Epigenetic analyses of planarian stem cells demonstrate conservation of bivalent histone modifications in animal stem cells.

Anish Dattani*[1], Damian Kao*[1], Yuliana Mihaylova[1], Prasad Abnave[1], Samantha Hughes[1], Alvina Lai[1], Sounak Sahu[1] and Aziz Aboobaker[1**]

1. Department of Zoology, Tinbergen Building, South Parks Road, Oxford, OX13PS, UK

* authors contributed equally

** corresponding author: aziz.aboobaker@zoo.ox.ac.uk

Abstract

Planarian flatworms have an indefinite capacity to regenerate missing or damaged body parts owing to a population of pluripotent adult stems cells called neoblasts (NBs). Currently, little is known about the importance of the epigenetic status of NBs and how histone modifications regulate homeostasis and cellular differentiation. We have developed an improved and optimized ChIP-seq protocol for NBs in *Schmidtea mediterranea* and have generated genome-wide profiles for the active marks H3K4me3 and H3K36me3, and suppressive marks H3K4me1 and H3K27me3. The genome-wide profiles of these marks were found to correlate well with NB gene expression profiles. We found that genes with little transcriptional activity in the NB compartment but which switch on in post-mitotic progeny during differentiation are bivalent, being marked by both H3K4me3 and H3K27me3 at promoter regions. In further support of this hypothesis bivalent genes also have a high level of paused RNA Polymerase II at the promoter-proximal region. Overall, this study confirms that epigenetic control is important for the maintenance of a NB transcriptional program and makes a case for bivalent promoters as a conserved feature of animal stem cells and not a vertebrate specific innovation. By establishing a robust ChIP-seq protocol and analysis methodology, we further promote planarians as a promising model system to investigate histone modification mediated regulation of stem cell function and differentiation.

Introduction

The promoters of developmental genes in mammalian embryonic stem cells (ESCs) are frequently marked with both the silencing H3K27me3 mark and active H3K4me3 marks. It has been proposed that this 'bivalent' state precedes resolution into full transcriptional activation or repression depending on ultimate cell type commitment (Bernstein et al. 2006; Voigt et al. 2013; Harikumar and Meshorer 2015). The advantage is that bivalency represents a poised or transcription-ready state, whereby a developmental gene is silenced in ESCs, but can be readily rendered active during differentiation to a defined lineage. Evidence for this comes from the finding that 51% of bivalent promoters in ESCs are bound by paused polymerase (RNAPII-Ser5P), compared with 8% of non-bivalent promoters (Lesch and Page 2014; Brookes et al. 2012); demonstrating a strong but not complete association. Bivalency may also protect promoters against less reversible suppressive mechanisms, such as DNA methylation (Lesch and Page 2014). Bivalent chromatin has also been discovered in male and female germ cells at many of the gene promoters that regulate somatic development, and may underpin the gametes' ability to generate a zygote capable of producing all cellular lineages (Yamaguchi et al. 2013; Cui et al. 2009; Hattori et al. 2013; Sachs et al. 2013; Lesch et al. 2013; Lesch and Page 2014).

It remains unclear whether the poised bivalent promoters of developmental genes are an epigenetic signature of vertebrates or arose earlier in the ancestor of all animals. Recently, the orthologs of bivalent genes that sit at the top of transcriptional hierarchies in mammalian development, were also found to be poised in chicken

male germ cells (Lesch et al. 2016). Sequential ChIP has also established H3K4me3/H3K27me3 co-occupancy of promoters in zebrafish blastomeres (Vastenhouw et al. 2010). Conversely, comparatively few bivalent domains were identified in Xenopus embryos undergoing the midblastula transition (Akkers et al. 2009). Xenopus genes which appear to have signals for both H3K4me3 and H3K27me3 originate from cells in distinct areas of the embryo, and as such the observed bivalency can be explained by cellular heterogeneity (Akkers et al. 2009).

Given that bivalency correlates with pluripotency in ESCs, planarian pluripotent adult stem cells or neoblasts (NBs) represent one possible scenario where poised promoters could have an important role in invertebrates, if this regulatory feature is conserved. Planarian NBs are a population of adult dividing cells that collectively produce all differentiated cells during homeostatic turnover and regeneration (Aboobaker 2011; Rink 2013). Several RNA-binding proteins, such as *piwi* and *vasa*, typically associated with nuage of germ cells are also expressed in planarian NBs where they function in the maintenance of pluripotency (Reddien et al. 2005; Palakodeti et al. 2008; Solana 2013; Shibata et al. 2016; Lai and Aboobaker 2018). Moreover, the ability of NBs to differentiate upon demand must also require well-regulated transcriptional and epigenetic processes, and poised, bivalent promoters may constitute an effective way of coordinating the differentiation of these stem cells.

Here we develop an optimized ChIP-seq methodology for planarian NBs and combine this with informatics approaches to establish robust approaches for studying histone modifications at transcriptional starts sites (TSSs). This enabled us to identify genes with inactive/low expression in the NB population, but with greatly

increased expression in post-mitotic NB progeny that are actively differentiating, with bivalent promoters in planarian NBs by combining transcriptomic and epigenetic analyses. Our findings indicate that bivalent promoters in pluripotent stem cells are not just a facet of vertebrates, but may have a role in regulating pluripotency in embryonic and adult stem cells across animals.

Results

Genome-wide annotation of transcribed loci in asexual *Schmidtea mediterranea* genome and categorization by proportional expression in FACS populations

We sought to produce an annotation of all transcribed loci on the asexual *Schmidtea mediterranea* genome (SmedAsxl v1.1) utilising both *de novo* assembled transcriptomes and 164 independent RNA-seq datasets covering RNAi knockdown-, regenerating-, whole worm-, and cell compartment-specific datasets (**Figure 1A, Supplementary File 1**). The inclusion of these diverse datasets was to improve the overall representation of the genome, and is useful for discovering potential non-coding RNAs and protein-coding genes expressed at low levels, both of which may not have been fully covered by individual studies limited by read number, or reliant on homology based annotation processes such as MAKER (Cantarel et al. 2008).

Our new expression-based annotation identified 38,711 expressed loci, 21,772 of which are predicted to be coding (**Figure 1A**). Moreover, compared to the current available annotation of the *Schmidtea mediterranea* asexual genome (Smed GD 2.0)

(Robb et al. 2008), our annotation discovered 10,210 new potential protein coding loci that are expressed at similar overall levels to previously annotated protein coding genes. A total of 6,300 genes from the existing MAKER homology based annotation were not present in our expression driven annotation. Further analysis of these MAKER-specific genes shows that they generally have no or very little potential expression within the 164 RNA-seq libraries utilised for our annotation (**Supplementary Figure 1**).

In the absence of transgenic approaches and antibodies for confirmed cell lineage markers, Fluorescence Activated Cell Sorting (FACS) gating cell populations stained with Hoechst and calcein is the best available tool for isolating NBs, progeny, and differentiated cells (Hayashi et al. 2006; Romero et al. 2012). FACS allows for two irradiation sensitive compartments to be discerned: the ‘X1’ gate representative of S/G2/M-phase NBs with >2C DNA content; and the ‘X2’ gate representative of G1 phase NBs and post-mitotic progeny with 2C DNA content. The third FACS population, ‘Xins’, represents an irradiation-insensitive population with a higher cytoplasmic to nuclear ratio (**Figure 1B and 1C**). These cell compartments are heterogeneous, with subpopulations of NBs expressing epidermal, gut and other lineage-specific markers present within the X1 population (Scimone et al. 2014; Wurtzel et al. 2015; Van Wolfswinkel et al. 2014), and the X2 compartment consisting of an amalgam of G1 NBs and lineage-committed post-mitotic progeny (Baguñá and Romero 1981; Hayashi et al. 2006; Zhu et al. 2015; Molinaro and Pearson 2016).

We used the publicly available RNA-seq datasets for these three different FACS populations in order to compare the expression of our annotated loci in these three distinct compartments (Önal et al. 2012; Labbé et al. 2012; Van Wolfswinkel et al. 2014; Zhu et al. 2015; Duncan et al. 2015). We first looked at the normalized TPM expression levels for annotated loci in our annotated genome in the FACS population datasets originating from four different planarian labs **(Supplementary Figure 2).** This revealed a rough congruence between different FACS populations from different labs (**Supplementary Figure 3A**).

We transformed absolute TPM expression values into proportional values for each FACS compartment in each of the datasets (**Figure 1D; Supplementary Figure 3B**). These proportional values were then averaged across datasets, to produce a final set of X1:X2:Xins proportions for 27,206 loci (18,010 of which are predicted to be protein-coding) that had at least 10 reads mapped in at least one FACS RNA-seq library.

We were now able to sort all annotated genes according to whether their predominant expression (i.e. => 50% expression) is in X1 (S/G2/M-phase NBs), X2 (NBs and stem cell progeny) or Xins (differentiated cells) (**Figure 1E**). We confirmed this analysis by Gene Ontology (GO) analyses and verification of the proportional expression profiles for known genes previously described as being enriched in X1, X2 or Xins (**Supplementary Figure 4**) (Solana et al. 2012; Önal et al. 2012; Labbé et al. 2012). We also re-analysed FACS single-cell RNA-seq datasets in the context of our genome annotation, and visualisation of the data by breakdown into our defined FACS expression categories was entirely consistent with this data

(**Supplementary Figure 5**). In particular we note that those genes with the highest proportion of expression the X2 compartment are indicative of genes expressed in post-mitotic undifferentiated NB progeny, with only very little expression in NBs themselves **(Supplementary Figure 5).**

Together these analyses provide a set of annotations and expression values that are directly related to the genome assembly, allowing integration of ChIP-seq data to investigate correlations between epigenetic marks and gene expression in the different planarians cell FACS compartments.

An optimized ChIP-seq protocol reveals H3K4me3 and H3K36me3 levels correlate with gene expression in planarian NBs

Research into the epigenetic mechanisms governing stem cell pluripotency in planarian NBs is still in its infancy (Dattani et al. 2018). Previous work has uncovered a lack of endogenous DNA methylation in the *Schmidtea mediterranea* genome, and characterized loss of function phenotypes for members of the NURD complex (Scimone et al. 2010; Jaber-Hijazi et al. 2013; Vásquez-Doorman and Petersen 2016), COMPASS and COMPASS-like families (Hubert et al. 2013; Duncan et al. 2015; Mihaylova et al. 2017) The first study to utilize ChIP-seq in planarians documented the effects of *mll1/2* and *set1* RNAi with respect to the activation mark H3K4me3 (Duncan et al. 2015). However, we revisited this data and noted that the total number of ChIP-seq reads from -1 million X1 sorted NBs was comparatively low in comparison to those from *Drosophila melanogaster* S2 ‘carrier’ cells.

We developed an optimized ChIP-seq protocol for FACS sorted X1 NBs without the addition of excess 'carrier' cells. Instead, a -3% *Drosophila* 'S2' spike-in was added to our chromatin before immunoprecipitations (IP) simply as a method to normalize any technical differences in IPs across our replicate libraries (Orlando et al. 2014). We were able to generate high quality uniquely mapped reads to our annotated *Schmidtea mediterranea* genome using only 150-200,000 X1 cells per IP – 5 to 7 times less material than the previously established planarian protocol (Duncan et al. 2015) . With our protocol, *Drosophila* 'spike-in' reads accounted for an average of 27% of X1 H3K4me3 libraries compared to an average of 87% in the previous study's X1 H3K4me3 libraries. Moreover, *Drosophila* 'spike-in' reads accounted for 9% of our X1 H3K36me3 libraries compared with 99% of the single X1 H3K36me3 replicate included in a previous study (Duncan et al. 2015) (**Supplementary Figure 6**).

We tested the robustness of our ChIP-seq protocol with reference to both H3K4me3 and H3K36me3 – epigenetic marks that are known to positively correlate with gene expression in other model systems. H3K4me3 is laid down by the trithorax group (trxG) complexes containing SET or MLL enzymes at active promoters near TSSs (Bledau et al. 2014; Denissov et al. 2014; Sirén et al. 2014; Hu et al. 2013). H3K36me3 is a mark of transcriptional elongation, and is deposited on histones as they are displaced by RNA polymerase II and as such this modification is enriched towards the 3' end of genes (Wagner and Carpenter 2012; Li et al. 2002). H3K36me3 is hypothesized to prevent spurious transcriptional initiation at cryptic promoter-like sequences within exons and, in yeast, this is achieved by the

209 recruitment of histone deacteylase complexes (HDAC) that erases elongation-
210 associated acetylation (Carrozza et al. 2005; Joshi and Struhl 2005).
211
212 As predicted, ChIP-seq of H3K4me3 in X1 NBs revealed a high average peak
213 around the TSSs of genes characterized as being X1 enriched (**Figure 2A**).
214 Conversely, we observed comparatively lower H3K4me3 deposition at the TSSs of
215 Xins enriched genes not expressed or expressed only at very low levels in X1 cells.
216 Intermediate levels of H3K4me3 in the X2 compartment are consistent with this
217 FACS population being a mixture of NBs and post-mitotic progeny. Indeed, genes
218 with the highest proportion of X2 expression (i.e. 'high ranking X2 genes') indicative
219 of expression in post-mitotic progeny but not NBs had lower levels of H3K4me3 in
220 X1 cells compared with low ranking X2 genes that retain expression in cycling NBs
221 **(Figure 2B).** A base by base Spearman's Rank correlation of ChIP-seq signal to
222 FACS proportional expression values of annotated loci across a 2.5 kb region either
223 side of the TSS, shows a positive correlation between genes defined by high X1
224 proportional expression and the levels of H3K4me3 deposition close to the TSS
225 (**Supplementary Figure 7A**). On the other hand, there is a negative correlation
226 between H3K4me3 deposition and genes with high Xins proportional expression
227 across the same region. Thus, a high H3K4me3 ChIP-seq signal reflects higher
228 expression of a locus in X1 NBs, whereas lower H3K4me3 signal reflects lower X1
229 NB gene expression but higher expression in the differentiated Xins compartment.
230
231 ChIP-seq plots of H3K36me3 split by FACS gene expression revealed, as predicted,
232 a higher average peak around X1 enriched genes when compared with the X2 and
233 Xins FACS enrichment categories (**Figure 2C**). Importantly, the average peak for X1

234 genes is located towards the 3' end of genes, whereas the smaller Xins peak is
235 promoter-proximal by comparison. This can be explained by a higher level of
236 transcriptional elongation of X1 transcripts in NBs compared with Xins genes that
237 have a predominant expression in the differentiated compartment. When splitting X2
238 enriched genes by rank order, we observe that genes with highest expression in the
239 X2 compartment and, as a consequence lowest transcript abundance in NBs, have
240 an enrichment for H3K36me3 at the promoter-proximal end of the gene (**Figure 2D**).
241 Conversely, with decreasing X2 proportional expression and a concomitant increase
242 in transcriptional activity in the NB compartment, the average peak of H3K36me3 is
243 shifted downstream of the TSS towards the 3' ends of genes.
244
245 We also looked at the individual H3K4me3 and H3K36me3 profiles of genes known
246 to be highly expressed in NBs, and compared this to the signal for the suppressive
247 marks H3K4me1 and H3K27me3 (see later). We confirmed that known metazoan
248 genes associated with stem cell maintenance, such as cell-cycle and replication
249 related genes (i.e. *mcm2, cyclin-B1, wee1, ctd1*), RNA-binding proteins (*piwi-1,*
250 *ddx52*), DNA-damage response (DDR) genes (*errc6-like, exonuclease 1*) and
251 epigenetic-related genes (*setd8-1)*, all have high levels of H3K4me3 at the promoter-
252 proximal end and H3K36me3 in the gene body, but a comparatively low signal for
253 the suppressive marks H3K4me1 and H3K27me3 (**Figure 2E**).
254
255 **Levels of repressive histone marks H3K27me3 and H3K4me1 at TSSs in NBs**
256 **correlate with gene expression**
257

Utilising our optimized ChIP-seq protocol, we investigated the occurrence of two additional histone modifications: H3K27me3, a repressive promoter mark catalysed by the PRC2 complex, and H3K4me1, a mark mediated by the MLL3/4 family of histone methyltransferases that correlates both with active enhancers and inactive promoter regions (Cheng et al. 2014; Calo and Wysocka 2013).

Genes that are categorized as being X1 enriched have low levels of H3K27me3 deposition at the TSS, compared with Xins enriched genes which are silenced in NBs (**Figure 3A**). A positive correlation is observed between the level of H3K27me3 and expression in the Xins compartment in a window from the TSS to 1kb downstream. This fairly broad domain of H3K27me3 deposition is consistent with previous studies in mammals (**Supplementary Figure 7b**) (Hawkins et al. 2011; Pauler et al. 2009). Conversely, a negative correlation at the TSS is observed between H3K27me3 signal and genes with high X1 expression (**Supplementary Figure 7b**). Consequently, the genome wide pattern for H327me3 is the opposite to that observed for H3K4me3. When splitting X2 genes by rank we note that genes with higher transcriptional enrichment in the post-mitotic compartment have a higher overall level of H3K27me3 at the promoter proximal region compared to genes that have NB expression (**Figure 3B)**.

The distribution of the H3K4me1 mark is noticeably different compared to that observed for either H3K27me3 or H3K4me3. Specifically, Xins loci have high levels of H3K4me1 at the TSS in X1 NBs, consistent with these genes being expressed at low levels in NBs, whereas X1 loci have H3K4me1 peaks that are on average -1kb downstream of the TSS **(Figure 3C)**. This data suggests that the H3K4me1 signal

283 shifts away from the TSS for genes that are actively expressed in NBs, in agreement
284 with previous observations in mammals (Cheng et al, 2014). Further evidence of this
285 peak shifting comes from analysis of X2 enriched genes sorted by rank order of
286 expression (**Figure 3D**). Highly ranked X2 genes are marked with H3K4me1 at the
287 promoter-proximal region. As the proportion of X2 enrichment decreases, indicative
288 of increasing expression the G1 NB compartment, the average H3K4me1 profile
289 becomes bimodal, eventually shifting downstream of the TSS (**Figure 3D**).
290
291 We plotted the epigenetic profiles of individual genes known to have high Xins
292 proportional expressions and that have validated expression patterns both by single-
293 cell RNA sequencing data and *in situ* hybridisations (**Figure 3E**) (Fincher et al. 2018;
294 Plass et al. 2018). For example, these genes are expressed almost exclusively in the
295 muscle (*COL21A1, slit1*), parenchyma (*glipr1, tolloid-like 1*), cathepsin+ cells (*dd961,*
296 *aquaporin 1*), non-cilliated neurons (*tph*, *dd8060*), and protonephridia (*Na/Ca*
297 *exchanger-like*), all have high H3K27me3 signal at the TSS consistent with these
298 genes being silenced in NBs. Moreover, these Xins enriched genes all have a high
299 H3K4me1 signal at the TSS that anti-correlates with H3K4me3 deposition, in support
300 of an earlier hypothesis that H3K4me1 limits the role of H3K4me3 interacting
301 proteins (Cheng et al. 2014). We also observe an atypical placement of H3K36me3
302 at the TSS of individual Xins genes which supports the previous suggestion that that
303 H3K36me3 may silence genic loci when placed at a promoter-proximal region of a
304 gene (Wu et al. 2011).
305
306 **Correlations of H3K7me3 and H3K4me3 profiles against FACS proportions**
307 **provide evidence for promoter bivalency in NBs**

Having demonstrated that known active and suppressive marks correlate with gene expression in planarian NBs, we investigated whether promoter bivalency could act to keep genes in a poised state prior to the onset of differentiation. Bivalent promoters are characterized by the presence of both the activating mark H3K4me3 and repressive mark H3K27me3. The simultaneous presence of both these marks keeps the gene in a poised transcriptional state, with low or no expression, and upon differentiation resolves such that only one of the two marks is dominant. We reasoned that loci that are off or have relatively low proportional expression in X1 NBs, but which are upregulated during the differentiation process in post-mitotic progeny (high X2 expression), would be good candidates for potential regulation by bivalent promoters in NBs. Additionally, in the absence of sequential or co-ChIP-seq technologies for planarians, using genes no or very low expression in NBs greatly reduces the likelihood that any bivalent signals are due to cell heterogeneity. This is because these genes would not be expected to have high levels of H3K4me3 in any (or at least very few cells) in the X1 NB compartment.

We plotted the percentage of maximum coverage for both H3K4me3 and H3K27me3 for the top 1000 genes for each of the three FACS enrichment categories (**Figure 4A-C**). A plot for the top 1000 X1 genes shows that these genes have a higher level of H3K4me3 compared to H3K27me3 (**Figure 4A**), whereas the top 1000 Xins genes have on average a much higher H3K27me3 signal compared to H3K4me3 (**Figure 4C)**. Consistent with our hypothesis, the top 1000 X2 genes, have peaks that are of similar magnitude for both of these functionally opposing epigenetic marks (**Figure 4B)**.

We also plotted the epigenetic profiles of genes that are downregulated following RNAi of the planarian homolog of the RNA-binding protein MEX3 (Zhu et al. 2015). Previously, *mex3-1* has been shown to be necessary for generating the differentiated cells of multiple lineages, and consistent with a role in the differentiation process we found that the downregulated genes (downregulated 2-fold; p-value <=0.05) had a higher average X2 proportional expression value (62.4%) compared with that of X1 (12.5%) (**Supplementary File 2**). As expected, we note a paired H3K4me3 and H3K27me3 ChIP-seq signal for these *mex3-1* downregulated genes (**Figure 4D)**.

One possibility is that that our observations are a result of genes with high X2 expression only having the H3K4me3 mark only whilst other genes exist in a H3K27me3-only state in NBs. This would produce an average profile that appears bivalent when many genes are looked at simultaneously. To rule out this possibility, we plotted the distribution of Pearson correlation coefficients between H3K4me3 and H3K27me3 for the top 500 ranked X1, X2 and 285 *mex3-1(RNAi)* downregulated loci. This showed a strong positive correlation between H3K4me3 and H3K27me3 for top 500 X2 loci and *mex3-1* downregulated loci, compared to a weak or no average correlation for X1 loci (**Figure 4E**). This is consistent with the interpretation that bivalency is present at promoters of genes that are highly enriched for expression in the X2 compartment.

Planarian orthologs to mammalian bivalent genes are marked by H3K4me3, H3K27me3 and paused RNA Pol II at the promoter-proximal region

RNA Polymerase II (RNAPII) pausing at genes that are highly inducible has been hypothesized to play a pivotal role in preparing genes for rapid induction in response to environmental or developmental stimuli. In a number of mammalian cellular contexts, bivalent genes have been shown to have a high density of paused RNA Pol II at the promoter-proximal region compared to genes which are actively transcribed, therefore allowing genes to be maintained in a transcriptionally poised state (Stock et al. 2007; Ferrai et al. 2017; Liu et al. 2017). Paused RNA Pol II can be distinguished from other forms by a phosphorylation at Ser5 (Ser5P) of the YSPTSPS heptad repeat at the C-terminus of the largest subunit of the Pol II complex. This heptad repeat is conserved across metazoans (Corden 2013), and is found in *S. mediterranea*.

ChIP-seq for RNAPII-Ser5P in NBs revealed that X2 enriched genes have a higher level of paused RNA Pol II at the promoter proximal region compared to X1 genes (**Figure 5A**). More significantly, highly ranked X2 genes with high expression in post-mitotic progeny and little expression in NBs have the highest amount of paused RNA Poll II close to the TSS, and with increasing expression in NBs the enrichment for this mark decreases (**Figure 5B**).

We calculated the pausing index (PI) for all annotated genes in our genome that have a total annotated length of ≥ 1kb. For our particular genome annotation, we calculated the PI as the read coverage (normalized to input) +/- 500bp either side of the annotated TSS divided by the normalized read coverage from +500bp to +2500bp from the TSS (**Figure 5C**). We applied a conservative definition of a gene as being significantly stalled for transcription if the PI ≥ 1. As expected, individual

383 genes highly expressed in the NB compartment had both a low PI and were not
384 enriched for RNAPII-Ser5P at the promoter-proximal region, thereby confirming our
385 methodology was accurate at the gene level (**Figure 5D**). We also found that X2
386 genes with high PI scores had, on average, higher Pearson correlation coefficients
387 between H3K4me3 and H3K27me3 (indicative of a bivalent state) compared with
388 both X1 and X2 genes that have lower PI scores (**Supplementary Figure 8**). Given
389 this correlation, we chose individual X2 enriched genes with high PI values and
390 plotted the ChIP-seq profiles for H3K4me3, H3K27me3 and RNAPII-Ser5P as a
391 percentage of maximum coverage for each mark.
392
393 Amongst genes enriched for these three signatures of bivalent promoters were those
394 that have orthology to transcription factor (TF) families and include the Hox (*hoxb9*),
395 Nkx (*nkx1.2*), Even-skipped (*evx-1*), Paired-like (*phox2A)* and T-box (*tbx2*) and Tlx
396 (*tlx1-like*) gene classes (**Figure 5E**). Indeed, previous studies in both mouse ESCs
397 (Bernstein et al. 2006) and quiescent muscle stem cells (Liu et al. 2013) have shown
398 that members of these gene families are typically marked by both H3K4me3 and
399 H3K27me3. A paired level of these marks at the TSS for these individual genes
400 suggests the existence of bivalent chromatin states at these conserved
401 developmental genes and confirms our correlational analysis of X2 loci (**Figure 4E**).
402 Moreover, single-cell sequencing data and pseudotime analyses plots made from
403 single cell data show that these genes are expressed at detectable levels in very
404 few, if any, *smedwi-1*+ cells (the archetypal NB marker) and are instead enriched in
405 post-mitotic cells of specific lineages (**Supplementary Figure 9**) (Fincher et al.
406 2018; Plass et al. 2018).
407

One caveat of our analyses is that the bivalent profiles of X2 enriched differentiation related genes may, for some individual genes that appear bivalent, reflect admixture of transcriptionally active and repressed states within the X1 NB compartment. For example, previous work has shown that the X1 compartment is highly heterogeneous with subsets of *piwi-1*+ NBs expressing lineage specific TFs (Van Wolfswinkel et al. 2014). These genes, such as *SoxP-3* and *egr-1,* which are in fact X2 enriched according to our dataset and others (Labbé et al. 2012), appear to have a paired H3K4me3 and H3K27me3 signal (**Figure 5E**). Given that they are known to be expressed in a subset of cells in the X1 compartment and are definitive markers of lineage-primed NB subsets that will go through one more cell division (as validated by *in situ* hybridisation, condensin knockdown studies (Van Wolfswinkel et al. 2014; Lai et al. 2018) and single-cell RNA-seq data (Wurtzel et al. 2015; Plass et al. 2018; Fincher et al. 2018) no definitive conclusions concerning bivalency of these particular genes can be reached.

Discussion

In this study, we have produced a *Schmidtea mediterranea* asexual genome annotation based on gene expression, and integrated FACS RNA-seq datasets from different laboratories to calculate consensus proportional expression values for each annotated locus in the X1, X2 and Xins cellular compartments. We have developed an optimized ChIP-seq protocol, and employed this to generate robust genome-wide profiles of the active H3K4me3 and H3K36me3 marks and repressive H3K4me1 and H3K27me3 marks in planarian NBs.

We find that the active marks H3K4me3 and H3K36me3, and suppressive H3K4me1 and H3K27me3 marks in X1 NBs correlate with the proportion of total transcript expression of these loci in X1 cells, validating our NB ChIP-seq methodology. These analyses showed that genes associated with stem cell differentiation, and which are expressed at low levels in X1 population but activated at high levels in the X2 population, are marked with both H3K4me3 and H3K27me3 marks at comparable levels at the TSS. Moreover, these genes were also highly marked with paused RNA polymerase (RNAPII-Ser5P) at the promoter region consistent with the definition of transcriptionally poised bivalent genes. Although we cannot entirely rule out cell heterogeneity within the X1 NB population as a factor contributing to our observation of promoter bivalency, our focus on both genes with high X2 expression (post-mitotic NB progeny) and orthology to vertebrate transcription factors known to have bivalent profiles, provide strong evidence that bivalent histone marks may be involved in poising of genes for activation upon NB commitment and differentiation in planarians.

The existence of promoter bivalency in invertebrates, prior to our work here, has been contentious. For example, the mammalian orthologs of bivalent genes in *Drosophila* germ cells were found to have only repressive H3K27me3 deposited at their promoters (Schuettengruber et al. 2009; Gan et al. 2010; Lesch et al. 2016). However, in a more recent study using fly embryos, the Pc-repressive complex 1 (PRC1) that binds to H3K27me3 was shown to co-purify with both the Fsh1 (ortholog of mammalian BRD4) that binds to acetylated histone marks and to Enok/Br140 (orthologs to subunits of mammalian MAZ/MORF histone acetyltransferase complex). ChIP-seq identified two groups of PRC1/Br140 genomic binding sites that were either defined by strong H3K27me3 signal or strong H3K27ac signal (i.e.

actively transcribed genes). Both groups were also marked with narrow peaks of H3K4me3 at the TSS (Kang et al. 2017). These recent findings also argue for the existence of bivalent-like promoters outside of vertebrates, at least with respect to the binding of chromatin regulatory complexes, and extends the model to suggest that acetylation may be important in the resolution of bivalent protein complexes during development.

One key role of bivalency is thought to be to allow the maintenance of pluripotency in ESCs, by having genes involved in differentiation and commitment both silent but competent to switch on if the right signals are received. Our data suggest that this mechanism is likely to be important for pluripotency in planarian pluripotent NBs, as genes that can switch on rapidly upon differentiation appear to be the bivalent. Indeed, these genes also included planarian orthologs to mammalian TFs that have been documented to be bivalent in ESCs. Consequently, we are able to present a case for promoter bivalency in planarian NBs and in doing so demonstrate that this process is not necessarily vertebrate-specific. This novel finding adds to the growing body of evidence which suggests a deep conservation of regulatory mechanisms involved in stem cell function (Juliano et al. 2010; Alié et al. 2015; Solana 2013; Solana et al. 2016; Lai and Aboobaker 2018) as well as combinatorial patterns of post-translational modifications (Schwaiger et al. 2014; Sebé-Pedrós et al. 2016; Gaiti et al. 2017). Epigenetic studies in the unicellular relative of metazoans, *Capsaspora owczarzaki*, could not find any evidence of bivalency given the absence of H3K27me3, and epigenetic studies in the sponge *Amphimedon queenslandica* (Gaiti et al. 2017) and the cnidarian *Nematostella vectensis* (Schwaiger et al. 2014)

have also not revealed any evidence for this approach to gene regulation. Further work will be required to establish when bivalent chromatin evolved in animals.

Overall our development of a robust ChIP-seq protocol for use with planarian sorted NBs, together with good coverage for four definitive and essential epigenetic marks, establishes a resource for the future planarian studies investigating the epigenetic regulation of stem cell function.

Materials and Methods

Reference assembly and annotations

Previous transcriptome assemblies (Oxford (ox_Smed_v2), Dresden (dd_smed_v4), SmedGD Asexual, Smed GD Unigenes) were downloaded from and PlanMine (Brandl et al. 2016) and Smed GD 2.0 (Robb et al. 2008). NCBI complete CDS sequences for *Schmidtea mediterranea* were also downloaded. Sequences were aligned to the SmedGD Asexual 1.1 genome with GMAP (Wu and Watanabe 2005) and consolidated with PASA. An independent reference assembly was also performed by mapping 164 available RNA-seq datasets with HISAT2 (Sirén et al. 2014) and assembly was performed with StringTie. The PASA consolidated and StringTie annotations were merged with StringTie.

An intron jaccard score (intersection of introns / union of introns) was calculated for all overlapping transcripts. Pair-wise jaccard similarity scores of 0.9 or greater were used to create a graph of similar annotations. From the resultant cliques of

transcripts, one was chosen to be the representative transcript for the locus, by prioritizing transcript length, ORF length, and BLAST homology.

Strand information for annotations was assigned by utilising in house strand-specific RNA-seq libraries, BLAST homology, and longest ORF length. Transdecoder was run utilizing PFAM and UNIPROT evidence to identify protein-coding transcripts in the annotations. Detail methods are recorded in an IPython notebook (**Supplementary File 3**).

FACS proportional expression value generation for annotated loci

Kallisto (Bray et al. 2016) was used to pseudo-align RNA-seq libraries originating from four labs (Önal et al. 2012; Labbé et al. 2012; Van Wolfswinkel et al. 2014; Zhu et al. 2015; Duncan et al. 2015) (**Supplementary Figure 2**) to our expression-based annotation of the asexual *S.mediterranea* genome. This generated TPM values for each annotated locus. Sleuth was used to calculate a normalization factor for each library. For each locus, the TPM values of member transcripts (potential isoforms) were summed to generate a consensus TPM value and then normalized accordingly. Replicates within each lab dataset were then averaged.

Normalized TPM values for each lab dataset were converted to a proportional value as a representation of expression in FACS categories. We next calculated three sets of pairwise ratios (X1:X2, X1:Xins, X2:Xins) using these proportional values. Given two of the three ratios, a third ratio can be ‘predicted’. Consequently, we calculated 3 ‘observed’ ratios and 3 ‘predicted ratios’. A good Spearman’s rank correlation was

observed for the X2:Xins ratio and as such we kept these observed proportions, and calculated an inferred X1 proportion. Detailed methodology is documented in **Supplementary File 4** and full list of X1,X2, and Xins proportional values is available in **Supplementary File 5**.

ChIP-seq protocol

For each experimental replicate, 600'000-700'000 planarian X1 cells were isolated, (sufficient for ChIP-seq of 3 histone marks and an input control) by utilisation of a published FACS protocol (Romero et al. 2012). We dissociated cells from an equal number of head, pharyngeal, or tail pieces from 3-day regenerating planarians. For whole worm ChIP-seq, wild-type worms were starved for 2 weeks prior to dissociation.

Following FACS, cells were pelleted. The pellet was re-suspended in Nuclei Extraction Buffer (0.5% NP40,0.25% Triton X-100, 10mM Tris-Cl pH7.5, 3mM CaCl2, 0.25mM Sucrose, 1mM DTT, phosphatase cocktail inhibitor 2, phosphatase cocktail inhibitor 3). A 3% *Drosophila* S2 cell spike-in was added at this point. This was followed by 1% formaldehyde fixation for 7mins, which was quenched with the addition of glycine to a final 125mM concentration. The nuclei pellet was re-suspended in SDS lysis buffer (1% SDS, 50mM Tris-Cl pH8.0, 10mM EDTA) and incubated on ice, followed by the addition of ChIP dilution buffer. Samples were sonicated and 1/10th volume of Triton X-100 was added to dilute SDS in the solution. Samples were pellet, and supernatant was collected that contained the sonicated chromatin. Test de-crosslinking was performed on 1/8th of the sonicated chromatin, and analysed using a TapeStation DNA HS tape to verify the DNA fragment range

was between 100-500bp. Commercial *Drosophila* S2 chromatin (Active Motif 53083) spike-in was added at this point (at 3% of the amount of amount of *S. mediterranea* prepared chromatin) if S2 cells had been added earlier before chromatin preparation.

Protein A-covered Dynabeads were used for immunoprecipitation (IP). 50µl of Dynabeads were incubated overnight at 4c with 7µg of antibody (H3K4me3 Abcam ab8580; H3K36me3 Abcam ab9050; H3K4me1 Abcam ab8895; H3K27me3 Abcam ab6002; RNAPII-Ser5P ab5131) diluted in 0.5% BSA/PBS. Following incubation, Dynabeads were washed with 0.5% BSA/PBS, and ¼ of the total isolated chromatin was added per IP. Following overnight incubation, washes were performed 6 times with RIPA buffer (50mM HEPES-KOH pH 8.0, 500mM LiCl, 1mM EDTA, 1% NP-40, 0.7% DOC, protease inhibitors). Dynabeads were washed with TE buffer and re-suspended in Elution Buffer (50mM Tris-Cl pH 8.0, 10mM EDTA, 1% SDS). Protein was separated from Dynabeads by incubating for 15mins at 65c on a shaking heating block at 1400rpm. Eluates were de-crosslinked at 65c overnight. Input chromatin (1/8th of the total chromatin amount) was also de-crosslinked at this point. Following incubation, RNAseA (0.2µg) and Proteinase K (0.2µg) was added to each sample and incubated for 1hr. DNA was purified with phenol:chloroform extraction followed by ethanol precipitation. DNA is re-suspended in TE and quantified with Qubit dsDNA HS kit. NEB Ultra II kit was used for library preparation, and clean-up was performed with Agencourt Ampure XP beads. Samples were paired-end sequenced at a length of 75 nucleotides on the Illumina NextSeq.

ChIP-seq analysis

Reads were trimmed with Trimmomatic (Bolger et al. 2014) and aligned to a concatenated file containing both our annotated *Schmidtea mediterranea* genome as well as the *Drosophila melanogaster* release 6 reference genome (Hoskins et al. 2015) using BWA mem 0.7.12 (Li and Durbin 2009). Only uniquely mapping reads were considered further. Paired reads that map to both species were also removed. Picard tools-1.115 was used to remove duplicate reads. Reads were separated into sets that mapped to *Drosophila* or *S. mediterranea* using custom python scripts (documented in IPython notebook in **Supplementary File 6**). The number of reads aligning to the *Drosophila* genome were calculated for use in normalization calculations. For each paired or single map read, coordinates representing 100bp at the centre of sequenced were parsed and written to a BED file.

The genomecov function was used in BEDTools 2.27.0 (Quinlan and Hall 2010) to generate coverage tracks in bedgraph format. The resultant bedgraph file was converted to bigwig format UCSC's bedgraphtoBigWig tool (Kent et al. 2010). Deeptools2's computeMatrix was used to extract coverage around 2.5kb or 5kb either side of the annotated TSS for each annotated locus in 50bp windows for each sample and corresponding input (Ramírez et al. 2016). A normalization factor was calculated using the number of mapped reads corresponding to the *Drosophila* spike-in to control for between IP technical variation (Orlando et al. 2014). A scaling factor for input ChIP-seq libraries was calculated using the deepTools2 python API that uses the SES method (Diaz et al. 2012). The mean normalized coverage was calculated for each sample and input. The normalized input coverage was subtracted from the normalized sample coverage to generate a final coverage track for downstream visualization and analyses. Individual gene profiles for given ChIP-seq

tracks could then be visualised and sequences for those genes plotted in this paper are given in **Supplementary File 7**.

To calculate the correlation of ChIP-seq signal coverage to proportional FACS expression, two vector values were calculated. The first vector was proportional FACS expression for all genomic loci. The second vector was ChIP-seq coverage at each 50bp position 2.5kb either side of the TSS. A Spearman's rank correlation was performed on both vectors yielding a correlation value for the assayed position. The correlation value for each non-overlapping 50bp window was then plotted on a graph.

For comparison of profiles between different epigenetic marks a percentage coverage was calculated for each mark. The maximum coverage was found across all 5kb or 10kb regions for all loci. Absolute normalized coverage for each 50bp window was then divided by the maximum coverage observed for that mark in the genome, resulting in a percentage coverage in each 50bp window for each mark.

For calculation of pausing index, we added normalized coverage to input 500bp either side of the annotated TSS for each gene and divided this value by the total coverage between 500bp and 2.5kb downstream of the TSS.

Detailed methods for ChIP-seq analysis are documented in **Supplementary File 6**.

Data access:

ChIP short read data have been deposited in the NCBI BioProject under the accession (BioProject; https://www.ncbi.nlm.nih.gov/bioproject/) numbers

630 PRJNA471851 and PRJNA338116. Annotations made on the *S. mediterranea*
631 genome and used in this study are available as compressed GFF file
632 (**Supplementary File 8**)
633
634 **Acknowledgements:**
635 This work was funded grants from th the MRC (MR/M000133/1) and BBSRC
636 (BB/K007564/1) awarded to AA. AD is funded by a BBSRC studentship
637 (BB/J014427/1).
638
639 **Author contributions:**
640 AAA originally conceived and designed the study, upon which AD innovated. AD,
641 YM, and PA performed ChIP-seq experiments. SH, SS, and AL assisted with
642 optimization of ChIP-seq and RNA-seq protocols. AD and DK performed
643 bioinformatic analyses. AD, DK and AAA wrote, reviewed, and revised the
644 manuscript.
645

646 **Figure legends**

647

648 **Figure 1: A**. Overview of methodology for annotating the *Schmidtea mediterranea*
649 asexual genome based on expression. 164 RNA-seq datasets, 4 *de novo*
650 transcriptome assemblies, NCBI complete CDS sequences, and Smed Unigenes
651 were mapped to the SmedGD Asxl v1.1 genome. Reference assemblies were
652 merged, cleaned to remove potential splice variant redundancies, and the best
653 representative transcript for each genomic locus was chosen. Strand information
654 was obtained by BLAST to Uniprot, prediction of longest ORF and data from strand-
655 specific libraries. This process yielded a total set of 38,771 loci. **B**. Methodology for
656 gating X1, X2, Xins cell populations based on nuclear to cytoplasmic ratio during
657 Fluorescent Activated Cell Sorting (FACS). **C**. Diagram depicting how X1, X2, Xins
658 FACS population gates relate to cell cycle and differentiation stage. **D.** Overview of
659 methodology for categorization of annotated loci based on FACS RNA-seq datasets.
660 FACS RNA-seq datasets were mapped to our annotated genome using Kallisto and
661 normalized with Sleuth. Normalization was done individually for each of the lab's
662 datasets. Normalized TPMs were converted to proportions between available FACS
663 categories of each lab, and a final consensus X1:X2:Xins proportion was calculated.
664 **E.** A table presenting number of loci in different FACS classification groups, as well
665 as number of protein-coding genes in each group based on Transdecoder evidence.

666

667 **Figure 2:** Histone marks for actively transcribed genes in X1 NBs. **A**. Average
668 H3K4me3 ChIP-seq coverage profiles across X1, X2 and Xins enriched loci in X1
669 NBs across biological replicates following outlier removal. Y-axis represents the
670 difference in coverage between sample and input. X-axis represents 2.5kb up- and

671 downstream of the TSS. H3K4me3 signal is highest around promoter-proximal
672 region close to the TSS for X1 enriched loci in NBs consistent with the role of
673 H3K4me3 in active transcription. **B.** H3K4me3 ChIP-seq profiles following outlier
674 removal for X2 genes ranked from high to low X2 proportional expression. H3K4me3
675 signal in NBs decreases with an increase in proportion of X2 expression, indicative
676 of high-ranking X2 genes having a predominant role in post-mitotic progeny as
677 opposed to NBs. **C.** Average H3K36me3 ChIP-seq profile across X1, X2 and Xins
678 enriched loci in X1 NBs across biological replicates following outlier removal. Y-axis
679 represents the difference in coverage between sample and input. X-axis represents
680 2.5kb upstream and downstream of the TSS. Signal for H3K36me3 is promoter-
681 proximal for Xins genes, whereas the magnitude of signal is greater and shifted 3' for
682 X1 genes. **D.** H3K36me3 ChIP-seq profiles following outlier removal for X2 genes
683 from high to low X2 proportional ranking. H3K36me3 signal in NBs shifts to the 3'
684 end with a decrease in X2 proportion, consistent with these lowly-ranked genes
685 having transcriptional activity in NBs. **E.** H3K4me3 and H3K36me3 (active marks)
686 and H3K4me1 and H327me3 (suppressive marks) ChIP-seq profiles for highly-
687 expressed X1 genes in NBs. Y-axis represents percentage coverage for each mark
688 and allows for the 4 epigenetic marks to be directly compared. X-axis represents 1.0
689 kb upstream and 2.5 kb downstream of the TSS. Pie Charts represent proportional
690 expression for each gene in X1 (dark blue), X2 (light blue) and Xins (orange).
691
692 **Figure 3:** Histone marks for inactive genes in X1 NBs. **A**. Average H3K27me3 ChIP-
693 seq profile across X1, X2 and Xins enriched loci in X1 NBs across 3 biological
694 replicates following outlier removal. Y-axis represents the difference in coverage
695 between sample and input. X-axis represents signal 2.5kb upstream and

696 downstream of the TSS. **B.** H3K27me3 ChIP-seq profiles following outlier removal
697 for X2 genes from high to low X2 proportional ranking. H3K27me3 signal increases
698 with an increase in proportion of X2 gene expression, indicative of these high-
699 ranking X2 genes being transcriptionally silenced or lowly expressed in NBs. **C.**
700 Average H3K4me1 ChIP-seq profiles following outlier removal across X1, X2 and
701 Xins enriched loci in X1 NBs**.** Y-axis represents the absolute difference in coverage
702 between sample and input. X-axis represents signal 2.5kb upstream and
703 downstream of the TSS. **D.** H3K4me1 ChIP-seq profiles following outlier removal for
704 X2 genes from high to low X2 proportional ranking. Highly ranked X2 genes have a
705 H3K4me1 signal at the promoter-proximal region, and a decrease in X2 ranking
706 coincides with a peak shift -1kb downstream of the TSS. **E.** H3K4me3, H3K36me3,
707 H3K4me1 and H327me3 NB ChIP-seq profiles for highly-expressed Xins genes. The
708 Y-axis scale represents percentage coverage for each mark. X-axis represents 1.0kb
709 upstream and 2.5 kb downstream of the TSS. Pie Charts represent proportional
710 expression for each gene in X1 (dark blue), X2 (light blue) and Xins (orange).
711
712
713 **Figure 4: A-E:** Average H3K4me3 and H3K27me3 ChIP-seq profiles in X1 NBs
714 across 3 biological replicates. Y-axis is percentage coverage after normalization to
715 input to allow both ChIP-seq profiles to be directly compared. Plots are shown for: **A.**
716 Top 1000 ranked X1 genes by expression. **B.** Top 1000 ranked X2 genes. **C.** Top
717 1000 ranked Xins genes. **D.** 285 *Smed-mex3-1* down-regulated loci with >2-fold
718 change ($p<0.05$). **E.** A distribution of Pearson correlation values for top 500 X1, X2,
719 and 285 =>2-fold downregulated *mex-3-1* downregulated loci. The Pearson
720 Correlation coefficient was calculated between the H3K4me3 and H3K27me3 values

721 at each 50bp window starting from -1000bp downstream of the TSS and 1500bp
722 upstream of the TSS.
723
724 **Figure 5: A.** Average paused RNAPII-Ser5P ChIP-seq profile across X1 and X2
725 enriched loci in X1 NBs across biological replicates following outlier removal. Y-axis
726 represents the difference in coverage between sample and input. X-axis represents
727 signal 2.5 upstream and downstream of the TSS. **B.** RNAPII-Ser5P ChIP-seq
728 profiles following outlier removal for X2 genes from high to low X2 proportional
729 ranking. RNAPII-Ser5P signal increases with an increase in proportion of X2 gene
730 expression, indicative of these high-ranking X2 genes being transcriptionally silenced
731 but maintained in a permissive state for rapid induction. **C.** Calculation for Pausing
732 Index (PI) of genes => 1kb. We divided normalized coverage between +/- 500bp
733 TSS by normalized coverage +500bp to +2.5 kb. For genes under 2.5 kb, we
734 inspected RNAPolI-Ser5P profiles visually to confirm whether Pol II pausing was
735 enriched at the promoter-proximal region. **D.** Individual profiles for H3K34me3 and
736 H3K27me3 of highly enriched NB X1 genes. X1 genes have a high level of
737 H3K4me3 and levels of H3K27me3 correspond to intron regions and are not
738 enriched at the promoter-proximal region. RNAPII-Ser5P signal is not enriched at the
739 promoter-proximal region compared with the gene body, and as a result PI <1. **E.**
740 We selected highly enriched X2 genes with a PI => 1 that have both H3K4me3 and
741 H3K27me3 enriched at the promoter-proximal region, together with an enrichment of
742 RNAPII-Ser5P close to the TSS. Pie Charts represent proportional expression for
743 each gene in X1 (dark blue), X2 (light blue) and Xins (orange).
744
745

746 References:

747

748 Aboobaker AA. 2011. Planarian stem cells: A simple paradigm for regeneration.
749 *Trends Cell Biol* **21**: 304–311.
750 Akkers RC, van Heeringen SJ, Jacobi UG, Janssen-Megens EM, Françoijs KJ,
751 Stunnenberg HG, Veenstra GJC. 2009. A Hierarchy of H3K4me3 and
752 H3K27me3 Acquisition in Spatial Gene Regulation in Xenopus Embryos. *Dev*
753 *Cell* **17**: 425–434.
754 Alié A, Hayashi T, Sugimura I, Manuel M, Sugano W, Mano A, Satoh N, Agata K,
755 Funayama N. 2015. The ancestral gene repertoire of animal stem cells. *Proc*
756 *Natl Acad Sci* 201514789.
757 Baguñá J, Romero R. 1981. Quantitative analysis of cell types during growth,
758 degrowth and regeneration in the planarians Dugesia mediterranea and Dugesia
759 tigrina. *Hydrobiologia* **84**: 181–194.
760 Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A,
761 Wernig M, Plath K, et al. 2006. A Bivalent Chromatin Structure Marks Key
762 Developmental Genes in Embryonic Stem Cells. *Cell* **125**: 315–326.
763 Bledau AS, Schmidt K, Neumann K, Hill U, Ciotta G, Gupta A, Torres DC, Fu J,
764 Kranz A, Stewart AF, et al. 2014. The H3K4 methyltransferase Setd1a is first
765 required at the epiblast stage, whereas Setd1b becomes essential after
766 gastrulation. *Development* **141**: 1022–1035.
767 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina
768 sequence data. *Bioinformatics* **30**: 2114–2120.
769 Brandl H, Moon HK, Vila-Farré M, Liu SY, Henry I, Rink JC. 2016. PlanMine - A
770 mineable resource of planarian biology and biodiversity. *Nucleic Acids Res* **44**:

771 D764–D773.
772 Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-
773 seq quantification. *Nat Biotechnol* **34**: 525–7.
774 Brookes E, De Santiago I, Hebenstreit D, Morris KJ, Carroll T, Xie SQ, Stock JK,
775 Heidemann M, Eick D, Nozaki N, et al. 2012. Polycomb associates genome-
776 wide with a specific RNA polymerase II variant, and regulates metabolic genes
777 in ESCs. *Cell Stem Cell* **10**: 157–170.
778 Calo E, Wysocka J. 2013. Modification of Enhancer Chromatin: What, How, and
779 Why? *Mol Cell* **49**: 825–837.
780 Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Alvarado AS,
781 Yandell M. 2008. MAKER: An easy-to-use annotation pipeline designed for
782 emerging model organism genomes. *Genome Res* **18**: 188–196.
783 Carrozza MJ, Li B, Florens L, Suganuma T, Swanson SK, Lee KK, Shia WJ,
784 Anderson S, Yates J, Washburn MP, et al. 2005. Histone H3 methylation by
785 Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious
786 intragenic transcription. *Cell* **123**: 581–592.
787 Cheng J, Blum R, Bowman C, Hu D, Shilatifard A, Shen S, Dynlacht BD. 2014. A
788 role for H3K4 monomethylation in gene repression and partitioning of chromatin
789 readers. *Mol Cell* **53**: 979–992.
790 Corden JL. 2013. RNA polymerase II C-terminal domain: Tethering transcription to
791 transcript and template. *Chem Rev* **113**: 8423–8455.
792 Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K. 2009. Chromatin
793 Signatures in Multipotent Human Hematopoietic Stem Cells Indicate the Fate of
794 Bivalent Genes during Differentiation. *Cell Stem Cell* **4**: 80–93.
795 Dattani A, Sridhar D, Aziz Aboobaker A. 2018. Planarian flatworms as a new model

system for understanding the epigenetic regulation of stem cell pluripotency and differentiation. *Semin Cell Dev Biol*.

Denissov S, Hofemeister H, Marks H, Kranz A, Ciotta G, Singh S, Anastassiadis K, Stunnenberg HG, Stewart AF. 2014. Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development* **141**: 526–537.

Diaz A, Park K, Lim DA, Song JS. 2012. Normalization, bias correction, and peak calling for ChIP-seq. *Stat Appl Genet Mol Biol* **11**.

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 Rep* **13**: 2741–55.

Ferrai C, Torlai Triglia E, Risner-Janiczek JR, Rito T, Rackham OJ, de Santiago I, Kukalev A, Nicodemi M, Akalin A, Li M, et al. 2017. RNA polymerase II primes Polycomb-repressed developmental genes throughout terminal neuronal differentiation. *Mol Syst Biol* **13**: 946.

Fincher CT, Wurtzel O, de Hoog T, Kravarik KM, Reddien PW. 2018. Cell type transcriptome atlas for the planarian Schmidtea mediterranea. *Science* **360**: eaaq1736.

Gaiti F, Jindrich K, Fernandez-Valverde SL, Roper KE, Degnan BM, Tanurdžić M. 2017. Landscape of histone modifications in a sponge reveals the origin of animal cis-regulatory complexity. *Elife* **6**.

Gan Q, Schones DE, Ho Eun S, Wei G, Cui K, Zhao K, Chen X. 2010. Monovalent and unpoised status of most genes in undifferentiated cell-enriched Drosophila testis. *Genome Biol* **11**: R42.

Harikumar A, Meshorer E. 2015. Chromatin remodeling and bivalent histone

821 modifications in embryonic stem cells. *EMBO Rep* **in press**: 1609–1619.
822 Hattori N, Niwa T, Kimura K, Helin K, Ushijima T. 2013. Visualization of multivalent
823 histone modification in a single cell reveals highly concerted epigenetic changes
824 on differentiation of embryonic stem cells. *Nucleic Acids Res* **41**: 7231–7239.
825 Hawkins RD, Hon GC, Yang C, Antosiewicz-Bourget JE, Lee LK, Ngo QM, Klugman
826 S, Ching KA, Edsall LE, Ye Z, et al. 2011. Dynamic chromatin states in human
827 ES cells reveal potential regulatory sequences and genes involved in
828 pluripotency. *Cell Res* **21**: 1393–1409.
829 Hayashi T, Asami M, Higuchi S, Shibata N, Agata K. 2006. Isolation of planarian X-
830 ray-sensitive stem cells by fluorescence-activated cell sorting. *Dev Growth Differ*
831 **48**: 371–380.
832 Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, Booth BW, Pfeiffer
833 BD, George RA, Svirskas R, et al. 2015. The Release 6 reference sequence of
834 the Drosophila melanogaster genome. *Genome Res* **25**: 445–458.
835 Hu D, Garruss AS, Gao X, Morgan MA, Cook M, Smith ER, Shilatifard A. 2013. The
836 Mll2 branch of the COMPASS family regulates bivalent promoters in mouse
837 embryonic stem cells. *Nat Struct Mol Biol* **20**: 1093–1097.
838 Hubert A, Henderson JM, Ross KG, Cowles MW, Torres J, Zayas RM. 2013.
839 Epigenetic regulation of planarian stem cells by the SET1/MLL family of histone
840 methyltransferases. *Epigenetics* **8**: 79–91.
841 Jaber-Hijazi F, Lo PJKP, Mihaylova Y, Foster JM, Benner JS, Tejada Romero B,
842 Chen C, Malla S, Solana J, Ruzov A, et al. 2013. Planarian MBD2/3 is required
843 for adult stem cell pluripotency independently of DNA methylation. *Dev Biol* **384**:
844 141–153.
845 Joshi AA, Struhl K. 2005. Eaf3 chromodomain interaction with methylated H3-K36

links histone deacetylation to pol II elongation. *Mol Cell* **20**: 971–978.

Juliano CE, Swartz SZ, Wessel GM. 2010. A conserved germline multipotency program. *Development* **137**: 4113–4126.

Kang H, Jung YL, McElroy KA, Zee BM, Wallace HA, Woolnough JL, Park PJ, Kuroda MI. 2017. Bivalent complexes of PRC1 with orthologs of BRD4 and MOZ/MORF target developmental genes in Drosophila. *Genes Dev*.

Kent WJ, Zweig AS, Barber G, Hinrichs AS, Karolchik D. 2010. BigWig and BigBed: Enabling browsing of large distributed datasets. *Bioinformatics* **26**: 2204–2207.

Labbé RM, Irimia M, Currie KW, Lin A, Zhu SJ, Brown DDR, Ross EJ, Voisin V, Bader GD, Blencowe BJ, et al. 2012. A Comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells* **30**: 1734–1745.

Lai AG, Aboobaker AA. 2018. EvoRegen in animals : Time to uncover deep conservation or convergence of adult stem cell evolution and regenerative processes. *Dev Biol* **433**: 118–131.

Lai AG, Kosaka N, Abnave P, Sahu S, Aboobaker AA. 2018. The abrogation of condensin function provides independent evidence for defining the self-renewing population of pluripotent stem cells. *Dev Biol* **433**: 218–226.

Lesch BJ, Dokshin GA, Young RA, McCarrey JR, Page DC. 2013. A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. *Proc Natl Acad Sci* **110**: 16061–16066.

Lesch BJ, Page DC. 2014. Poised chromatin in the mammalian germ line. *Development* **141**: 3619–3626.

Lesch BJ, Silber SJ, McCarrey JR, Page DC. 2016. Parallel evolution of male germline epigenetic poising and somatic development in animals. *Nat Genet* **48**:

888–894.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li J, Moazed D, Gygi SP. 2002. Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. *J Biol Chem* **277**: 49383–8.

Liu J, Wu X, Zhang H, Pfeifer GP, Lu Q. 2017. Dynamics of RNA Polymerase II Pausing and Bivalent Histone H3 Methylation during Neuronal Differentiation in Brain Development. *Cell Rep* **20**: 1307–1318.

Liu L, Cheung TH, Charville GW, Hurgo BMC, Leavitt T, Shih J, Brunet A, Rando TA. 2013. Chromatin Modifications as Determinants of Muscle Stem Cell Quiescence and Chronological Aging. *Cell Rep* **4**: 189–204.

Mihaylova Y, Abnave P, Kao D, Hughes S, Lai A, Jaber-Hijazi F, Kosaka N, Aboobaker A. 2017. Conservation of epigenetic regulation by the MLL3/4 tumour suppressor in planarian pluripotent stem cells. *Bioarxiv*.

Molinaro AM, Pearson BJ. 2016. In silico lineage tracing through single cell transcriptomics identifies a neural stem cell population in planarians. *Genome Biol* **17**: 87.

Önal P, Grün D, Adamidi C, Rybak A, Solana J, Mastrobuoni G, Wang Y, Rahn HP, Chen W, Kempa S, et al. 2012. Gene expression of pluripotency determinants is conserved between mammalian and planarian stem cells. *EMBO J* **31**: 2755–2769.

Orlando DA, Chen MW, Brown VE, Solanki S, Choi YJ, Olson ER, Fritz CC, Bradner JE, Guenther MG. 2014. Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep* **9**: 1163–1170.

Palakodeti D, Smielewska M, Lu Y-C, Yeo GW, Graveley BR. 2008. The PIWI proteins SMEDWI-2 and SMEDWI-3 are required for stem cell function and piRNA expression in planarians. *RNA* **14**: 1174–1186.

Pauler FM, Sloane MA, Huang R, Regha K, Koerner M V., Tamir I, Sommer A, Aszodi A, Jenuwein T, Barlow DP. 2009. H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome. *Genome Res* **19**: 221–233.

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**: eaaq1723.

Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842.

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 Res* **44**: W160–W165.

Reddien PW, Oviedo NJ, Jennings JR, Jenkin JC, Sánchez Alvarado A. 2005. SMEDWI-2 is a PIWI-like protein that regulates planarian stem cells. *Science* **310**: 1327–30.

Rink JC. 2013. Stem cell systems and regeneration in planaria. *Dev Genes Evol* **223**: 67–84.

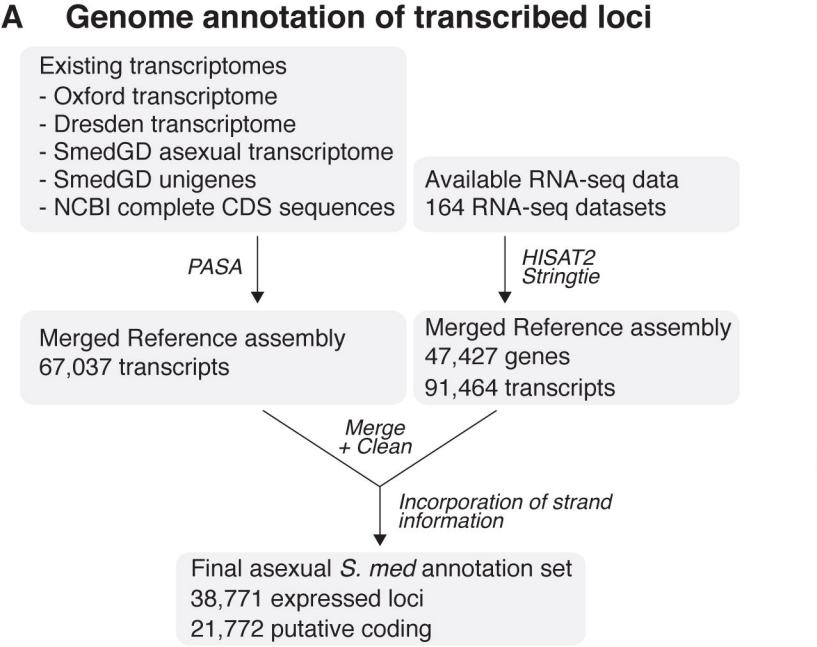
Robb SMC, Ross E, Alvarado AS. 2008. SmedGD: The Schmidtea mediterranea genome database. *Nucleic Acids Res* **36**.

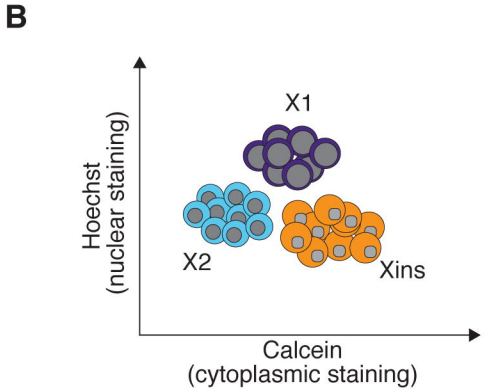
Romero BT, Evans DJ, Aboobaker AA. 2012. FACS analysis of the planarian stem cell compartment as a tool to understand regenerative mechanisms. *Methods Mol Biol* **916**: 167–179.

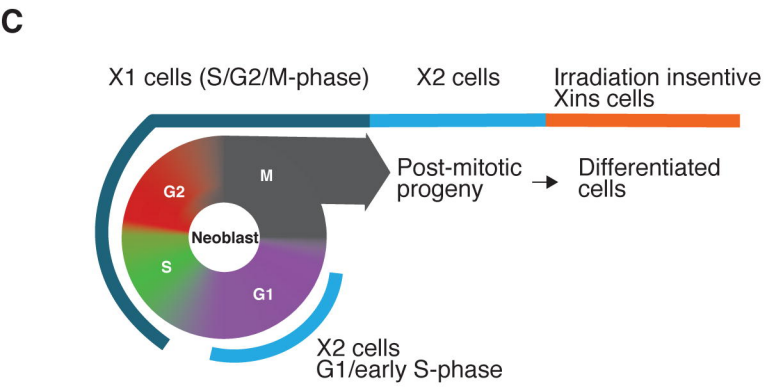
921 Sachs M, Onodera C, Blaschke K, Ebata K, Song J, Ramalho-Santos M. 2013.
922 Bivalent Chromatin Marks Developmental Regulatory Genes in the Mouse
923 Embryonic Germline InVivo. *Cell Rep* **3**: 1777–1784.
924 Schuettengruber B, Ganapathi M, Leblanc B, Portoso M, Jaschek R, Tolhuis B, Van
925 Lohuizen M, Tanay A, Cavalli G. 2009. Functional anatomy of polycomb and
926 trithorax chromatin landscapes in Drosophila embryos. *PLoS Biol* **7**.
927 Schwaiger M, Schönauer A, Rendeiro AF, Pribitzer C, Schauer A, Gilles AF, Schinko
928 JB, Renfer E, Fredman D, Technau U. 2014. Evolutionary conservation of the
929 eumetazoan gene regulatory landscape. *Genome Res* **24**: 639–650.
930 Scimone ML, Kravarik KM, Lapan SW, Reddien PW. 2014. Neoblast specialization in
931 regeneration of the planarian schmidtea mediterranea. *Stem Cell Reports* **3**:
932 339–352.
933 Scimone ML, Meisel J, Reddien PW. 2010. The Mi-2-like Smed-CHD4 gene is
934 required for stem cell differentiation in the planarian Schmidtea mediterranea.
935 *Development* **137**: 1231–1241.
936 Sebé-Pedrós A, Ballaré C, Parra-Acero H, Chiva C, Tena JJ, Sabidó E, Gómez-
937 Skarmeta JL, Di Croce L, Ruiz-Trillo I. 2016. The Dynamic Regulatory Genome
938 of Capsaspora and the Origin of Animal Multicellularity. *Cell* **165**: 1224–1237.
939 Shibata N, Kashima M, Ishiko T, Nishimura O, Rouhana L, Misaki K, Yonemura S,
940 Saito K, Siomi H, Siomi MC, et al. 2016. Inheritance of a Nuclear PIWI from
941 Pluripotent Stem Cells by Somatic Descendants Ensures Differentiation by
942 Silencing Transposons in Planarian. *Dev Cell* **37**: 226–237.
943 Sirén J, Välimäki N, Mäkinen V. 2014. HISAT2 - Fast and sensitive alignment
944 against general human population. *IEEE/ACM Trans Comput Biol Bioinforma*
945 **11**: 375–388.

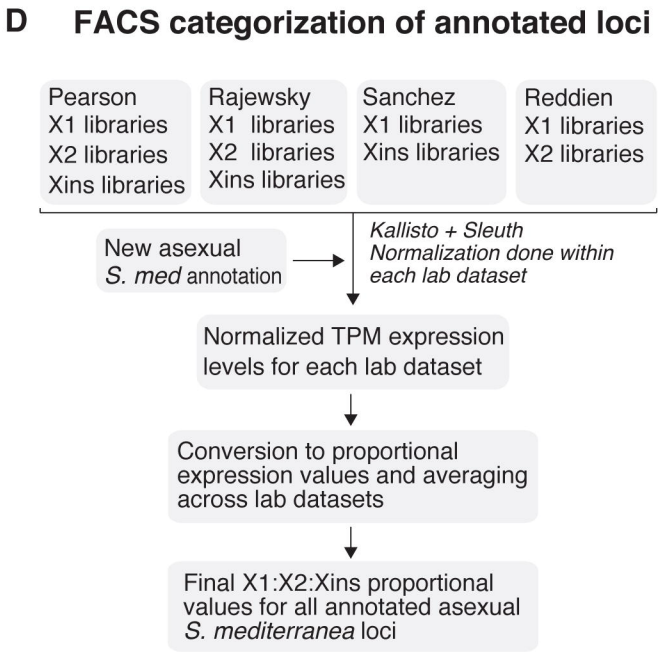
946 Solana J. 2013. Closing the circle of germline and stem cells: the Primordial Stem
947 Cell hypothesis. *Evodevo* **4**: 2.
948 Solana J, Irimia M, Ayoub S, Orejuela MR, Zywitza V, Jens M, Tapial J, Ray D,
949 Morris Q, Hughes TR, et al. 2016. Conserved functional antagonism of CELF
950 and MBNL proteins controls stem cell-specific alternative splicing in planarians.
951 *Elife* **5**.
952 Solana J, Kao D, Mihaylova Y, Jaber-Hijazi F, Malla S, Wilson R, Aboobaker A.
953 2012. Defining the molecular profile of planarian pluripotent stem cells using a
954 combinatorial RNAseq, RNA interference and irradiation approach. *Genome Biol*
955 **13**: R19.
956 Stock JK, Giadrossi S, Casanova M, Brookes E, Vidal M, Koseki H, Brockdorff N,
957 Fisher AG, Pombo A. 2007. Ring1-mediated ubiquitination of H2A restrains
958 poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat Cell Biol* **9**:
959 1428–1435.
960 Van Wolfswinkel JC, Wagner DE, Reddien PW. 2014. Single-cell analysis reveals
961 functionally distinct classes within the planarian stem cell compartment. *Cell*
962 *Stem Cell* **15**: 326–339.
963 Vásquez-Doorman C, Petersen CP. 2016. The NuRD complex component p66
964 suppresses photoreceptor neuron regeneration in planarians. *Regen (Oxford,*
965 *England)* **3**: 168–78.
966 Vastenhouw NL, Zhang Y, Woods IG, Imam F, Regev A, Liu XS, Rinn J, Schier AF.
967 2010. Chromatin signature of embryonic pluripotency is established during
968 genome activation. *Nature* **464**: 922–926.
969 Voigt P, Tee WW, Reinberg D. 2013. A double take on bivalent promoters. *Genes*
970 *Dev* **27**: 1318–1338.

971 Wagner EJ, Carpenter PB. 2012. Understanding the language of Lys36 methylation
972 at histone H3. *Nat Rev Mol Cell Biol* **13**: 115–126.
973 Wu SF, Zhang H, Cairns BR. 2011. Genes for embryo development are packaged in
974 blocks of multivalent chromatin in zebrafish sperm. *Genome Res* **21**: 578–589.
975 Wu TD, Watanabe CK. 2005. GMAP: A genomic mapping and alignment program for
976 mRNA and EST sequences. *Bioinformatics* **21**: 1859–1875.
977 Wurtzel O, Cote LE, Poirier A, Satija R, Regev A, Reddien PW. 2015. A Generic and
978 Cell-Type-Specific Wound Response Precedes Regeneration in Planarians. *Dev*
979 *Cell* **35**: 632–645.
980 Yamaguchi S, Hong K, Liu R, Inoue A, Shen L, Zhang K, Zhang Y. 2013. Dynamics
981 of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell
982 reprogramming. *Cell Res* **23**: 329–339.
983 Zhu SJ, Hallows SE, Currie KW, Xu C, Pearson BJ. 2015. A mex3 homolog is
984 required for differentiation during planarian stem cell lineage development. *Elife*
985 **4**: 1–23.
986



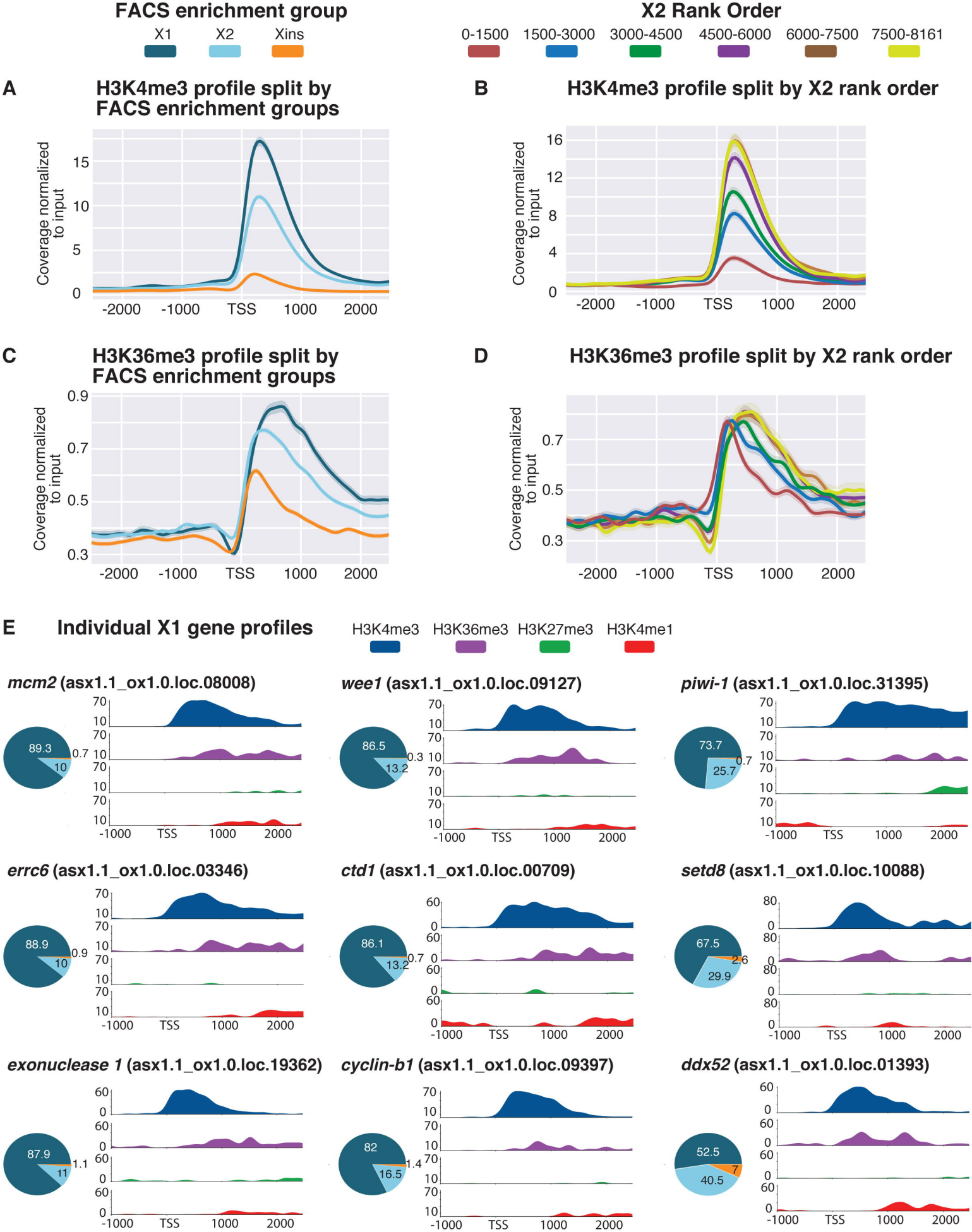


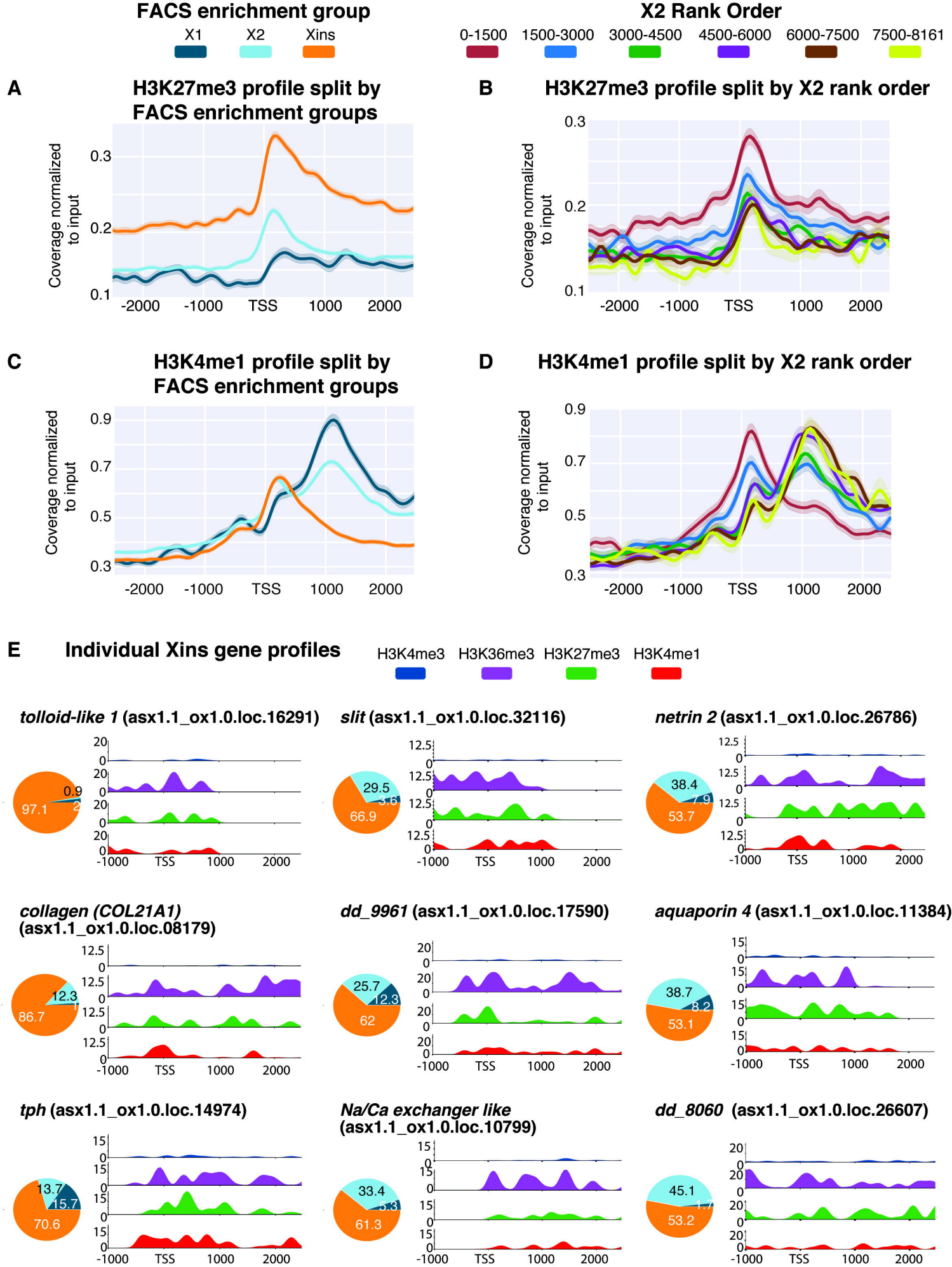


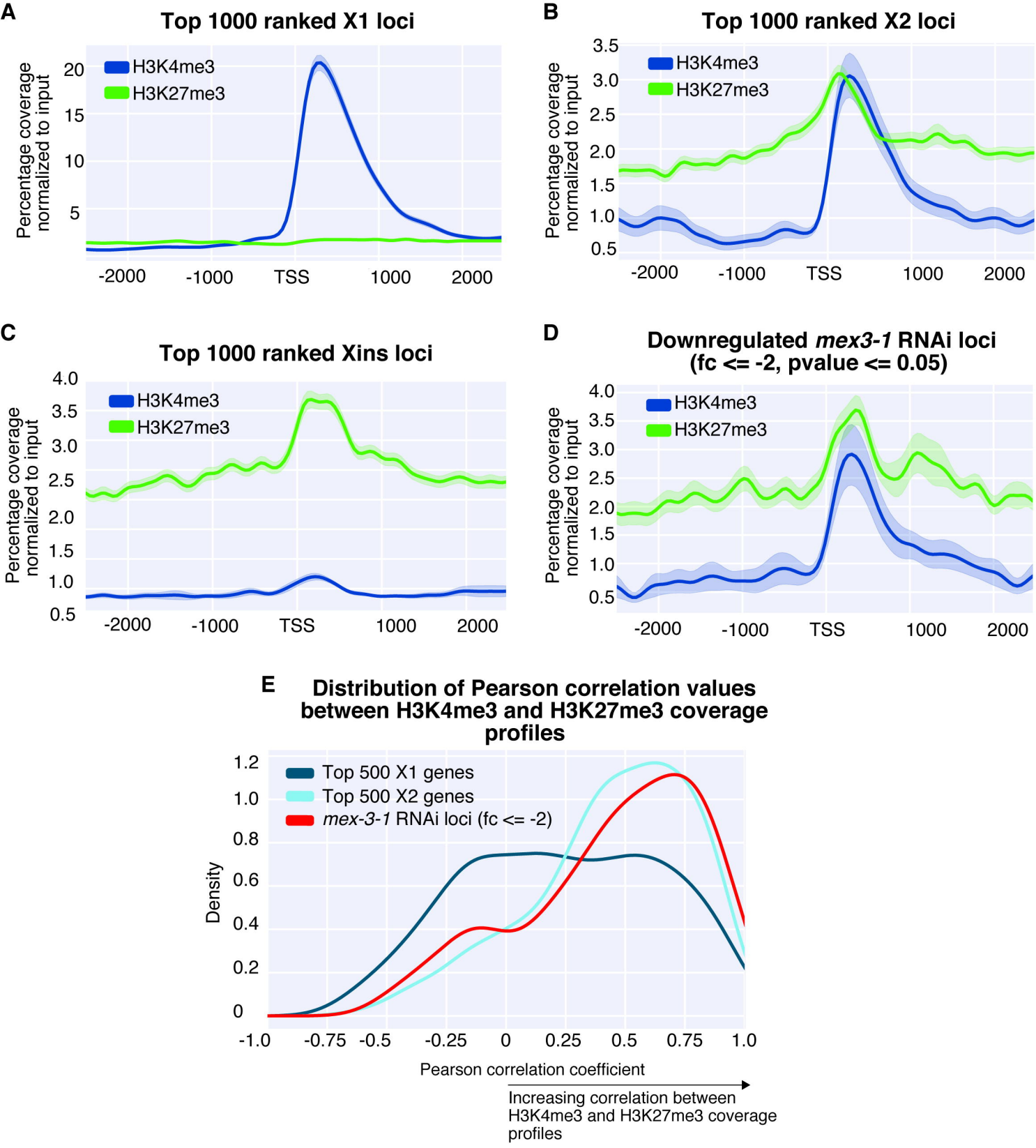


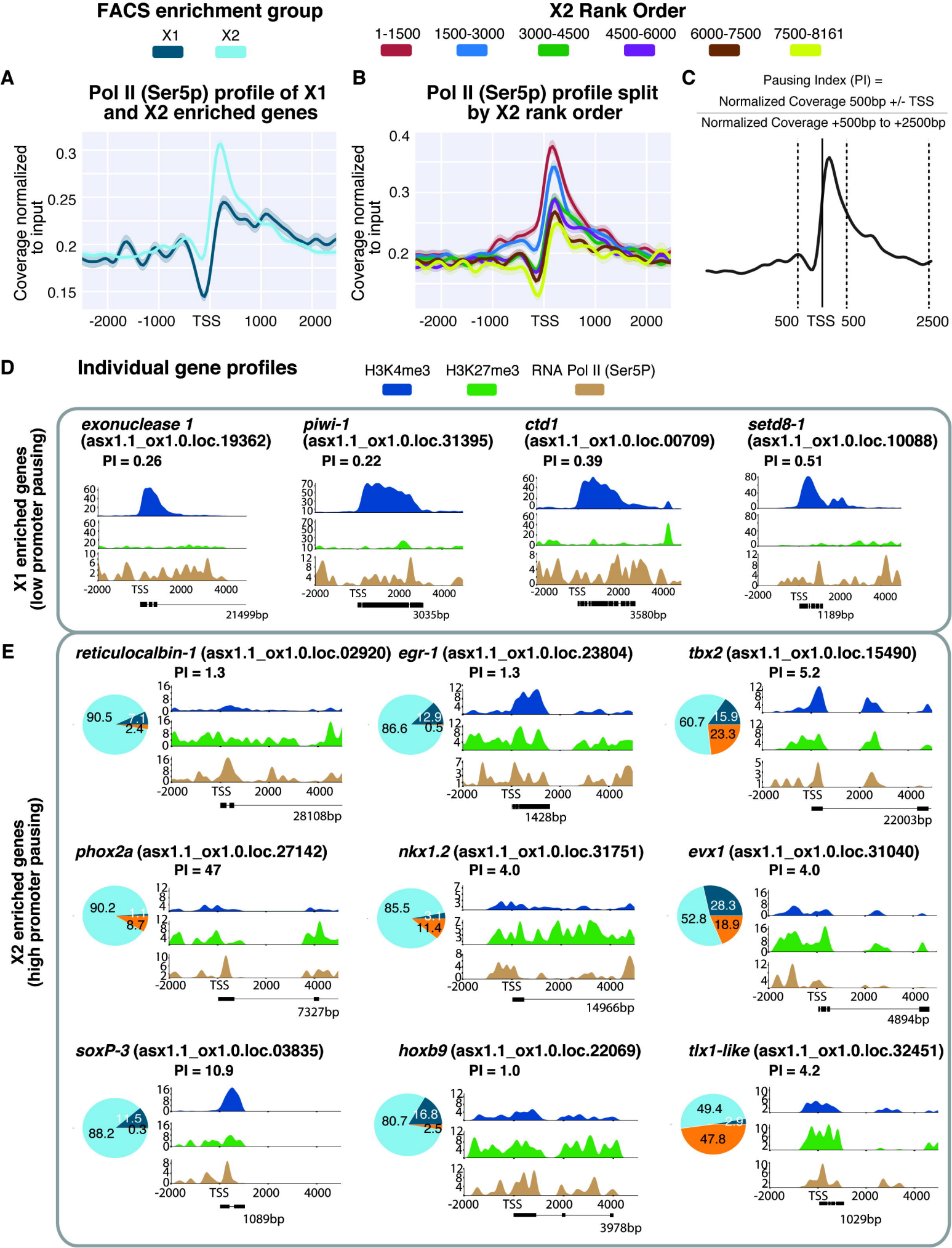
E FACS categorization of 38,711 total loci

Category	Criteria	Loci	Coding loci (% of category)
X1 enriched	X1 proportional expression =>50%	2,253	1,544 (68%)
X2 enriched	X2 proportional expression =>50%	8,444	4,781 (57%)
Xins enriched	Xins proportional expression =>50%	5,119	3,877 (76%)
X1/X2 enriched	X1 + X2 proportional expression =>75% Neither enriched in X1 nor X2	4,538	3,107 (68%)
X2/Xins enriched	X2 + Xins proportional expression =>75% Neither enriched in X2 nor Xins	3,652	2,688 (74%)
X1/Xins enriched	X1 + Xins proportional expression =>75% Neither enriched in X2 nor Xins	303	0 (0%)
Ubiquitous	Loci with roughly equal proportion in X1, X2, and Xins	2,897	2,003 (69%)
Unclassified	Loci with <10 reads in all FACS RNA-seq libraries	11,565	3,762 (33%)


FACS enrichment group
X1
X2
Xins
X2 Rank Order
0-1500
1500-3000
3000-4500
4500-6000
6000-7500
7500-8161
A
H3K4me3 profile split by FACS enrichment groups
B
H3K4me3 profile split by X2 rank order
C
H3K36me3 profile split by FACS enrichment groups
D
H3K36me3 profile split by X2 rank order
Coverage normalized to input
-2000
-1000
TSS
1000
2000
E
Individual X1 gene profiles
H3K4me3
H3K36me3
H3K27me3
H3K4me1
mcm2 (asx1.1_ox1.0.loc.08008)
89.3
10
0.7
wee1 (asx1.1_ox1.0.loc.09127)
86.5
13.2
0.3
piwi-1 (asx1.1_ox1.0.loc.31395)
73.7
25.7
0.7
errc6 (asx1.1_ox1.0.loc.03346)
88.9
10
0.9
ctd1 (asx1.1_ox1.0.loc.00709)
86.1
13.2
0.7
setd8 (asx1.1_ox1.0.loc.10088)
67.5
29.9
2.6
exonuclease 1 (asx1.1_ox1.0.loc.19362)
87.9
11
1.1
cyclin-b1 (asx1.1_ox1.0.loc.09397)
82
16.5
1.4
ddx52 (asx1.1_ox1.0.loc.01393)
52.5
40.5
7


FACS enrichment group
X1
X2
Xins
X2 Rank Order
0-1500
1500-3000
3000-4500
4500-6000
6000-7500
7500-8161
A
H3K27me3 profile split by FACS enrichment groups
B
H3K27me3 profile split by X2 rank order
C
H3K4me1 profile split by FACS enrichment groups
D
H3K4me1 profile split by X2 rank order
Coverage normalized to input
-2000
-1000
TSS
1000
2000
E
Individual Xins gene profiles
H3K4me3
H3K36me3
H3K27me3
H3K4me1
tolloid-like 1 (asx1.1_ox1.0.loc.16291)
slit (asx1.1_ox1.0.loc.32116)
netrin 2 (asx1.1_ox1.0.loc.26786)
collagen (COL21A1) (asx1.1_ox1.0.loc.08179)
dd_9961 (asx1.1_ox1.0.loc.17590)
aquaporin 4 (asx1.1_ox1.0.loc.11384)
tph (asx1.1_ox1.0.loc.14974)
Na/Ca exchanger like (asx1.1_ox1.0.loc.10799)
dd_8060 (asx1.1_ox1.0.loc.26607)


A
Top 1000 ranked X1 loci
H3K4me3
H3K27me3
Percentage coverage normalized to input
20
15
10
5
-2000
-1000
TSS
1000
2000
B
Top 1000 ranked X2 loci
H3K4me3
H3K27me3
Percentage coverage normalized to input
3.5
3.0
2.5
2.0
1.5
1.0
0.5
-2000
-1000
TSS
1000
2000
C
Top 1000 ranked Xins loci
H3K4me3
H3K27me3
Percentage coverage normalized to input
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
-2000
-1000
TSS
1000
2000
D
Downregulated mex3-1 RNAi loci
(fc <= -2, pvalue <= 0.05)
H3K4me3
H3K27me3
Percentage coverage normalized to input
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
-2000
-1000
TSS
1000
2000
E
Distribution of Pearson correlation values
between H3K4me3 and H3K27me3 coverage
profiles
Top 500 X1 genes
Top 500 X2 genes
mex-3-1 RNAi loci (fc <= -2)
Density
1.2
1.0
0.8
0.6
0.4
0.2
0
-1.0
-0.75
-0.5
-0.25
0
0.25
0.5
0.75
1.0
Pearson correlation coefficient
Increasing correlation between
H3K4me3 and H3K27me3 coverage
profiles


FACS enrichment group
X1
X2
X2 Rank Order
1-1500
1500-3000
3000-4500
4500-6000
6000-7500
7500-8161
A
Pol II (Ser5p) profile of X1 and X2 enriched genes
Coverage normalized to input
0.3
0.25
0.2
0.15
-2000
-1000
TSS
1000
2000
B
Pol II (Ser5p) profile split by X2 rank order
Coverage normalized to input
0.4
0.3
0.2
-2000
-1000
TSS
1000
2000
C
Pausing Index (PI) =
Normalized Coverage 500bp +/- TSS
Normalized Coverage +500bp to +2500bp
500
TSS
500
2500
D
Individual gene profiles
H3K4me3
H3K27me3
RNA Pol II (Ser5P)
X1 enriched genes (low promoter pausing)
exonuclease 1 (asx1.1_ox1.0.loc.19362)
PI = 0.26
21499bp
piwi-1 (asx1.1_ox1.0.loc.31395)
PI = 0.22
3035bp
ctd1 (asx1.1_ox1.0.loc.00709)
PI = 0.39
3580bp
setd8-1 (asx1.1_ox1.0.loc.10088)
PI = 0.51
1189bp
E
X2 enriched genes (high promoter pausing)
reticulocalbin-1 (asx1.1_ox1.0.loc.02920)
PI = 1.3
90.5
7.1
2.4
28108bp
egr-1 (asx1.1_ox1.0.loc.23804)
PI = 1.3
86.6
12.9
0.5
1428bp
tbx2 (asx1.1_ox1.0.loc.15490)
PI = 5.2
60.7
15.9
23.3
22003bp
phox2a (asx1.1_ox1.0.loc.27142)
PI = 47
90.2
1.1
8.7
7327bp
nkx1.2 (asx1.1_ox1.0.loc.31751)
PI = 4.0
85.5
3.1
11.4
14966bp
evx1 (asx1.1_ox1.0.loc.31040)
PI = 4.0
52.8
28.3
18.9
4894bp
soxP-3 (asx1.1_ox1.0.loc.03835)
PI = 10.9
88.2
11.5
0.3
1089bp
hoxb9 (asx1.1_ox1.0.loc.22069)
PI = 1.0
80.7
16.8
2.5
3978bp
tlx1-like (asx1.1_ox1.0.loc.32451)
PI = 4.2
49.4
2.9
47.8
1029bp