicted enhancer regions.
- Supplementary file 5. Overview of number of enhancers, predicted bound motifs (footprints), and unbound motifs for fate-specific transcription factors (FSTF) expressed in adult stem cells.
- Supplementary file 6. Full list of transcription factors (TFs), involved in a putative stem cell gene regulatory network (GRN). Includes TF with expression >1/3 in the X1 compartment and includes other TFs implicated in regulating or being regulated by these TFs.
- Supplementary file 7. Differential expression analysis of transcription factor (TF) RNAi with RNA-seq datasets.
- Supplementary file 8. Differentially expressed transcription factor (TFs) after TF RNAi.
- MDAR checklist
- Source code 1. Full set of codes used for analyses in this study.

Data availability

Data and analyses are available at <https://jakke-neuro.github.io/Oxplatys>, (copy archived at [swh:1:rev:51a4d412daf02503691c903cd64a6f6e78dc1b25](https://swh.1:rev:51a4d412daf02503691c903cd64a6f6e78dc1b25)). All analysis code is provided in supplementary file New sequence data are available at the NCBI under Bioproject ID PRJNA832235, <https://www.ncbi.nlm.nih.gov/bioproject/832235>.

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Neiro J, Sridhar D, Dattani A, Aboobaker A	2022	Identification of enhancer-like elements defines regulatory networks active in planarian adult stem cells	https://www.ncbi.nlm.nih.gov/bioproject/832235	NCBI BioProject, PRJNA832235

References

- Abnave P**, Aboukhatwa E, Kosaka N, Thompson J, Hill MA, Aboobaker AA. 2017. Epithelial-mesenchymal transition transcription factors control pluripotent adult stem cell migration in vivo in planarians. *Development* **144**:3440–3453. DOI: <https://doi.org/10.1242/dev.154971>, PMID: 28893948
- Aboobaker AA**. 2011. Planarian stem cells: a simple paradigm for regeneration. *Trends in Cell Biology* **21**:304–311. DOI: <https://doi.org/10.1016/j.tcb.2011.01.005>, PMID: 21353778
- Adler CE**, Sánchez Alvarado A. 2015. Types or states? cellular dynamics and regenerative potential. *Trends in Cell Biology* **25**:687–696. DOI: <https://doi.org/10.1016/j.tcb.2015.07.008>, PMID: 26437587
- Albà MM**. 2017. Zinc-finger domains in metazoans: evolution gone wild. *Genome Biology* **18**:1–3. DOI: <https://doi.org/10.1186/s13059-017-1307-y>, PMID: 28877710
- Alexa A**, Rahnenfuhrer J. 2020. TopGO: enrichment analysis for gene ontology. R Package Version 2.40.0.
- Altschul SF**, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**:403–410. DOI: [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2), PMID: 2231712
- An Y**, Kawaguchi A, Zhao C, Toyoda A, Sharifi-Zarchi A, Mousavi SA, Bagherzadeh R, Inoue T, Ogino H, Fujiyama A, Chitsaz H, Baharvand H, Agata K. 2018. Draft genome of *Dugesia japonica* provides insights into conserved regulatory elements of the brain restriction gene *nou-darake* in planarians. *Zoological Letters* **4**:24. DOI: <https://doi.org/10.1186/s40851-018-0102-2>, PMID: 30181897

- Andrews S**, Krueger F, Seconds-Pichon A, Biggins F, Wingett S. 2010. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics.
- Arnold CD**, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. 2013. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**:1074–1077. DOI: <https://doi.org/10.1126/science.1232542>, PMID: 23328393
- Arnold CP**, Lozano AM, Mann FG, Nowotarski SH, Haug JO, Lange JJ, Seidel CW, Alvarado AS. 2021. Hox genes regulate asexual reproductive behavior and tissue segmentation in adult animals. *Nature Communications* **12**:6706. DOI: <https://doi.org/10.1038/s41467-021-26986-2>, PMID: 34795249
- Azlan A**, Obeidat SM, Yunus MA, Azzam G. 2019. Systematic identification and characterization of aedes aegypti long noncoding rnas (lncrnas). *Scientific Reports* **9**:1–9. DOI: <https://doi.org/10.1038/s41598-019-47506-9>, PMID: 31434910
- Baldini A**, Fulcoli FG, Illingworth E. 2017. Tbx1: transcriptional and developmental functions. *Current Topics in Developmental Biology* **122**:223–243. DOI: <https://doi.org/10.1016/bs.ctdb.2016.08.002>, PMID: 28057265
- Bembom O**. 2019. SeqLogo: sequence logos for DNA sequence alignments. R Package Version 1.52.0.
- Bentsen M**, Goymann P, Schultheis H, Klee K, Petrova A, Wiegandt R, Fust A, Preussner J, Kuenne C, Braun T, Kim J, Looso M. 2020. ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. *Nature Communications* **11**:1–11. DOI: <https://doi.org/10.1038/s41467-020-18035-1>, PMID: 32848148
- Bolger AM**, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**:2114–2120. DOI: <https://doi.org/10.1093/bioinformatics/btu170>, PMID: 24695404
- Bonn S**, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczyński B, Riddell A, Furlong EEM. 2012. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nature Genetics* **44**:148–156. DOI: <https://doi.org/10.1038/ng.1064>, PMID: 22231485
- Brandl H**, Moon H, Vila-Farré M, Liu SY, Henry I, Rink JC. 2016. PlanMine—a mineable resource of planarian biology and biodiversity. *Nucleic Acids Research* **44**:D764–D773. DOI: <https://doi.org/10.1093/nar/gkv1148>, PMID: 26578570
- Buenrostro JD**, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* **10**:1213–1218. DOI: <https://doi.org/10.1038/nmeth.2688>, PMID: 24097267
- Calo E**, Wysocka J. 2013. Modification of enhancer chromatin: what, how, and why? *Molecular Cell* **49**:825–837. DOI: <https://doi.org/10.1016/j.molcel.2013.01.038>, PMID: 23473601
- Cassandri M**, Smirnov A, Novelli F, Pitolli C, Agostini M, Malewicz M, Melino G, Raschella G. 2017. Zinc-finger proteins in health and disease. *Cell Death Discovery* **3**:17071. DOI: <https://doi.org/10.1038/cddiscovery.2017.71>, PMID: 29152378
- Cowles MW**, Brown DDR, Nisperos SV, Stanley BN, Pearson BJ, Zayas RM. 2013. Genome-wide analysis of the bh1h gene family in planarians identifies factors required for adult neurogenesis and neuronal regeneration. *Development* **140**:4691–4702. DOI: <https://doi.org/10.1242/dev.098616>, PMID: 24173799
- Cowles MW**, Omuro KC, Stanley BN, Quintanilla CG, Zayas RM. 2014. COE loss-of-function analysis reveals a genetic program underlying maintenance and regeneration of the nervous system in planarians. *PLOS Genetics* **10**:e1004746. DOI: <https://doi.org/10.1371/journal.pgen.1004746>, PMID: 25356635
- Creyghton MP**, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. 2010. Histone h3k27ac separates active from poised enhancers and predicts developmental state. *PNAS* **107**:21931–21936. DOI: <https://doi.org/10.1073/pnas.1016071107>, PMID: 21106759
- Currie KW**, Brown DDR, Zhu S, Xu C, Voisin V, Bader GD, Pearson BJ. 2016. HOX gene complement and expression in the planarian *Schmidtea mediterranea*. *EvoDevo* **7**:1–11. DOI: <https://doi.org/10.1186/s13227-016-0044-8>, PMID: 27034770
- Dattani A**, Kao D, Mihaylova Y, Abnave P, Hughes S, Lai A, Sahu S, Aboobaker AA. 2018. Epigenetic analyses of planarian stem cells demonstrate conservation of bivalent histone modifications in animal stem cells. *Genome Research* **28**:1543–1554. DOI: <https://doi.org/10.1101/gr.239848.118>, PMID: 30143598
- Dattani A**, Sridhar D, Aziz Aboobaker A. 2019. Planarian flatworms as a new model system for understanding the epigenetic regulation of stem cell pluripotency and differentiation. *Seminars in Cell & Developmental Biology* **87**:79–94. DOI: <https://doi.org/10.1016/j.semcdb.2018.04.007>, PMID: 29694837
- Daugherty AC**, Yeo RW, Buenrostro JD, Greenleaf WJ, Kundaje A, Brunet A. 2017. Chromatin accessibility dynamics reveal novel functional enhancers in *C. elegans*. *Genome Research* **27**:2096–2107. DOI: <https://doi.org/10.1101/gr.226233.117>, PMID: 29141961
- Duncan EM**, Chitsazan AD, Seidel CW, Sánchez Alvarado A. 2015. Set1 and MLL1/2 target distinct sets of functionally different genomic loci in vivo. *Cell Reports* **13**:2741–2755. DOI: <https://doi.org/10.1016/j.celrep.2015.11.059>, PMID: 26711341
- Duren Z**, Chen X, Jiang R, Wang Y, Wong WH. 2017. Modeling gene regulation from paired expression and chromatin accessibility data. *PNAS* **114**:E4914–E4923. DOI: <https://doi.org/10.1073/pnas.1704553114>
- Duren Z**, Chen X, Xin J, Wang Y, Wong WH. 2020. Time course regulatory analysis based on paired expression and chromatin accessibility data. *Genome Research* **30**:622–634. DOI: <https://doi.org/10.1101/gr.257063.119>, PMID: 32188700
- Emerson RO**, Thomas JH. 2011. Gypsy and the birth of the SCAN domain. *Journal of Virology* **85**:12043–12052. DOI: <https://doi.org/10.1128/JVI.00867-11>, PMID: 21865395

- Ernst J**, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**:43–49. DOI: <https://doi.org/10.1038/nature09906>, PMID: 21441907
- Felix DA**, Aboobaker AA. 2010. The TALE class homeobox gene *smad-prep* defines the anterior compartment for head regeneration. *PLOS Genetics* **6**:e1000915. DOI: <https://doi.org/10.1371/journal.pgen.1000915>, PMID: 20422023
- Fincher CT**, Wurtzel O, de Hoog T, Kravarik KM, Reddien PW. 2018. Cell type transcriptome atlas for the planarian *Schmidtea mediterranea*. *Science* **360**:eaq1736. DOI: <https://doi.org/10.1126/science.aaq1736>, PMID: 29674431
- Fornes O**, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranašić D, Santana-Garcia W, Tan G, Chèneby J, Ballester B, Parcy F, Sandelin A, Lenhard B, Wasserman WW, Mathelier A. 2020. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research* **48**:D87–D92. DOI: <https://doi.org/10.1093/nar/gkz1001>, PMID: 31701148
- Forsthoefel DJ**, Cejda NI, Khan UW, Newmark PA. 2020. Cell-type diversity and regionalized gene expression in the planarian intestine. *eLife* **9**:e52613. DOI: <https://doi.org/10.7554/eLife.52613>, PMID: 32240093
- Fraguas S**, Cárcel S, Vivancos C, Molina MD, Ginés J, Mazariegos J, Sekaran T, Bartscherer K, Romero R, Cebrià F. 2021. CREB-binding protein (CBP) gene family regulates planarian survival and stem cell differentiation. *Developmental Biology* **476**:53–67. DOI: <https://doi.org/10.1016/j.ydbio.2021.02.008>, PMID: 33774010
- Fujimura N**, Vacik T, Machon O, Vlcek C, Scalabrin S, Speth M, Diep D, Krauss S, Kozmik Z. 2007. Wnt-mediated down-regulation of *sp1* target genes by a transcriptional repressor *sp5*. *The Journal of Biological Chemistry* **282**:1225–1237. DOI: <https://doi.org/10.1074/jbc.M605851200>, PMID: 17090534
- Gaiti F**, Calcino AD, Tanurdžić M, Degnan BM. 2017a. Origin and evolution of the metazoan non-coding regulatory genome. *Developmental Biology* **427**:193–202. DOI: <https://doi.org/10.1016/j.ydbio.2016.11.013>, PMID: 27880868
- Gaiti F**, Jindrich K, Fernandez-Valverde SL, Roper KE, Degnan BM, Tanurdžić M. 2017b. Landscape of histone modifications in a sponge reveals the origin of animal cis-regulatory complexity. *eLife* **6**:e22194. DOI: <https://doi.org/10.7554/eLife.22194>, PMID: 28395144
- Gao T**, Qian J. 2020a. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. *Nucleic Acids Research* **48**:D58–D64. DOI: <https://doi.org/10.1093/nar/gkz980>, PMID: 31740966
- Gao B**, Wang Y, Diaby M, Zong W, Shen D, Wang S, Chen C, Wang X, Song C. 2020b. Evolution of pogo, a separate superfamily of IS630-tc1-mariner transposons, revealing recurrent domestication events in vertebrates. *Mobile DNA* **11**:1–15. DOI: <https://doi.org/10.1186/s13100-020-00220-0>, PMID: 32742312
- García-Castro H**, Kenny NJ, Iglesias M, Álvarez-Campos P, Mason V, Elek A, Schönauer A, Sleight VA, Neiro J, Aboobaker A, Permanyer J, Irimia M, Sebé-Pedrós A, Solana J. 2021. ACME dissociation: a versatile cell fixation-dissociation method for single-cell transcriptomics. *Genome Biology* **22**:1–34. DOI: <https://doi.org/10.1186/s13059-021-02302-5>, PMID: 33827654
- Gehrke AR**, Neverett E, Luo YJ, Brandt A, Ricci L, Hulett RE, Gompers A, Ruby JG, Rokhsar DS, Reddien PW, Srivastava M. 2019. Acoel genome reveals the regulatory landscape of whole-body regeneration. *Science* **363**:eaau6173. DOI: <https://doi.org/10.1126/science.aau6173>, PMID: 30872491
- Goldman JA**, Kuzu G, Lee N, Karasik J, Gemberling M, Foglia MJ, Karra R, Dickson AL, Sun F, Tolstorukov MY, Poss KD. 2017. Resolving heart regeneration by replacement histone profiling. *Developmental Cell* **40**:392–404. DOI: <https://doi.org/10.1016/j.devcel.2017.01.013>, PMID: 28245924
- Gorkin DU**, Lee D, Reed X, Fletez-Brant C, Bessling SL, Loftus SK, Beer MA, Pavan WJ, McCallion AS. 2012. Integration of chip-seq and machine learning reveals enhancers and a predictive regulatory sequence vocabulary in melanocytes. *Genome Research* **22**:2290–2301. DOI: <https://doi.org/10.1101/gr.139360.112>, PMID: 23019145
- Grohme MA**, Schloissnig S, Rozanski A, Pippel M, Young GR, Winkler S, Brandl H, Henry I, Dahl A, Powell S, Hiller M, Myers E, Rink JC. 2018. The genome of *Schmidtea mediterranea* and the evolution of core cellular mechanisms. *Nature* **554**:56–61. DOI: <https://doi.org/10.1038/nature25473>, PMID: 29364871
- Haas BJ**, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. *Nature Protocols* **8**:1494–1512. DOI: <https://doi.org/10.1038/nprot.2013.084>, PMID: 23845962
- Hahne F**, Ivanek R. 2016. Visualizing genomic data using gviz and bioconductor. *Methods in Molecular Biology* **1418**:335–351. DOI: https://doi.org/10.1007/978-1-4939-3578-9_16, PMID: 27008022
- Hamilton NE**, ggtern FM. 2018. Ternary diagrams using ggplot2. *Journal of Statistical Software* **87**:1–7. DOI: <https://doi.org/10.18637/jss.v087.c03>
- Harris RE**, Setiawan L, Saul J, Hariharan IK. 2016. Localized epigenetic silencing of a damage-activated Wnt enhancer limits regeneration in mature *Drosophila* imaginal discs. *eLife* **5**:e11588. DOI: <https://doi.org/10.7554/eLife.11588>, PMID: 26840050
- Hayashi T**, Asami M, Higuchi S, Shibata N, Agata K. 2006. Isolation of planarian X-ray-sensitive stem cells by fluorescence-activated cell sorting. *Development, Growth & Differentiation* **48**:371–380. DOI: <https://doi.org/10.1111/j.1440-169X.2006.00876.x>, PMID: 16872450

- Heintzman ND**, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics* **39**:311–318. DOI: <https://doi.org/10.1038/ng1966>, PMID: 17277777
- Hoff KJ**, Stanke M. 2013. WebAUGUSTUS—a web service for training AUGUSTUS and predicting genes in eukaryotes. *Nucleic Acids Research* **41**:W123–W128. DOI: <https://doi.org/10.1093/nar/gkt418>, PMID: 23700307
- Hoff KJ**, Stanke M. 2019. Predicting genes in single genomes with augustus. *Current Protocols in Bioinformatics* **65**:e57. DOI: <https://doi.org/10.1002/cpbi.57>, PMID: 30466165
- Hopitzan AA**, Baines AJ, Ludosky MA, Recouvreur M, Kordeli E. 2005. Ankyrin-G in skeletal muscle: tissue-specific alternative splicing contributes to the complexity of the sarcolemmal cytoskeleton. *Experimental Cell Research* **309**:86–98. DOI: <https://doi.org/10.1016/j.yexcr.2005.04.013>, PMID: 15953600
- Jaber-Hijazi F**, Lo PJKP, Mihaylova Y, Foster JM, Benner JS, Tejada Romero B, Chen C, Malla S, Solana J, Ruzov A, Aziz Aboobaker A. 2013. Planarian MBD2/3 is required for adult stem cell pluripotency independently of DNA methylation. *Developmental Biology* **384**:141–153. DOI: <https://doi.org/10.1016/j.ydbio.2013.09.020>, PMID: 24063805
- Jänes J**, Dong Y, Schoof M, Serizay J, Appert A, Cerrato C, Woodbury C, Chen R, Gemma C, Huang N, Kissiov D, Stempor P, Steward A, Zeiser E, Sauer S, Ahringer J. 2018. Chromatin accessibility dynamics across *C. elegans* development and ageing. *eLife* **7**:e37344. DOI: <https://doi.org/10.7554/eLife.37344>, PMID: 30362940
- Janssens J**, Aibar S, Taskiran II, Ismail JN, Gomez AE, Aughey G, Spanier KI, De Rop FV, González-Blas CB, Dionne M, Grimes K, Quan XJ, Papasokrati D, Hulselmans G, Makhzami S, De Waegeneer M, Christiaens V, Southall T, Aerts S. 2022. Decoding gene regulation in the fly brain. *Nature* **601**:630–636. DOI: <https://doi.org/10.1038/s41586-021-04262-z>
- Jegla T**, Nguyen MM, Feng C, Goetschius DJ, Luna E, van Rossum DB, Kamel B, Pisupati A, Milner ES, Rolls MM. 2016. Bilaterian giant ankyrins have a common evolutionary origin and play a conserved role in patterning the axon initial segment. *PLOS Genetics* **12**:e1006457. DOI: <https://doi.org/10.1371/journal.pgen.1006457>, PMID: 27911898
- Jones P**, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**:1236–1240. DOI: <https://doi.org/10.1093/bioinformatics/btu031>, PMID: 24451626
- Kang J**, Hu J, Karra R, Dickson AL, Tornini VA, Nachtrab G, Gemberling M, Goldman JA, Black BL, Poss KD. 2016. Modulation of tissue repair by regeneration enhancer elements. *Nature* **532**:201–206. DOI: <https://doi.org/10.1038/nature17644>, PMID: 27049946
- Khan SF**, Damerell V, Omar R, Du Toit M, Khan M, Maranyane HM, Mlaza M, Bleloch J, Bellis C, Sahn BDB, Peres J, ArulJothi KN, Prince S. 2020. The roles and regulation of TBX3 in development and disease. *Gene* **726**:144223. DOI: <https://doi.org/10.1016/j.gene.2019.144223>, PMID: 31669645
- Kim D**, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* **12**:357–360. DOI: <https://doi.org/10.1038/nmeth.3317>, PMID: 25751142
- Klemm SL**, Shipony Z, Greenleaf WJ. 2019. Chromatin accessibility and the regulatory epigenome. *Nature Reviews. Genetics* **20**:207–220. DOI: <https://doi.org/10.1038/s41576-018-0089-8>, PMID: 30675018
- Koenecke N**, Johnston J, He Q, Meier S, Zeitlinger J. 2017. *Drosophila* poised enhancers are generated during tissue patterning with the help of repression. *Genome Research* **27**:64–74. DOI: <https://doi.org/10.1101/gr.209486.116>
- Labbé RM**, Irimia M, Currie KW, Lin A, Zhu SJ, Brown DDR, Ross EJ, Voisin V, Bader GD, Blencowe BJ, Pearson BJ. 2012. A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells* **30**:1734–1745. DOI: <https://doi.org/10.1002/stem.1144>, PMID: 22696458
- Langmead B**, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**:R25. DOI: <https://doi.org/10.1186/gb-2009-10-3-r25>, PMID: 19261174
- Li H**, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and samtools. *Bioinformatics* **25**:2078–2079. DOI: <https://doi.org/10.1093/bioinformatics/btp352>, PMID: 19505943
- Li Z**, Schulz MH, Look T, Begemann M, Zenke M, Costa IG. 2019. Identification of transcription factor binding sites using ATAC-seq. *Genome Biology* **20**:45. DOI: <https://doi.org/10.1186/s13059-019-1642-2>, PMID: 30808370
- Lopez AY**, Wang X, Xu M, Maheshwari A, Curry D, Lam S, Adesina AM, Noebels JL, Sun Q-Q, Cooper EC. 2017. Ankyrin-G isoform imbalance and interneuronopathy link epilepsy and bipolar disorder. *Molecular Psychiatry* **22**:1464–1472. DOI: <https://doi.org/10.1038/mp.2016.233>, PMID: 27956739
- Lowe EK**, Cuomo C, Voronov D, Arnone MI. 2019. Using ATAC-seq and RNA-seq to increase resolution in GRN connectivity. *Methods in Cell Biology* **151**:115–126. DOI: <https://doi.org/10.1016/bs.mcb.2018.11.001>, PMID: 30948003
- Magner R**. 2022. Picard. swb:1:rev:fc87aed37f657bc844c7271a91b190ac1a77a9ac. Software Heritage. <https://archive.softwareheritage.org/swb:1:dir:c4a976f952df723d9f815ff16d8f6de52d45a1ce;origin=https://github.com/broadinstitute/picard;visit=swb:1:snp:4ab926082574490562e2205c801f7e9ec2260463;anchor=swb:1:rev:fc87aed37f657bc844c7271a91b190ac1a77a9ac>

- Mihaylova Y**, Abnave P, Kao D, Hughes S, Lai A, Jaber-Hijazi F, Kosaka N, Aboobaker AA. 2018. Conservation of epigenetic regulation by the MLL3/4 tumour suppressor in planarian pluripotent stem cells. *Nature Communications* **9**:1–7. DOI: <https://doi.org/10.1038/s41467-018-06092-6>, PMID: 30194301
- Miraldi ER**, Pokrovskii M, Watters A, Castro DM, De Veaux N, Hall JA, Lee J-Y, Ciofani M, Madar A, Carriero N, Littman DR, Bonneau R. 2019. Leveraging chromatin accessibility for transcriptional regulatory network inference in T helper 17 cells. *Genome Research* **29**:449–463. DOI: <https://doi.org/10.1101/gr.238253.118>, PMID: 30696696
- Molina MD**, Saló E, Cebrià F. 2007. The BMP pathway is essential for re-specification and maintenance of the dorsoventral axis in regenerating and intact planarians. *Developmental Biology* **311**:79–94. DOI: <https://doi.org/10.1016/j.ydbio.2007.08.019>, PMID: 17905225
- Murad R**, Macias-Muñoz A, Wong A, Ma X, Mortazavi A. 2019. Integrative Analysis of Hydra Head Regeneration Reveals Activation of Distal Enhancer-like Elements. *bioRxiv*. DOI: <https://doi.org/10.1101/544049>
- Najafabadi HS**, Garton M, Weirauch MT, Mnaimneh S, Yang A, Kim PM, Hughes TR. 2017. Non-base-contacting residues enable kaleidoscopic evolution of metazoan C2H2 zinc finger DNA binding. *Genome Biology* **18**:1–15. DOI: <https://doi.org/10.1186/s13059-017-1287-y>, PMID: 28877740
- Neiro J**. 2020. Oxplatys. swh:1:rev:51a4d412daf02503691c903cd64a6f6e78dc1b25. Software Heritage. <https://archive.softwareheritage.org/swh:1:dir:d0b31ae353719430dbcd0785d1a62a2ce5297a6d;origin=https://github.com/jakke-neiro/Oxplatys;visit=swh:1:snp:c490a9c5c87bc69eb23a62fd05f80b6d5b201a51;anchor=swh:1:rev:51a4d412daf02503691c903cd64a6f6e78dc1b25>
- Nelson AD**, Jenkins PM. 2017. Axonal membranes and their domains: assembly and function of the axon initial segment and node of ranvier. *Frontiers in Cellular Neuroscience* **11**:136. DOI: <https://doi.org/10.3389/fncel.2017.00136>, PMID: 28536506
- Nepusz T**. 2022. Rigraph. swh:1:rev:eda5db6952d19d4043522ff1daf82ccff8e71d84. Software Heritage. <https://archive.softwareheritage.org/swh:1:dir:88c2167c0a67ce01671a7b4783a4e7282163d3e1;origin=https://github.com/igraph/igraph;visit=swh:1:snp:d5e6cddde440f706979d912e610cebecf1a09ea99;anchor=swh:1:rev:eda5db6952d19d4043522ff1daf82ccff8e71d84>
- Onal P**, Grün D, Adamidi C, Rybak A, Solana J, Mastrobuoni G, Wang Y, Rahn H-P, Chen W, Kempa S, Ziebold U, Rajewsky N. 2012. Gene expression of pluripotency determinants is conserved between mammalian and planarian stem cells. *The EMBO Journal* **31**:2755–2769. DOI: <https://doi.org/10.1038/emboj.2012.110>, PMID: 22543868
- Pascual-Carreras E**, Marín-Barba M, Castillo-Lara S, Coronel-Córdoba P, Magri MS, Wheeler GN, Abril JF, Gomez-Skarmeta JL, Saló E, Adell T. 2020. Genomic Analyses Reveal FoxG as an Upstream Regulator of Wnt1 Required for Posterior Identity Specification in Planarians. *bioRxiv*. DOI: <https://doi.org/10.1101/2020.12.08.416008>
- Patro R**, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**:417–419. DOI: <https://doi.org/10.1038/nmeth.4197>, PMID: 28263959
- Pertea M**, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* **33**:290–295. DOI: <https://doi.org/10.1038/nbt.3122>, PMID: 25690850
- Pertea M**, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, stringtie and ballgown. *Nature Protocols* **11**:1650–1667. DOI: <https://doi.org/10.1038/nprot.2016.095>, PMID: 27560171
- Pimentel H**, Bray NL, Puente S, Melsted P, Pachter L. 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods* **14**:687. DOI: <https://doi.org/10.1038/nmeth.4324>
- Plass M**, Solana J, Wolf FA, Ayoub S, Misios A, Glažar P, Obermayer B, Theis FJ, Kocks C, Rajewsky N. 2018. Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. *Science* **360**:eaq1723. DOI: <https://doi.org/10.1126/science.aaq1723>, PMID: 29674432
- Pliner HA**, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, Srivatsan S, Qiu X, Jackson D, Minkina A, Adey AC, Steemers FJ, Shendure J, Trapnell C. 2018. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. *Molecular Cell* **71**:858–871. DOI: <https://doi.org/10.1016/j.molcel.2018.06.044>, PMID: 30078726
- Pundhir S**, Bagger FO, Lauridsen FB, Rapin N, Porse BT. 2016. Peak-valley-peak pattern of histone modifications delineates active regulatory elements and their directionality. *Nucleic Acids Research* **44**:4037–4051. DOI: <https://doi.org/10.1093/nar/gkw250>, PMID: 27095194
- Quinlan AR**. 2014. BEDTools: the swiss-army tool for genome feature analysis. *Current Protocols in Bioinformatics* **47**:11–12. DOI: <https://doi.org/10.1002/0471250953.bi1112s47>
- Rada-Iglesias A**, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**:279–283. DOI: <https://doi.org/10.1038/nature09692>, PMID: 21160473
- Ramírez F**, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. DeepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**:W160–W165. DOI: <https://doi.org/10.1093/nar/gkw257>, PMID: 27079975
- Raz AA**, Wurtzel O, Reddien PW. 2021. Planarian stem cells specify fate yet retain potency during the cell cycle. *Cell Stem Cell* **28**:1307–1322. DOI: <https://doi.org/10.1016/j.stem.2021.03.021>, PMID: 33882291

- Reddien PW**, Bermange AL, Kicza AM, Sánchez Alvarado A. 2007. BMP signaling regulates the dorsal planarian midline and is needed for asymmetric regeneration. *Development* **134**:4043–4051. DOI: <https://doi.org/10.1242/dev.007138>, PMID: 17942485
- Reddien PW**. 2018. The cellular and molecular basis for planarian regeneration. *Cell* **175**:327–345. DOI: <https://doi.org/10.1016/j.cell.2018.09.021>, PMID: 30290140
- Rink JC**, Gurley KA, Elliott SA, Sánchez Alvarado A. 2009. Planarian HH signaling regulates regeneration polarity and links HH pathway evolution to cilia. *Science* **326**:1406–1410. DOI: <https://doi.org/10.1126/science.1178712>, PMID: 19933103
- Rink JC**. 2013. Stem cell systems and regeneration in planaria. *Development Genes and Evolution* **223**:67–84. DOI: <https://doi.org/10.1007/s00427-012-0426-4>, PMID: 23138344
- Romero BT**, Evans DJ, Aboobaker AA. 2012. FACS analysis of the planarian stem cell compartment as a tool to understand regenerative mechanisms. *Progenitor Cells* **916**:167–179. DOI: https://doi.org/10.1007/978-1-61779-980-8_13
- Schoenfelder S**, Sugar R, Dimond A, Javierre B-M, Armstrong H, Mifsud B, Dimitrova E, Matheson L, Tavares-Cadete F, Furlan-Magaril M, Segonds-Pichon A, Jurkowski W, Wingett SW, Tabbada K, Andrews S, Herman B, LeProust E, Osborne CS, Koseki H, Fraser P, et al. 2015. Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nature Genetics* **47**:1179–1186. DOI: <https://doi.org/10.1038/ng.3393>, PMID: 26323060
- Schwaiger M**, Schönauer A, Rendeiro AF, Pribitzer C, Schauer A, Gilles AF, Schinko JB, Renfer E, Fredman D, Technau U. 2014. Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Research* **24**:639–650. DOI: <https://doi.org/10.1101/gr.162529.113>, PMID: 24642862
- Scimone ML**, Kravarik KM, Lapan SW, Reddien PW. 2014. Neoblast specialization in regeneration of the planarian *Schmidtea mediterranea*. *Stem Cell Reports* **3**:339–352. DOI: <https://doi.org/10.1016/j.stemcr.2014.06.001>, PMID: 25254346
- Scimone ML**, Cote LE, Reddien PW. 2017. Orthogonal muscle fibres have different instructive roles in planarian regeneration. *Nature* **551**:623–628. DOI: <https://doi.org/10.1038/nature24660>, PMID: 29168507
- Sebé-Pedrós A**, Ballaré C, Parra-Acero H, Chiva C, Tena JJ, Sabidó E, Gómez-Skarmeta JL, Di Croce L, Ruiz-Trillo I. 2016. The dynamic regulatory genome of capsaspora and the origin of animal multicellularity. *Cell* **165**:1224–1237. DOI: <https://doi.org/10.1016/j.cell.2016.03.034>, PMID: 27114036
- Shlyueva D**, Stampfel G, Stark A. 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nature Reviews. Genetics* **15**:272–286. DOI: <https://doi.org/10.1038/nrg3682>, PMID: 24614317
- Siebert S**, Farrell JA, Cazet JF, Abeykoon Y, Primack AS, Schnitzler CE, Juliano CE. 2019. Stem cell differentiation trajectories in *hydra* resolved at single-cell resolution. *Science* **365**:eaav9314. DOI: <https://doi.org/10.1126/science.aav9314>, PMID: 31346039
- Solana J**, Kao D, Mihaylova Y, Jaber-Hijazi F, Malla S, Wilson R, Aboobaker A. 2012. Defining the molecular profile of planarian pluripotent stem cells using a combinatorial rnaseq, RNA interference and irradiation approach. *Genome Biology* **13**:R19. DOI: <https://doi.org/10.1186/gb-2012-13-3-r19>, PMID: 22439894
- Spicuglia S**, Vanhille L. 2012. Chromatin signatures of active enhancers. *Nucleus* **3**:126–131. DOI: <https://doi.org/10.4161/nucl.19232>, PMID: 22555596
- Stegmaier P**, Kel AE, Wingender E. 2004. Systematic DNA-binding domain classification of transcription factors. *Genome Informatics* **15**:276–286. PMID: 15706513.
- Stelman CR**, Smith BM, Chandra B, Roberts-Galbraith RH. 2021. CBP/p300 homologs CBP2 and CBP3 play distinct roles in planarian stem cell function. *Developmental Biology* **473**:130–143. DOI: <https://doi.org/10.1016/j.ydbio.2021.02.004>, PMID: 33607113
- Stower's Institute**. 2016. RNA seq analysis in flatworm following 6, 12, or 18 day treatment with rna against pax5, sox8, or zfp1, relative to unc22 to identify transcriptional targets. Stower's Institute.
- Swapna LS**, Molinaro AM, Lindsay-Mosher N, Pearson BJ, Parkinson J. 2018. Comparative transcriptomic analyses and single-cell RNA sequencing of the freshwater planarian *Schmidtea mediterranea* identify major cell types and pathway conservation. *Genome Biology* **19**:1–22. DOI: <https://doi.org/10.1186/s13059-018-1498-x>
- Takahashi K**, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**:663–676. DOI: <https://doi.org/10.1016/j.cell.2006.07.024>, PMID: 16904174
- Takahashi K**, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S. 2007. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**:861–872. DOI: <https://doi.org/10.1016/j.cell.2007.11.019>, PMID: 18035408
- Tewari AG**, Owen JH, Petersen CP, Wagner DE, Reddien PW. 2019. A small set of conserved genes, including sp5 and hox, are activated by wnt signaling in the posterior of planarians and acoels. *PLOS Genetics* **15**:e1008401. DOI: <https://doi.org/10.1371/journal.pgen.1008401>, PMID: 31626630
- Thomas S**, Li X-Y, Sabo PJ, Sandstrom R, Thurman RE, Canfield TK, Giste E, Fisher W, Hammonds A, Celniker SE, Biggin MD, Stamatoyannopoulos JA. 2011. Dynamic reprogramming of chromatin accessibility during *Drosophila* embryo development. *Genome Biology* **12**:R43. DOI: <https://doi.org/10.1186/gb-2011-12-5-r43>, PMID: 21569360
- Tomoyasu Y**, Halfon MS. 2020. How to study enhancers in non-traditional insect models. *The Journal of Experimental Biology* **223**:jeb212241. DOI: <https://doi.org/10.1242/jeb.212241>, PMID: 32034049
- van den Boogaart KG**, Tolosana-Delgado R. 2008. "Compositions": a unified R package to analyze compositional data. *Comput Geosci* **34**:320–338. DOI: <https://doi.org/10.1016/j.cageo.2006.11.017>

- van Wolfswinkel JC, Wagner DE, Reddien PW. 2014. Single-cell analysis reveals functionally distinct classes within the planarian stem cell compartment. *Cell Stem Cell* **15**:326–339. DOI: <https://doi.org/10.1016/j.stem.2014.06.007>, PMID: 25017721
- Vásquez-Doorman C, Petersen CP. 2014. Zic-1 expression in planarian neoblasts after injury controls anterior pole regeneration. *PLOS Genetics* **10**:e1004452. DOI: <https://doi.org/10.1371/journal.pgen.1004452>, PMID: 24992682
- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**:854–858. DOI: <https://doi.org/10.1038/nature07730>, PMID: 19212405
- Vogg MC, Owlarn S, Pérez Rico YA, Xie J, Suzuki Y, Gentile L, Wu W, Bartscherer K. 2014. Stem cell-dependent formation of a functional anterior regeneration pole in planarians requires zic and forkhead transcription factors. *Developmental Biology* **390**:136–148. DOI: <https://doi.org/10.1016/j.ydbio.2014.03.016>, PMID: 24704339
- Wagner DE, Wang IE, Reddien PW. 2011. Clonogenic neoblasts are pluripotent adult stem cells that underlie planarian regeneration. *Science* **332**:811–816. DOI: <https://doi.org/10.1126/science.1203983>, PMID: 21566185
- Wang IE, Lapan SW, Scimone ML, Clandinin TR, Reddien PW. 2016. Hedgehog signaling regulates gene expression in planarian glia. *eLife* **5**:e16996. DOI: <https://doi.org/10.7554/eLife.16996>, PMID: 27612382
- Wang W, Hu C-K, Zeng A, Alegre D, Hu D, Gotting K, Ortega Granillo A, Wang Y, Robb S, Schnittker R, Zhang S, Alegre D, Li H, Ross E, Zhang N, Brunet A, Sánchez Alvarado A. 2020. Changes in regeneration-responsive enhancers shape regenerative capacities in vertebrates. *Science* **369**:eaaz3090. DOI: <https://doi.org/10.1126/science.aaz3090>, PMID: 32883834
- Wang LH, Aberin MAE, Wu S, Wang SP. 2021. The mll3/4 h3k4 methyltransferase complex in establishing an active enhancer landscape. *Biochemical Society Transactions* **49**:1041–1054. DOI: <https://doi.org/10.1042/BST20191164>, PMID: 34156443
- Weidinger G, Thorpe CJ, Wuennenberg-Stapleton K, Ngai J, Moon RT. 2005. The sp1-related transcription factors sp5 and sp5-like act downstream of wnt/beta-catenin signaling in mesoderm and neuroectoderm patterning. *Current Biology* **15**:489–500. DOI: <https://doi.org/10.1016/j.cub.2005.01.041>, PMID: 15797017
- West JA, Cook A, Alver BH, Stadtfeld M, Deaton AM, Hochedlinger K, Park PJ, Tolstorukov MY, Kingston RE. 2014. Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. *Nature Communications* **5**:1–2. DOI: <https://doi.org/10.1038/ncomms5719>
- Wilson D, Charoensawan V, Kummerfeld SK, Teichmann SA. 2008. DBD—taxonomically broad transcription factor predictions: new content and functionality. *Nucleic Acids Research* **36**:D88–D92. DOI: <https://doi.org/10.1093/nar/gkm964>
- Wu Y, Cheng T, Liu C, Liu D, Zhang Q, Long R, Zhao P, Xia Q, Ling E. 2016. Systematic identification and characterization of long non-coding rnas in the silkworm, *bombyx mori*. *PLOS ONE* **11**:e0147147. DOI: <https://doi.org/10.1371/journal.pone.0147147>
- Yan F, Powell DR, Curtis DJ, Wong NC. 2020. From reads to insight: a hitchhiker’s guide to ATAC-seq data analysis. *Genome Biology* **21**:22. DOI: <https://doi.org/10.1186/s13059-020-1929-3>, PMID: 32014034
- Yang K, Kang J. 2019. Tissue regeneration enhancer elements: a way to unlock endogenous healing power. *Developmental Dynamics* **248**:34–42. DOI: <https://doi.org/10.1002/dvdy.24676>, PMID: 30291668
- Yazawa S, Umesono Y, Hayashi T, Tarui H, Agata K. 2009. Planarian hedgehog/patched establishes anterior-posterior polarity by regulating Wnt signaling. *PNAS* **106**:22329–22334. DOI: <https://doi.org/10.1073/pnas.0907464106>, PMID: 20018728
- Yu G, Wang LG, He QY. 2015. ChIPseeker: an R/bioconductor package for chip peak annotation, comparison and visualization. *Bioinformatics* **31**:2382–2383. DOI: <https://doi.org/10.1093/bioinformatics/btv145>, PMID: 25765347
- Zerbino DR, Johnson N, Juettemann T, Wilder SP, Flicek P. 2014. WiggleTools: parallel processing of large collections of genome-wide datasets for visualization and statistical analysis. *Bioinformatics* **30**:1008–1009. DOI: <https://doi.org/10.1093/bioinformatics/btt737>, PMID: 24363377
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of chip-seq (MACS). *Genome Biology* **9**:R137. DOI: <https://doi.org/10.1186/gb-2008-9-9-r137>, PMID: 18798982
- Zhu B, Zhang W, Zhang T, Liu B, Jiang J. 2015. Genome-wide prediction and validation of intergenic enhancers in arabidopsis using open chromatin signatures. *The Plant Cell* **27**:2415–2426. DOI: <https://doi.org/10.1105/tpc.15.00537>, PMID: 26373455
- Zhu SJ, Pearson BJ. 2016. (Neo)blast from the past: new insights into planarian stem cell lineages. *Current Opinion in Genetics & Development* **40**:74–80. DOI: <https://doi.org/10.1016/j.gde.2016.06.007>, PMID: 27379899