Title

High resolution transcriptional profiling of *Anopheles gambiae* spermatogenesis reveals mechanism of sex chromosome regulation

Authors

Chrysanthi Taxiarchi¹², Nace Kranjc¹, Antonios Kriezis¹, Kyros Kyrou¹, Federica Bernardini¹, Steven Russell³, Tony Nolan¹⁴, Andrea Crisanti¹, Roberto Galizi^{1*}

Supplementary information

Transgenic line	% transgenic ¹	Insertion site ²	Chromosome band
β2mC-1	54.4 (90)	CTTCGGATTTTTAA-PB-TTAAGCGCAAGGTT	UNKN
β2mC-2	46.6 (60)	GAGGATGAAGTTAA-PB-TTAAGTACAGTGCG	UNKN
β2mC-X	53.3 (60)	ΤΑΑΑΑΑΑCCGTTAΑ-ΡΒ-ΤΤΑΑΑΤΑΑΑΑCCAA	X-5D

Supplementary Table S1: Characterisation of β 2mC transgenic lines.

(1) Percentage of transgenic offspring scored in the progeny of the first transgenic founders. The total number of larvae screened is given in parenthesis. (2) 14 base pairs flanking the piggyBac integration site and predicted chromosomal insertion based on current genome assembly.

Supplementary	y Table S2: Pearson correlation coefficient of all the biological replicates.
---------------	---

	l Dom 1	l Dom 2	l Dom 2	ll Dom 1	ll Dom 2	ll Dom 2	III Dom 1	III Dom 2		IV Den 1	IV Dom 2	IV Den 2
-	керт	кер 2	кер з	керт	кер 2	кер з	керт	кер 2	кер з	керт	кер 2	кер з
I Rep 1	1.0000											
l Rep 2	0.9823	1.0000										
I Rep 3	0.9789	0.9548	1.0000									
II Rep 1	0.8488	0.8294	0.7979	1.0000								
II Rep 2	0.7701	0.7447	0.7149	0.9862	1.0000							
II Rep 3	0.7830	0.7468	0.7419	0.9825	0.9920	1.0000						
III Rep 1	0.4599	0.4416	0.4000	0.8305	0.8931	0.8749	1.0000					
III Rep 2	0.4327	0.4147	0.3780	0.8138	0.8800	0.8621	0.9970	1.0000				
III Rep 3	0.4570	0.4359	0.4061	0.8303	0.8930	0.8835	0.9925	0.9950	1.0000			
IV Rep 1	0.3913	0.3669	0.3520	0.7732	0.8381	0.8263	0.9730	0.9816	0.9800	1.0000		
IV Rep 2	0.3783	0.3570	0.3354	0.7674	0.8343	0.8182	0.9753	0.9843	0.9791	0.9983	1.0000	
IV Rep 3	0.3848	0.3626	0.3432	0.7704	0.8376	0.8243	0.9749	0.9839	0.9818	0.9988	0.9986	1.0000

Correlation values between replicates (Rep) of the same sample are indicated in bold.

	Popul	Population I Population II Population III		Population IV				
FPKM threshold	X:A	2:3	X:A	2:3	X:A	2:3	X:A	2:3
0	0.350376	1.117451	0.085872	1.086674	0.025497	1.051632	0.042358	1.01867
	(917/10690)	(6183/4507)	(831/10633)	(6118/4515)	(686/10413)	(5990/4423)	(715/10057)	(5785/4272)
1	0.571388 (576/7587)	1.017861 (4416/3171)	0.20808 (428/7787)	1.005417 (4506/3281)	0.146643 (209/7836)	1.002019 (4525/3311)	0.166951 (255/7531)	1.010136 (4346/3185)
2	0.650212 (518/7142)	1.018933 (4169/2973)	0.272494 (333/7312)	0.983543 (4246/3066)	0.203202 (152/7324)	1.008115 (4229/3095)	0.234816 (183/6920)	0.966899 (4012/2908)
3	0.666549	1.019181	0.326781	0.974556	0.256033	0.969458	0.278578	0.943684
	(487/6816)	(4000/2816)	(283/6979)	(4071/2908)	(114/6911)	(3992/2919)	(148/6512)	(3772/2740)
4	0.705947	0.979023	0.36937	0.967305	0.27543	0.947079	0.307851	0.93863
	(461/6598)	(3872/2726)	(249/6720)	(3924/2796)	(101/6600)	(3829/2771)	(130/6167)	(3574/2593)
10	0.771018	0.945489	0.614083	0.960344	0.553159	0.915979	0.446783	0.926351
	(350/5565)	(3262/2303)	(149/5624)	(3277/2347)	(51/5387)	(3121/2266)	(75/5029)	(2898/2131)
20	0.992577	0.928266	0.765333	0.938786	0.576514	0.908194	0.678284	0.826995
	(263/4387)	(2569/1818)	(99/4506)	(2612/1894)	(31/4329)	(2492/1837)	(44/4108)	(2384/1724)
30	1.191703	0.926871	0.822087	0.916534	0.555921	0.882635	0.62428	0.871947
	(207/3663)	(2145/1518)	(71/3747)	(2167/1580)	(24/3638)	(2101/1537)	(34/3541)	(2050/1491)
40	1.33226	0.970425	0.795713	0.896114	0.920404	0.919766	0.503921	0.933909
	(167/3129)	(1814/1315)	(56/3169)	(1830/1339)	(16/3165)	(1797/1368)	(30/3108)	(1780/1328)

autosomes (X:A) and between autosomes (2:3) as a function of increasing FPKM thresholds.

Numbers in brackets indicate the fraction of genes (above the applied FPKM threshold) of the total

number of annotated genes for each chromosome used to calculate ratios

Supplementary Figure S1: Fluorescent microscopy of $\beta 2mC^+$, YVasG⁺ and $\beta 2mC-2^+/YVasG^+$ transgenic testes.

	Transmitted	mCherry	eGFP	Overlay
β2mC-1			N/A	N/A
β2mC-2			N/A	N/A
β2mC-X	e e		N/A	N/A
YVasG		N/A		N/A
β2mC-2+/YVasG+				

Images of testes dissected from 1-day old adult mosquitoes of each transgenic line. From left to right: transmitted, mCherry, eGFP channels and overlay of mCherry and eGFP fluorescent signal. Scale bars: 100 μm.





Flow cytometry dot plot of FlowJo biexponential scale values of (a) TO-PRO-3 iodide and FSC-A to

exclude dead cells and **(b)** mCherry and eGFP fluorescence intensity of live cells. **(c)** Linear FSC-A and SSC-A dot plots of the four selected populations indicating gates for cell sorting to exclude debris with low FSC-A. FSC-A/SSC-A gating was also applied to capture the majority of spermatocytes, excluding smaller spermatids, in population III; and to capture postmeiotic spermatids, excluding the smaller mature spermatozoa and the larger spermatocytes, in population IV. Indicative brightfield images of sorted cells are also shown for each sorted population. Scale bars: 10 μm.

Supplementary Figure S3: Expression profiles of β2-mCherry, vasa2:eGFP and endogenous *β2 tubulin* and *vasa* genes.



Line graph of vasa-eGFP, AGAP008578, β 2-mCherry and AGAP008622 transcript levels across the four populations.

Supplementary Figure S4: Expression profiles of conserved meiotic genes and *D. melanogaster* orthologs.



Expression levels of **(a)** conserved meiotic genes showing transcriptional peak in meiotic populations and **(b)** orthologues of *Drosophila melanogaster* testis genes previously tested through in situ hybridisation²⁴.



(a) Expression levels of significantly switched transcript isoforms differentially used among populations (adjusted P < 0.05). (b) Examples of population enriched (fold change higher than 2 and adjusted P < 0.05 after Sleuth analysis⁴¹) and differentially used transcript isoforms (adjusted *P*-value below 0.05 after IsoformSwitchAnalyzeR analysis⁴²). Isoform A represents the annotated transcript and B the novel. The predicted functional features are shown at the top of each gene dataset where ticker rectangles indicate the coding sequence of the gene and domains with predicted function are indicated in light brown. Line graphs and bar charts represent respectively the expression profile and the usage across the four populations for each of the transcript isoform. Only genes that showed expression above 1 FPKM in at least one of the cell populations were included in the de novo transcriptome analysis.

differentially expressed gene clusters.

а Cytoskeleton organisation 0.6 0.4 0.2 z-score 0.0 -0.2 -0.4 -0.6 -ii - iii ī actin filament organization GO:0007015 actin cytoskeleton organization GO:0030036 actin filament-based process GO:0030029 RNA processing and RNA catabolism 2 z-score 0 -1 -2 1İI İV Ű nuclear-transcribed mRNA catabolic process GO:0000956 mRNA catabolic process GO:0006402 RNA catabolic process GO:0006401 rRNA 3'-end processing GO:0031125 ncRNA 3'-end processing GO:0043628 RNA 3'-end processing GO:0031123 RNA splicing GO:0008380 + С Gene expression and protein maturation 1.5 10 9.5 0.0 2 2 0.5 0.5 -1.0 -1.5 -Ú - IİI İŻ RNA phosphodiester bond hydrolysis, endonucleolytic GO:0090502 protein maturation GO:0051604 Mitochondrial transport 2 1 z-score 0 ıi. лiг. iv mitochondrial transport GO:0006839 mitochondrial electron transport NADH to ubiquinone GO:0006120 intracellular protein transmembrane transport GO:0065002



- translation GO:0006412
- peptide biosynthetic process GO:0043043
- - amide biosynthetic process GO:0043604
 - peptide metabolic process GO:0006518

rRNA processing/modification



- maturation of SSU-rRNA from
- (SSU-rRNA, 5.8S rRNA, LSU-rRNA) GO:0000462
- maturation of SSU-rRNA GO:0030490
- maturation of LSU-rRNA GO:0000470
 maturation of 5.8S rRNA GO:0000460

Transmembrane transport

rRNA modification GO:0000154



Protein folding and

- regulation of protein catabolic process GO:0042176
- chaperone-mediated protein folding
 GO:0061077
 - Ribonucleoprotein complex



- 1İ ТŪ īV ribosome biogenesis GO:0042254
- ribonucleoprotein complex biogenesis GO:0022613 ribosomal large subunit assembly GO:000027
- GO:000027
 ribosomal large subunit biogenesis
 GO:0042273
 ribosome assembly GO:0042255
 ribonucleoprotein complex assembly
 GO:0022618
- ribonucleoprotein complex subunit organization GO:0071826 ribosomal small subunit assembly GO:0000028
- .
- ribosomal small subunit biogenesis GO:0042274

Cellular localisation



ribonucleoprotein complex export from nucleus GO:0071426

ribonucleoprotein complex localization GO:0071166

ATP metabolism 2

- -2 ıi. - iii ī Î.

- mitochondrial electron transport
 cytochrome c to oxygen GO:0006123
 mitochondrial ATP synthesis coupled electron transport GO:0042775
 oxidative phosphorylation GO:0006119
- ATP synthesis coupled electron transport GO:0042773
- ATP metabolic process GO:0046034
- ATP synthesis coupled proton transport GO:0015986
- ATP biosynthetic process GO:0006754
- ATP hydrolysis coupled transmembrane transport GO:0090662
- ATP hydrolysis coupled cation transmembrane transport GO:0099132
- ATP hydrolysis coupled ion transmembrane transport GO:0099131 ATP hydrolysis coupled proton transport GO:0015991
- mitochondrial electron transport ubiquinol to cytochrome c GO:0006122



- positive regulation of cellular component biogenesis GO:0044089 positive regulation of protein complex
- assembly GO:0031334 regulation of cellular component biogenesis GO:0044087

b



Cellular respiration



- aerobic electron transport chain GO:0019646 aerobic respiration GO:0009060
- respiratory electron transport chain GO:0022904
- cellular respiration GO:0045333
 tricarboxylic acid cycle GO:0006099 cellular respiration GO:0045333

Cellular metabolism 1.5



- -
- purine ribonucleoside triphosphate metabolic process GO:0009205 purine nucleoside triphosphate metabolic process GO:0009144
- energy derivation by oxidation of organic compounds GO:0015980 ribonucleoside triphosphate metabolic process GO:0009199
- . electron transport chain GO:0022900 -
- generation of precursor metabolites and energy GO:0006091 nucleoside triphosphate metabolic process GO:0009141
- purine nucleoside triphosphate biosynthetic process GO:0009145 purine ribonucleoside triphosphate biosynthetic process GO:0009206

- biosynthetic process GO:0009206 ribonucleoside triphosphate biosynthetic process GO:0009201 nucleoside triphosphate biosynthetic process GO:0009142 cytochrome complex assembly GO:0017004
- citrate metabolic process GO:0006101 tricarboxylic acid metabolic process GO:0072350
- antibiotic metabolic process
 GO:0016999



- energy coupled proton transmembrane transport, against electrochemical gradient GO:0015988

Mitochondrion organisation



- protein localization to mitochondrion GO:0070585

energy coupled proton transport, down electrochemical gradient GO:0015985



2.0

1.5

1.0

0.5







- establishment of protein localization to mitochondrion GO:0072655
- mitochondrion organization GO:0007005 -
- protein targeting GO:0006605
- 1 0 -1

Average Z-score of genes associated to GO terms significantly enriched in each k-means cluster shown in **Fig. 3a**. **(a)** GO groups significantly enriched in the premeiotic clusters 1, 2 and 3. **(b)** GO term significantly enriched in the postmeiotic cluster 8. **(c)** GO groups significantly enriched in the postmeiotic clusters 9 and 10. Significance of the statistical overrepresentation test was analysed by applying Fisher's Exact test with False Discovery Rate correction.

Supplementary Figure S7: Germline expression dynamics of overrepresented biological processes in

testes-enriched genes.



Average Z-score of genes associated to GO terms significantly overrepresented among testes-enriched genes based on tissue enrichment calculation (tau-value \geq 0.8 and maximum expression in testis) and false-colour k-means clustering heatmap summarizing the expression patterns (as Z-score of log2(FPKM+1)) of genes included in each GO term.

metabolic pathways.



Average Z-score of genes associated to main KEGG metabolic pathways of relevance for male gametogenesis and false-colour k-means clustering heatmap summarizing the expression patterns (as Z-score of log2(FPKM+1)) of genes included in each KEGG pathway.