#### **ADDITIONAL FILE 2: Supplemental Figures S1-13**

#### Sex-specific chromatin landscapes in an ultra-compact chordate genome

Pavla Navratilova<sup>1§</sup>, Gemma B. Danks<sup>1§</sup>, Abby Long<sup>2</sup>, Stephen Butcher<sup>2</sup>, J. Robert Manak<sup>2</sup> and Eric M. Thompson<sup>1,3\*</sup>

# Supplemental Figure S1



**Supplemental Fig. S1.** Specification of the tissues sampled in this study. A) *Oikopleura dioica* in the late day 5 - early day 6 developmental stage that was used for all experiments. Scale bar in  $\mu$ m. B) Gonad maturation chronologies in the ovaries (meiotic nuclei - top, nurse nuclei - bottom) and in the testis. The sexes can be easily separated based on the size of nuclei (larger endocyling nuclei give a rougher appearance in the ovary) as shown by the stereoscope images of day 5-6 ovary and testis sections. The stages that were sampled are indicated by a red rectangle on the timeline.



**Supplemental Fig. S2.** Mean  $\log_2$  ChIP to input signals in a 1000 bp window centered around transcription start (TSS) and end (TES) sites for different gene categories. Genes were grouped by expression level (upper quartile of expression versus silent) and further subdivided based on whether or not their transcripts are *trans*-spliced with the spliced leader SL-RNA (A); the GC-content of their promoter regions: high GC = HGP (top 10% of all genes) and low GC = LGP (bottom 10% of all genes) (B); or by their chromosomal location (X, Y, or autosomal) (C,D). Error bars represent 95% confidence intervals on the mean calculated by bootstrapping.



**Supplemental Fig. S3.** Emission parameters from separately learnt models on the ovary and testis samples were compared to those of the jointly learnt model on both tissues combined. States from the separately learnt models as labeled (where O# = state # in the ovary model and T# = state # in the testis model) according to their closest match (lowest Euclidean distance) to the jointly learnt model states (1 to 15). The jointly learnt model captures all states of the separately learnt models shown here (cf. Fig. 2A).



**Supplemental Fig. S4.** Chromatin states in the *O. dioica* ovary and testis obtained by learning 15-state models separately for each data set reveal all states captured by the jointly learnt model and highlight their specificity to each tissue. Heatmaps visualize model emission parameters from the ovary (A) and testis (C) 15-state models (see also Fig. S3). The fold-enrichments at genomic features for separately learnt chromatin state in the ovary (B) and testis (D) are plotted as in Fig. 2. Gene bodies (orange) and promoter regions (purple) are split into active (green) and silent (red). Txn = transcription; TF = transcription factor; ZF = zinc finger protein; HD = homeodomain protein; high and low specificity genes according to breadth of expression across development (see Methods); TAR = transcriptionally active region; TE = transposable element; unannotated = regions lacking annotation or transcription; 5meC = methylcytosine; S5P = serine 5 phosphorylated RNA polymerase II.



**Supplemental Fig. S5.** A more detailed 50-state model for the *O. dioica* ovary and testes epigenomes. A) Heatmap of model emission parameters for 50 chromatin states learned across the genome, using both samples. Proportions of the genome in the ovary and testis covered by each chromatin state are indicated at the top. Chromatin state fold-enrichments for genomic features in the ovary (B) and testis (C), ordered according to clustering by correlation in the ovary. The table gives the numbers of each feature in each sample. Gene bodies (orange) and promoter regions (purple) are split into active (green) and silent (red). Txn = transcription; TF = transcription factor; ZF = zinc finger protein; HDwpred = homeodomain protein; high and low specificity genes according to breadth of expression across development (see Methods); TAR = transcriptionally active region; TE = transposable element; unannotated = regions lacking annotation or transcription; Cher = ChIP-enriched region; 5meC = methylcytosine; S5P = serine 5 phosphorylated RNA polymerase II; UTR5 = 5' UTR.

Supplemental Figure S6



**Supplemental Fig. S6.** Compact chromatin state domains in the *O. dioica* epigenome. Distributions of chromatin state domain widths are shown for each of 15 chromatin states in the testis and ovary in *O. dioica* (A) and in 8 human cell lines (B) [1]. The heterochromatic states (red boxes) are reduced in particular compared to corresponding states in the human epigenome.



**Supplemental Fig. S7**. Gene Ontology (GO) categories related to RNA pol II pausing and H3K27me3 ChIP-enriched regions. GO-terms over-represented in the set of silent genes with paused RNA pol II in the promoter in the ovary (A) and testis (B). GO terms over-represented in the set of genes with H3K27me3 chers in the ovary (C) and (D) testis.



**Supplemental Fig. S8.** Co-occurrence of ChIP-enriched regions among all ChIP samples in both testis and ovary. Labels indicate the sample and the number of the particular ChIP-enriched regions (chers) detected (using all replicates; see Methods) in the ovary and testes, respectively. Color indicates the proportion of overlapping chers between pairs of samples (blue = 0% overlap; yellow = 100% overlap).



**Supplemental Fig. S9.** Enrichment of 15 jointly-learnt chromatin states at HCNEs (highly conserved elements).



**Supplemental Fig. S10.** Distinct chromatin states were observed for large introns and skipped exons. Chromatin state fold-enrichment (converted to z-scores to highlight the relative fold-enrichments for each feature) heatmap for expressed (green) and silent (red) introns and exons in the ovary (pink) and testis (light blue). Introns were separated by length into short (< 50 nt), mid (50 - 100 nt) and long (> 100 nt). Features were clustered by correlation of fold-enrichments across chromatin states (darker blue = higher enrichment) of a particular state. Notably, the chromatin states of large introns cluster as a function of their expression level, irrespective of the sample being from the ovary or the testes.





**Supplemental Fig. S11.** Polycomb domains resolved by using a lower (75%) threshold for cher detection. (A) Genome browser screenshot showing histone PTMs over a 1.5 Mb region. Tracks of Polycomb-related histone PTM signal/input distributions and H3K27me3 chers using 95 (top of paired row) and 75% (bottom of paired row) thresholds were complementary to the H3K36me2 chers (95% threshold). (B) Distribution of cher breadth and (C) Relative coverage of the 75% threshold-detected H3K27me3 domains on autosomal (Acoverage) and X- and Y- chromosomes.

Testis autosome	Testis X-chromosome	Testis Y-chromosome
watering_1 50.000 100.000 200.000 200.000 300.000 300.000 400.000 400.000   Programming No.000	50,001 100,001 250,001 300,001 360,001 400,001   301 302,001 300,001 300,001 300,001 300,001   301 302,001 300,001 300,001 300,001 300,001 300,001   301 301,001 300,001 300,001 300,001 300,001 300,001 300,001   301 301,001 300,001	00.004 150.004 90.004 200.004 200.004 200.004 400.001 400.004 400
Way 3 convertis states 4 1112 2 Mar growth's states Mar growth's to y 2 Mar growth's to y 2	Mar 3 choosen boon Mar pond M, tr 2	Make 55 chronath state Mar growel 16 (p-2
Average and the second se	ar	Note 507 cct
15- March Ma		- dendrik wa muu shiki kikumi ka kiku mu maka sa ku sa mu na ku mu sa kiku sa ku sa sa kiku sa kiku sa mu
A real of a finite behavior of the second se	Man 1990 mar 1 - California and a state and a state and a state of the	na sheati na ina na n
A second se	e presentation de la constante d	and mining an and the last set in the many set of the set
Mare Have	ren kur med eine einer sich sich vollen fellen sich sich die Jahren beite auf sichte die steht	
	the second	
Mar 109276	and the second lite backs and the second second second second second second	
Main HORZTONE 2 - Main HORZTONE - Main HORZTON	Mar (Aldreit) Mar (Aldreit) - An ann an A	Man 1907 թու Լունա 1914 թունի 1914 թ. հային է հայիներին է հայիներին է հայիներին է հայիներին է հայիներին է հայիներին է հայինե
2 - Maio 1027/m3 o âr uit di tit, th thriftight the same ar our tite tite die tot, dat a single rate a tot biologic hair an	Maridiana A an an ann an	Main 1900 รายนี้ 1986 มีนักราวนี้นี้
Nove HOSSEP 2. Make HOSSEP 4. A statistical structure and a st	nan 1930) Kalamati di dan bernari bata sa silina sa dikasar din 1933 ng matalakasa din dika sa da bahada Dai wina akan da	ատություրը։ Աման հայտությունը այն ունելու ու ու ունելու ունելու ունելու հայտությունը ունելու ու ու ունելու ու ու ունելու հա
	an i shi shi disa ataa ku	Man visioner Լունակում է վերադատանակությունը կորությունը այստությունը հետ այս այստությունը։
	tan, J. (a. 1.)., anternet to allow with each links, allo, and a law (M. (	
the COSO	nainneiline (, , innenne in surein anteiline energie (	and it was able and the such any condition of the international sector about the feature base of the second sec
Marioner	na – 1 daala ka waxaa maxaa kaasa ahafaa sharaa daa aalaha ya kacaasa waxaa ka sharaa daa kabaa ka ka kata kata Marii waxaa	
Mar 1999 2	ماسيداني معرف المستعدية الأللية بمعانية بالمحتوية بعلام فسنابك سمعت بالمساهير والتألي	من من من الله الله الله عن المشرق عن من الله عنه من الله الله عنه الله الله عنه الله الله عنه المار عن الله ال الله المحمد الله الله عنه الله الله عنه الله الله عنه الله الله الله الله عنه الله الله عنه الله الله الله الله
Mare HORSINGS	and the second	in 1. It is an administration of the second statements and the statement in the second second second statements
the record of the field of the state of the first and state of the first of the state of the sta	t dame, de hein ble en stor. Et and o ministration of the built of storage date, due basel in built may be an the sto Manual Anna and a storage of the	de de la collection de la
Non-Hellows 	na dhe maada ay aharika ah ahariya na ahariya na ahariya na ahariya na ahariya na ahariya na ahariya ahariya a Maniya ahariya na ahariya	. Mil (L. 1.1) New Astronomical Action in a stability of the first system. The other states in the second
nerice a substantia terrar a substant a la falla f	n Malaka da sa sa sakaka na sana na ka na ka na ka na na na na kawa ka	ninger sonalisis alian oo ar ar ar alian ah ar alian ah aning shirani karana ah in ana alian iyo ar ar an an a
2 departed in Annual Annua	na – Andrei Cheineach, an Anna ann an Anna Anna ann an Anna ann an Anna ann an Anna Anna Anna Anna Anna Anna A Maraineach ann an Anna Anna Anna Anna Anna Anna A	
Pende H3CRes	Person Decement in the second se	
Construction of the second sec	an a	
Fernie HOOme3	Format Following	response 
Fende HDC7nm1	Press internet	Prove 10 <sup>10</sup>
Fernie 19527m3	(and ) with a sublider of a finite state of the sublider of the state	terme to the
renne verser	halan salah mananda taka matilita falim sadalari shakalari mananda a salar sakatari dakalar	
Pennie (1906) er 2 Pennie (1906) er 3 o "af ult "Marzina dotablia skradol och Jahr anderskin der Er 1, stadskiller det Er 2, stadskiller det er anderskiller der som bit seller ber	Fenne 1980au 1 16	Annual
	dale, sulasse e respectaval dour sedant politiciana, sur a asticles, retained de asticles	Processible:
Fender (150m) • La Julia da La	rende 1959/192   hankan juga kardan a sala kar a (kardan alkan kara) a sana juga kardan kardan a sala sala sala sala sala sala	
FenderHeldenni	and and a state of the state of	The second
PenterHotons	Pornal HACORES	
Fands Hilling and the second and the second and se	han an an an an an ann an ann an ann an a	Trees SIG

**Supplemental Fig. S12.** Genome browser screenshots showing histone PTMs over 0.5 Mb regions of sex chromosomes and autosomes together with chromatin state (15-state model) and selected annotation tracks. HL fpr 2 = processed tiling microarray probe intensity signal (expression data). Selected tracks in the sox2 genomic locus show differences in histone PTM distributions and chromatin states between ovary and testis. The sox2 gene with its promoter is boxed in red.



**Supplemental Fig. S13.** Expression value distributions of autosomal and X-chromosomal genes in *O. dioica* ovary and testis indicate a dosage compensatory mechanism.

#### References

1. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473:43–9.