Cell-type-specific genomics reveals histone modification dynamics in mammalian meiosis

Authors

Kwan-Wood Gabriel Lam, Kevin Brick, Gang Cheng, Florencia Pratto and R. Daniel Camerini-Otero*

Affiliations

Genetics and Biochemistry Branch, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, USA

*Corresponding author

Abstract

Meiosis is the specialized cell division during which parental genomes recombine to create genotypically unique gametes. Despite its importance, mammalian meiosis cannot be studied *in vitro,* greatly limiting mechanistic studies. *In vivo,* meiocytes progress asynchronously through meiosis and therefore the study of specific stages of meiosis is a challenge. Here, we describe a method for isolating pure sub-populations of nuclei that allows for detailed study of meiotic sub-stages. Interrogating the H3K4me3 landscape revealed dynamic chromatin transitions between sub-stages of meiotic prophase I, both at sites of genetic recombination and at gene promoters. We also leveraged this method to perform the first comprehensive, genome-wide survey of histone marks in meiotic prophase, revealing a heretofore unappreciated complexity of the epigenetic landscape at meiotic recombination hotspots. Ultimately, this study presents a straightforward, scalable framework for interrogating the complexities of mammalian meiosis.

Main

Meiosis is the specialized cell division required to form gametes. In meiosis the DNA of a parent cell is replicated, then haploid gametes are produced through two successive cell divisions. Distinct mechanisms are required to assure that chromosomes segregate accurately in each division and these events take place through a staged cascade of nuclear events. Unique to meiosis is the formation of programmed DNA double-strand breaks (DSBs) that result in recombination between parental haplotypes and whose repair tethers homologous chromosomes at the first meiotic division. Meiotic DSB formation, repair and recombination occur in a continuum of sub-stages (leptonema, zygonema, pachynema, and diplonema) known collectively as meiotic prophase I (MPI), and each stage is characterized by specific nuclear events[1]. Despite its importance, technical challenges preclude the isolation and study of specific meiotic sub-stages and many of the molecular mechanisms of meiosis remain poorly studied in higher eukaryotes. This is largely because mammalian meiosis cannot be adequately recapitulated *in vitro*[2]. *In vivo*, meiosis either occurs in a difficult-to-isolate niche, such as the fetal ovary, or in an asynchronously dividing tissue such as in adult testis.

Extant methods to study sub-stages of meiotic prophase have been dominated by the use of crude enrichment strategies. One common approach has been to study spermatocytes in juvenile mice, where the first wave of meiosis progresses relatively synchronously[3,4]. Nonetheless, most cells in juvenile testis are not at the stage of interest because this synchrony is not absolute, and there are many non-meiotic cells. Furthermore, differences between the first and the subsequent rounds of spermatogenesis[5–7] may preclude generalization to adults. Chemically induced synchronization of meiosis has recently been developed[8], however the application is technically cumbersome and the consequences of chemical treatment have not been explored. An alternative to enrichment is the isolation of meiotic sub-stages. Fluorescence-activated cell sorting (FACS) or sedimentation based methods have been used to isolate particular sub-stages[9–11]. However,

since these methods rely on imprecise metrics such as cell size and chromatin content, the isolation and distinction between many sub-stages is challenging. Thus, isolation-based approaches have not been broadly adapted to the study of the dynamic molecular events during meiosis.

We have devised a straightforward, yet highly specific strategy to isolate pure sub-populations of meiotic nuclei. This approach uses a variant of FACS[12] in which antibodies to intra-nuclear markers are used to select nuclei of interest. The proteins present in different meiotic sub-stages are well studied[13–16] and we leverage this knowledge to isolate nuclei in which specific combinations of proteins are expressed. This allowed for the isolation of meiotic sub-populations ranging from 74-96% purity. Nuclei are fixed before sorting, which preserves *in vivo* interactions and prevents degradation during sample processing. Importantly, this fixation also means our method is the first to allow for downstream interrogation of genome-wide protein-DNA interactions in meiotic sub-stages.

Although comprehensively studied in somatic cells, little is known about histone modification dynamics in meiotic cells. A case in point are the histone modifications that precede meiotic DSB formation, an early event in meiotic recombination. Programmed DSBs are targeted to sites defined by DNA binding and subsequent H3K4/K36-trimethylation by the PRDM9 protein[17,18]. PRDM9 gene expression[19], nuclear protein levels of PRDM9[20] and microscopy-based estimates of DSB formation[21,22] all imply that PRDM9 marks sites for DSB formation at the onset of MPI.

Using chromatin immunoprecipitation followed by high throughput sequencing (ChIP-Seq) on sorted populations of nuclei, we determined that Histone H3 Lysine 4 trimethylation (H3K4me3) varies from stage-to-stage at meiotic DSB hotspots, gene promoters and a vast number of heretofore unannotated sites. These are the first genome-wide analyses of the dynamic chromatin landscape in meiosis. Also, we found that the histone code at DSB hotspots is more complex than has been

96 previously appreciated[17,18], with multiple histone tail modifications exhibiting
97 enrichment. This includes histone acetylation marks unlikely to have been
98 deposited by the DSB-defining histone methyltransferase PRDM9.
99
100 Together, these data demonstrate the simplicity, utility and feasibility of our
101 approach. This method opens the door to a far better understanding of the
102 molecular mechanisms underlying meiosis and other complex developmental
103 processes.
104

105 ## Results

106 **Isolation of stage-specific spermatocyte nuclei**

107 We first devised a discriminative panel of intra-nuclear proteins to distinguish
108 between five, classically-defined MPI sub-stages; leptonema, zygonema, early
109 pachynema, late pachynema and diplonema (Fig. 1). Nuclei were extracted from
110 formaldehyde-fixed testes from adult male mice, then incubated with antibodies
111 against the desired proteins (see methods). Nuclei were also stained with DAPI to
112 allow measurement of DNA content. The combinatorial signals from these
113 fluorescent markers were subsequently used to isolate specific sub-populations of
114 spermatocyte nuclei using FACS (Fig. 2a and b; see methods).
115
116 We used the meiosis-specific protein SCP3 to distinguish post-replicative meiotic
117 prophase nuclei (4C) from fully replicated mitotic nuclei (also 4C). We then
118 combined SCP3 with other markers for FACS. Our combination of antibodies
119 required two separate sorts; one to isolate leptotene nuclei and a second to isolate
120 the other four sub-stages (Fig. 2a and b). In male mice, STRA8 regulates the entry
121 to meiotic S phase, and is expressed from primitive spermatogonia to leptotene
122 spermatocytes[16]. We therefore used STRA8 to isolate leptotene nuclei from later
123 stages. To isolate the four other meiotic stages, we used the presence or absence
124 of H1t and quantitative differences in SCP1 intensity (Fig. 1, Fig. 2a). H1t (testis-
125 specific histone H1) is present from mid pachynema through MPI[23], whereas

126 SCP1 is a marker of progressive chromosome synapsis from zygotene to
127 pachytene stage and the SCP1 signal is diminished in diplonema when
128 chromosomes desynapse[14] (Fig. 1). Thus, we collected seven populations,
129 assessed the purity of each population and determined the gating thresholds for
130 the four meiotic sub-stages (Fig. 2b and c).
131
132 The purity of each sorted population was assessed using immunofluorescence
133 microscopy to stage the nuclei according to the known localization of markers (Fig.
134 1, Fig. 2a and c, Supplementary Fig. 1, see methods)[13–16]. The purity of the sorted
135 leptotene population was 94.6% - 97.9% (n = 2), a >50-fold increase in purity
136 compared to the percentage of leptotene cells in whole testis (1.6% - 1.8% (n = 2),
137 Fig. 2c). The purity of the other four populations ranged from 73.9 - 90.4%, each
138 more than 20-fold enriched relative to the starting population from whole testis (Fig.
139 2c). The least pure sample was that of zygotene (71.3% - 78.7% (n = 3)), likely
140 because we rely entirely on quantitative differences in SCP1 to distinguish
141 between adjacent stages (Fig. 1, Fig. 2b and c). Narrowing of the gate can increase
142 purity but at the cost of yield (data not shown). Quantitative thresholding issues
143 also explain the 9.0% - 16.9% (n = 3) of zygotene nuclei found in the sorted early
144 pachytene population (Fig. 1, Fig. 2b and c).
145
146 From a single adult mouse, we obtained between 115,000 and 400,000 nuclei for
147 each meiotic sub-stage (Fig. 2d). This range broadly reflects differences in the
148 proportion of each sub-stage in the adult testis, where leptotene and zygotene
149 nuclei are less abundant. Thus, the different sub-stages are isolated in sufficient
150 numbers for experiments using standard genomics protocols[24].
151
152 **Stage-specific ChIP-Seq reveals histone modification dynamics in meiosis**
153 We next used ChIP-Seq to examine the dynamics of H3K4me3 in MPI. H3K4me3
154 is of key biological interest in meiosis because it regulates multiple independent
155 dynamic processes during MPI; H3K4me3 is found at gene promoters and its
156 presence correlates with active transcription[25,26], whereas PRDM9-mediated

157 H3K4me3 marks the future sites of meiotic DSBs genome-wide[27,28]. We therefore
158 performed H3K4me3 ChIP-Seq in the five aforementioned sorted populations to
159 assess the temporal dynamics of this histone mark in MPI.
160
161 We performed ChIP-Seq for H3K4me3 using a highly-specific antibody
162 (EpiCypher, 13-0028) (Supplementary Fig. 2). To allow for quantitative cross-
163 comparison of H3K4me3 signals from samples of differing quality, we used a
164 spike-in control containing H3K4me3-modified nucleosomes[29] (see methods). We
165 observed a highly dynamic H3K4me3 signal at hotspots in MPI; the H3K4me3
166 signal first appears at leptonema, and is maximal at zygonema, and gone by early
167 pachynema (Fig. 3a and b). The maximal signal at zygonema was somewhat
168 surprising because DSBs are formed and nuclear PRDM9 protein levels are
169 maximal at leptonema[20]. Intriguingly, at strong hotspots, H3K4me3 is relatively
170 weak at zygonema compared to leptonema (Supplementary Fig. 3). This implies
171 that DSB formation or repair removes H3K4me3-modified nucleosomes, similar to
172 findings in *Saccharomyces cerevisae*[30].
173
174 Most H3K4me3 ChIP-Seq studies, including all studies of H3K4me3 in meiosis
175 have used antibodies that extensively cross-react with H3K4me2[31]
176 (Supplementary Fig. 2). To allow cross-comparison with published data we
177 performed ChIP-Seq with the most commonly used antibody (Abcam ab8580; [31])
178 and observed a broadly similar dynamic of histone modification at hotspots as
179 compared to the specific antibody (Supplementary Fig. 4); the hotspot signal is still
180 maximal at zygotene, but the non-specific antibody gives a higher leptotene signal
181 and a residual signal in early pachynema. We hypothesize that these differences
182 reflect dynamic changes in H3K4 dimethylation at hotspots.
183
184 The H3K4me3 level at gene promoters is correlated with gene expression in mitotic
185 cells[26]. We therefore investigated if the temporal dynamics of H3K4me3 at
186 transcription start sites (TSSs) is predictive of meiotic gene expression patterns.
187 Stage-specific gene expression in MPI has been deduced from classical cell

188 sorting of meiotic cells followed by RNA-Seq[11,32]. We found that the H3K4me3
189 profiles at TSSs were positively correlated with gene expression through MPI
190 (Spearman's R = 0.27; see methods). This correlation was highly significant, far
191 higher than the maximum correlation observed in randomized comparisons
192 (empirical $P < 0.0001$; N = 10,000 bootstraps, $R_{max} = 0.03$, see methods). To
193 explore the relationship between H3K4me3 and gene expression in more detail,
194 we clustered transcripts based on the temporal H3K4me3 pattern (Fig. 3c; optimal
195 clusters = 5; k-means clustering using gap statistic[33]). There was significant
196 correlation with gene expression in all clusters ($P < 0.0001$; N = 10,000 bootstraps,
197 range; $0.12 \leq R \leq 0.44$) (Fig. 3c). Each cluster is composed of a large fraction of
198 transcripts where H3K4me3 and gene expression correlate very well
199 (Supplementary Fig. 5). Nonetheless, there are also many transcripts for which
200 H3K4me3 and expression are not correlated. For instance, in cluster 1, many
201 genes have higher gene expression late in MPI than we would predict from
202 H3K4me3 data (far right panels; Supplementary Fig. 5). This may be explained by
203 mRNA accumulation through MPI or by H3K4me3 in early MPI marking poised,
204 but not yet active promoters[34]. In contrast, genes in cluster 5 are less affected by
205 these confounding effects and the correlation is far higher for this cluster. This
206 seems to indicate a concerted wave of gene activation and expression as cells
207 approach the end of MPI. Together, these data demonstrate that ChIP-Seq of
208 H3K4me3 in stage-specific nuclei reveals a dynamic that parallels that of gene
209 expression. These data also reveal an intricate and poorly understood interplay
210 between H3K4me3 and gene expression during MPI.
211
212 **HIstone modification dynamics at other genomic elements**
213 Dynamic H3K4-trimethylation is observed at both gene promoters and DSB
214 hotspots in MPI. However, 16% of H3K4me3 peaks (4,664/28,528) are outside of
215 these sites (Supplementary Fig. 6a). These unannotated H3K4me3 peaks may
216 represent enhancers, cryptic promoters or other functional elements. To explore
217 these dynamics in greater detail, we used the H3K4me2/3 antibody (Abcam
218 ab8580), as this should capture H3K4me2 modifications that mark enhancers.

219 Indeed, using this antibody, 18,738/53,518 (35%) peaks occur at non-hotspot, non-
220 TSS sites (Supplementary Fig. 7a). Using peaks from this antibody, we derived a
221 MPI profile for each peak across the five MPI stages, then performed unbiased k-
222 means clustering. To validate this approach, we examined clustering of DSB
223 hotspots; as expected, hotspots primarily occur in a single large cluster with a
224 maximum signal at zygonema (cluster 3; Supplementary Fig. 7b-d; 89% of
225 hotspots occur in this cluster). 11% of unannotated peaks also occur in this cluster
226 (cluster 3; Supplementary Fig. 7d) and may represent sites of PRDM9 binding that
227 were not detected in hotspot mapping experiments. Alternatively, we cannot rule
228 out that a cluster of cryptic functional sites simply exhibit a similar dynamic. The
229 remaining unannotated peaks exhibit an MPI dynamic that broadly mirrors that at
230 TSS (Supplementary Fig. 7b, d). It seems likely therefore that functional elements
231 at these sites play a role in regulating meiotic progression. These dynamic, but
232 unannotated sites represent a completely unexplored aspect of the regulation of
233 mammalian meiosis.
234
235 Clustering also allowed us to study H3K4me3 dynamics at sites involved in other
236 important processes during MPI. In particular, we examined H3K4me3 sites that
237 are used for DSB targeting in mice that lack functional PRDM9[27,28]. H3K4me3
238 peaks used as "default" hotspot locations were predominantly those with maximal
239 signal in early MPI (Supplementary Fig. 6 and 7).
240
241 **Identifying other histone marks at hotspots**
242 At DSB hotspots defined by PRDM9, H3K4me3 is necessary for DSB formation[35].
243 Nonetheless, H3K4me3 alone is not sufficient to define DSB sites, as this histone
244 mark is also present at other functional sites, such as gene promoters. In mice,
245 H3K36me3 is the only other histone mark described genome-wide at DSB
246 hotspots[17], however other histone marks may also be involved in defining DSB
247 sites[36–38]. We therefore used our method to isolate target nuclei from adult mice
248 for a systematic survey of multiple histone modifications.
249

250 We performed ChIP-Seq using antibodies against 16 additional histone
251 methylation and acetylation marks in isolated nuclei (Fig. 4a and b). These marks
252 were either previously analyzed at individual hotspots in mice[36,38] or reported to be
253 enriched at DSB sites in other organisms[39,40]. In this exploratory phase, we did not
254 perform ChIP-Seq in all five sub-populations, but instead, in a combined
255 population that includes all stages with potential relevance to understanding
256 histone modifications at DSB sites (leptonema to early pachynema; SCP3+ and
257 H1t– population).
258
259 H4K8ac, H4K12ac, H4K20me3, H3K4ac, H3K79me1, H3K79me3, H3K27me1,
260 H3K9me2, H3K9me3 and H3K27me3 were not enriched at hotspots relative to
261 controls (Fig. 4a, b; Supplementary Fig. 8). With the exceptions of H3K27me1, and
262 H3K9me2, these experiments showed enrichment at the expected functional
263 genomic regions (Supplementary Fig. 9), lending confidence the lack of
264 enrichment at hotspots does not result from failed ChIP experiments. Nonetheless,
265 transient or weak signals at hotspots could still be missed.
266
267 H3K4me1, H3K27ac and H4ac5 showed marginal enrichment at hotspots (Fig. 4a,
268 b; Supplementary Fig. 8). The weak signals at hotspots are unlikely the result of
269 sub-optimal ChIP-quality, as strong signals are seen at gene promoters and
270 enhancers (Supplementary Fig. 9). Instead, these marks may be rapidly turned-
271 over or present only in a sub-population of early spermatocytes. Indeed, a weak
272 DSB-dependent H4ac5 signal was previously reported at two mouse hotspots[36].
273
274 PRDM9 can trimethylate both H3K4 and H3K36[35] and as expected, H3K4me3 and
275 H3K36me3 were both enriched at hotspots (Fig. 4a and b; Supplementary Fig. 8).
276 H3K4me2 and H3K9ac were also strongly enriched at DSB hotspots genome-
277 wide; this clearly resolves the ambiguity from previous studies where these marks
278 were enriched at one but not another hotspot[36]. H3K4me2 is a likely intermediate
279 of PRDM9 H3K4-trimethylation (see above), however since PRDM9 lacks a
280 histone acetyltransferase domain, it is unlikely that PRDM9 directly acetylates

281 H3K9. In stage-specific experiments, the dynamics of H3K9ac through MPI
282 resemble those of H3K4me3: a signal is first observed at hotspots in leptonema,
283 maximal at zygonema and absent by early pachynema (Fig. 4c). Thus, H3K9ac is
284 a *bona-fide* marker of meiotic DSB hotspots genome-wide, explicitly demonstrating
285 for the first time that proteins other than PRDM9 modify nucleosomes at the sites
286 of DSB hotspots.
287
288 H3K4me3 is positively correlated with DSB frequency[28,41], but most variation in
289 hotspot strength is not accounted for by changes in H3K4me3 (Spearman R^2 =
290 0.46) (Supplementary Fig. 10a). H3K36me3 was previously reported to slightly
291 improve the correlation between H3K4me3 and DSB strength[42], therefore we tried
292 to better predict hotspot strength using a combination of histone marks that are
293 enriched at DSB hotspots. Multiple linear regression with all histone marks only
294 slightly improved the correlation with hotspot strength (max R^2 = 0.52;
295 Supplementary Fig. 10b; see methods) suggesting that these extra histone marks
296 (including H3K36me3 and H3K9ac) do not offer an independent readout of DSB
297 usage as compared to H3K4me3.
298
299 **A histone code distinguishes DSB hotspots from PRDM9-independent**
300 **H3K4me3 sites**
301 In the absence of PRDM9, DSB hotspots occur at sites of non-PRDM9-mediated
302 H3K4me3[28]. Nonetheless, in *wild-type* mice, PRDM9-defined H3K4me3 sites are
303 used. In agreement with previous results[17,18], we found that H3K36me3 is a potent
304 discriminator between DSB hotspots and other H3K4me3-marked sites in the
305 genome (Fig. 4d). Among the other histones, H3K4me2 can best distinguish
306 hotspots (Fig. 4d; Supplementary Fig. 11). We excluded histone marks made by
307 PRDM9 (H3K4me3 and H3K36me3), then performed principal component
308 analysis to explore if any combination(s) of histone marks better define DSB
309 hotspots. The most discriminative PC (PC1) captured signal from H3K9ac and
310 other histone acetylation marks (Fig. 4e), but did not discriminate hotspots as well
311 as H3K4me2 or H3K36me3 alone. These data suggest that H3K36me3, H3K4me2

312 and to a lesser extent H3K9ac and other histone acetylation marks, distinguish the
313 chromatin at sites of PRDM9-marked DSB hotspots from that at other functional
314 sites where H3K4 is tri-methylated.

315

316 ## Discussion

317 In this study, we developed a method to isolate highly pure populations of meiotic
318 nuclei from whole testis. This method is rapid, requires very little starting material,
319 and resolves a major hurdle to studying meiosis in mammals. We have
320 demonstrated that we can efficiently sort five populations of MPI nuclei, however,
321 in principle, any number of specific stages can be purified if the requisite antibodies
322 are available. We obtained populations of up to 96% purity for a given meiotic sub-
323 stage, negating the need to artificially synchronize meiosis using chemicals or the
324 need to use juvenile mice to obtain enriched populations of meiocytes. This opens
325 the door to many detailed studies of meiosis and also to experiments in
326 challenging-to-breed knockout mice and importantly in humans, where other
327 synchronization strategies are impractical.

328

329 Importantly, fixation prior to sorting allows for genome-wide interrogation of
330 transient protein-DNA interactions in sorted populations. To this end, we profiled
331 H3K4me3 using ChIP-Seq in five MPI populations, capturing extensive dynamics
332 of this histone mark. The H3K4me3 marks made at DSB hotspots by PRDM9
333 appear in leptonema, are maximal in zygonema, and are removed by early
334 pachynema. Intriguingly, the presence of robust H3K4me3 signal at zygonema
335 implies that PRDM9 trimethylates more nucleosomes than are used during DSB
336 formation. An alternate explanation would require a large fraction of DSBs to form
337 in zygonema. This is unlikely because homolog synapsis, which occurs in
338 zygonema, begins to shut down the DSB machinery in yeast[43]. An analogous
339 system appears to operate in mammals[44,45]. These are the first data to
340 demonstrate excess H3K4-trimethylation by PRDM9 in *wild-type* mice, however

341 recent work in mice co-expressing both endogenous and transgenic PRDM9
342 reached a similar conclusion[35].
343
344 H3K4me3 at gene promoters is also dynamic through MPI. By comparing
345 H3K4me3 at promoters with gene expression, we found that transcript levels are
346 broadly predicted by H3K4me3. The most cohesive group of transcripts exhibit a
347 coordinate increase in H3K4me3 and in expression late in MPI. This likely reflects
348 the resumption of extensive transcription during mid-pachytene and its continued
349 increase into diplotene[46]. A large subset of transcripts exhibit gene expression that
350 appears decoupled from H3K4me3 levels. Some such transcripts exhibit a strong
351 H3K4me3 signal at the TSS in early MPI, but expression is maximal late in MPI;
352 this may result from accumulation of long-lived mRNA and/or from H3K4me3
353 marks at the sites of poised RNA polymerase[34]. Indeed, these transcripts are
354 intriguing candidates for further understanding gene regulation in MPI. Because
355 we isolate only nuclei, cytoplasmic mRNA cannot be sequenced in our sorted
356 populations. We therefore used mRNA-Seq data from a previously published
357 study[32]. However, we anticipate that recent developments in transcriptomics using
358 RNA from formaldehyde-fixed nuclei[47] will enable us to adapt our method to study
359 transcription in sorted populations of meiotic nuclei. This may help to clarify the
360 relationship between histone marks and active transcription. Finally, a large
361 fraction of dynamic histone marks (H3K4me2/3) in meiosis do not coincide with
362 gene promoters or DSB hotspots. The functions of these sites are unknown
363 although some are likely to be enhancers. Together, they represent a large,
364 heretofore unstudied set of functional sites that may be modulating mouse meiosis.
365
366 To examine the chromatin landscape at DSB hotspots, we also performed the first
367 systematic genome-wide survey on a wide repertoire of histone marks. Our ability
368 to efficiently sort early MPI nuclei eliminated the need to perform these
369 experiments in juvenile mice; this negates the logistical challenge of collecting
370 sufficient juvenile animals and assures consistent and high levels of enrichment.
371 Aside from H3K4me3[28] and H3K36me3[17,18], H3K9ac, H3K4me2, H3K4me1,

372 H3K27ac and H4ac5 were enriched at DSB hotspots genome-wide. Incorporating
373 the extra information about these histone marks only marginally increases our
374 ability to predict hotspot strength, implying that even the expanded chromatin
375 landscape at hotspots remains a poor predictor of DSB formation. The genome is
376 replete with H3K4me3 marks that are used for DSB formation in the absence of
377 PRDM9, however, enigmatically, these sites are not used if PRDM9 is functional.
378 We demonstrated that both H3K36me3[17,40,42], and multiple histone marks
379 distinguish between these sites. Although this unique histone "code" may help the
380 DSB machinery to distinguish PRDM9-defined H3K4me3 from other H3K4me3
381 marks, it is equally possible that direct interactions mediated by PRDM9 itself are
382 important for this distinction. It seems unlikely that PRDM9 is responsible for all of
383 the histone tail modifications at hotspots, therefore other histone remodelers are
384 likely acting at the sites of PRDM9-marked H3K4/K36me3. H3K9ac and H3K4me3
385 are strength correlated and have a similar temporal dynamic implying that they
386 occur at a comparable frequency at hotspots. In mitotic cells, H3K9 acetylation is
387 actively promoted at H3K4 trimethylated nucleosomes[48]. It is therefore likely that
388 H3K9ac is a constitutive response to H3K4 trimethylation. Nonetheless, histone
389 acetylation may be functionally important *per se* in the context of DSB formation;
390 for example, in fission yeast, H3K9 to H3A9 mutation eliminates H3K9ac and
391 consequently reduces DSB formation[40].
392
393 Overall, we demonstrated that our straightforward and robust method can explore
394 previously hidden dynamics of mammalian meiotic recombination. Importantly, our
395 method is widely applicable to other organisms, tissues and cell types, paving the
396 way for further understanding of the temporal dynamics of other developmental
397 processes.
398

Acknowledgements

We thank members from Camerini-Otero's lab for discussions, and Galina Petukhova for insightful comments. We are grateful to Mary Ann Handel for sharing H1t antibodies. This study used the high performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov). We thank members from the NHLBI Flow Cytometry Core for assistance with nuclei sorting, and members from the NIDDK Genomics Core for assistance with high-throughput sequencing. This work was funded by the NIDDK Intramural Research Program (R.D.C.O).

Author contributions

K-W.G.L., K.B., G.C., F.P. and R.D.C-O. conceived the study and designed the experiments. K-W.G.L. performed the experiments. K-W.G.L., K.B. and F.P. analyzed the data. K-W.G.L., K.B., F.P. and R.D.C-O. wrote the manuscript.

Ethical compliance

The authors confirm that all experiments were performed in accordance with the NIH Animal Care and Use regulations.

418 **Table 1.** Summary of cell-type-specific H3K4me3 ChIP-Seq.

419

Sub-stage	Experiment	Antibody	Purity (%)	Starting chromatin DNA (ng)	SPoT (HS; %)	SPoT (All; %)	Peaks (#)
Leptonema	1	EpiCypher	96.9	2,398	2.4	4.5	9,528
	2	EpiCypher	95.3	805	2.2	7.6	13,095
	2	Abcam	95.3	805	3.3	20.3	30,943
Zygonema	1	EpiCypher	79.7	2,688	4.7	6.4	15,699
	2	EpiCypher	64.6	755	3.6	4.2	11,314
	2	Abcam	64.6	755	5.3	17.0	29,114
Early Pachynema	1	EpiCypher	81.6	1,793	1.6	4.2	6,647
	2	EpiCypher	89.5	4,585	1.7	2.3	4,883
	2	Abcam	85.2	960	2.0	14.0	17,561
Late Pachynema	1	EpiCypher	84.1	4,936	1.7	2.6	5,189
	2	EpiCypher	71.3	915	1.6	3.8	9,124
	2	Abcam	71.3	915	1.5	20.1	19,711
Diplonema	1	EpiCypher	88.7	6,502	1.8	3.1	6,645
	2	EpiCypher	82.3	1,115	1.7	5.4	12,071
	2	Abcam	82.3	1,115	1.6	21.0	18,191

420

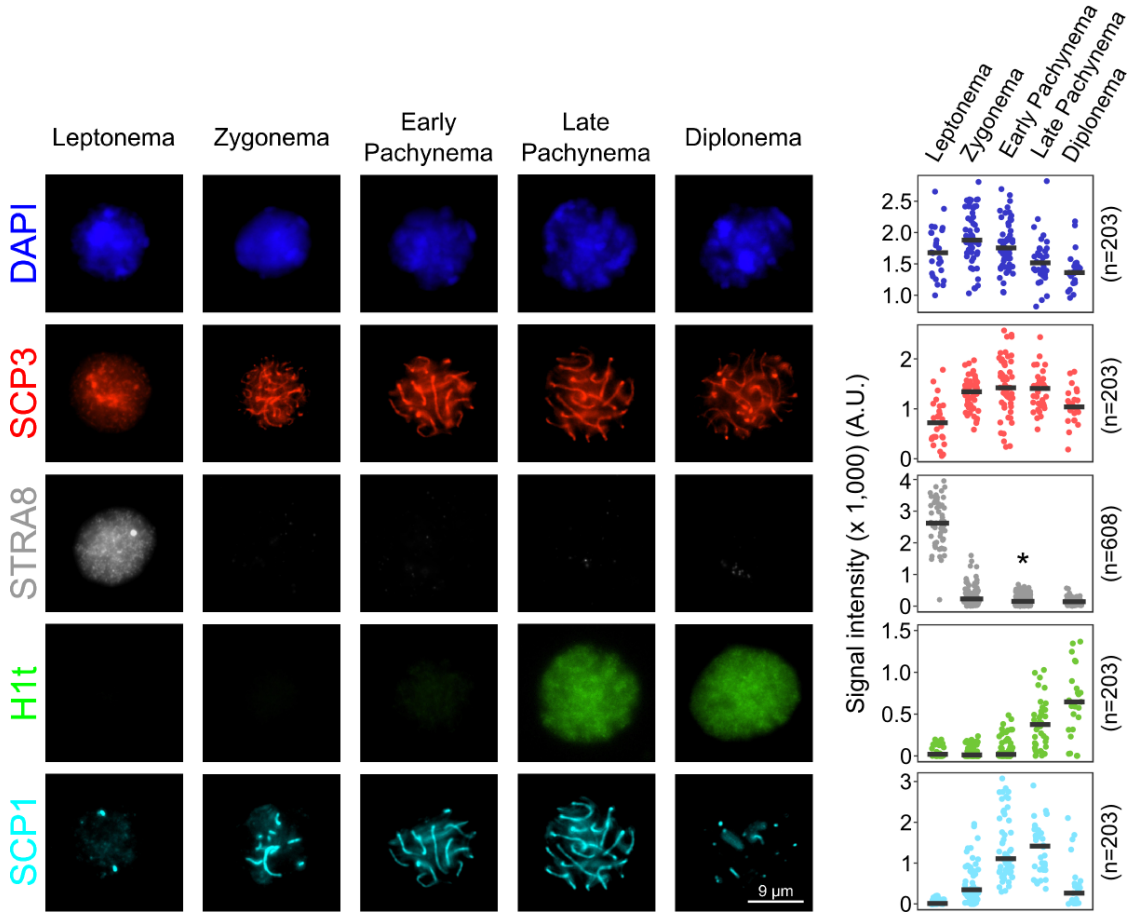


Fig. 1 Immunofluorescence staining of spermatocyte nuclei.

Immunofluorescence images and signal quantification of stage-specific spermatocyte nuclei through meiosis prophase I. Details for signal quantification are described in Methods. Microscopic images are selected from two independent experiments in which two different combinations of primary antibodies are used; one using SCP3, H1t and SCP1, another one using SCP3 and STRA8. *Early and late pachytene nuclei cannot be unambiguously differentiated in the absence of H1t staining, and are therefore merged for counting and signal quantification. Source data are provided as a Source Data file.

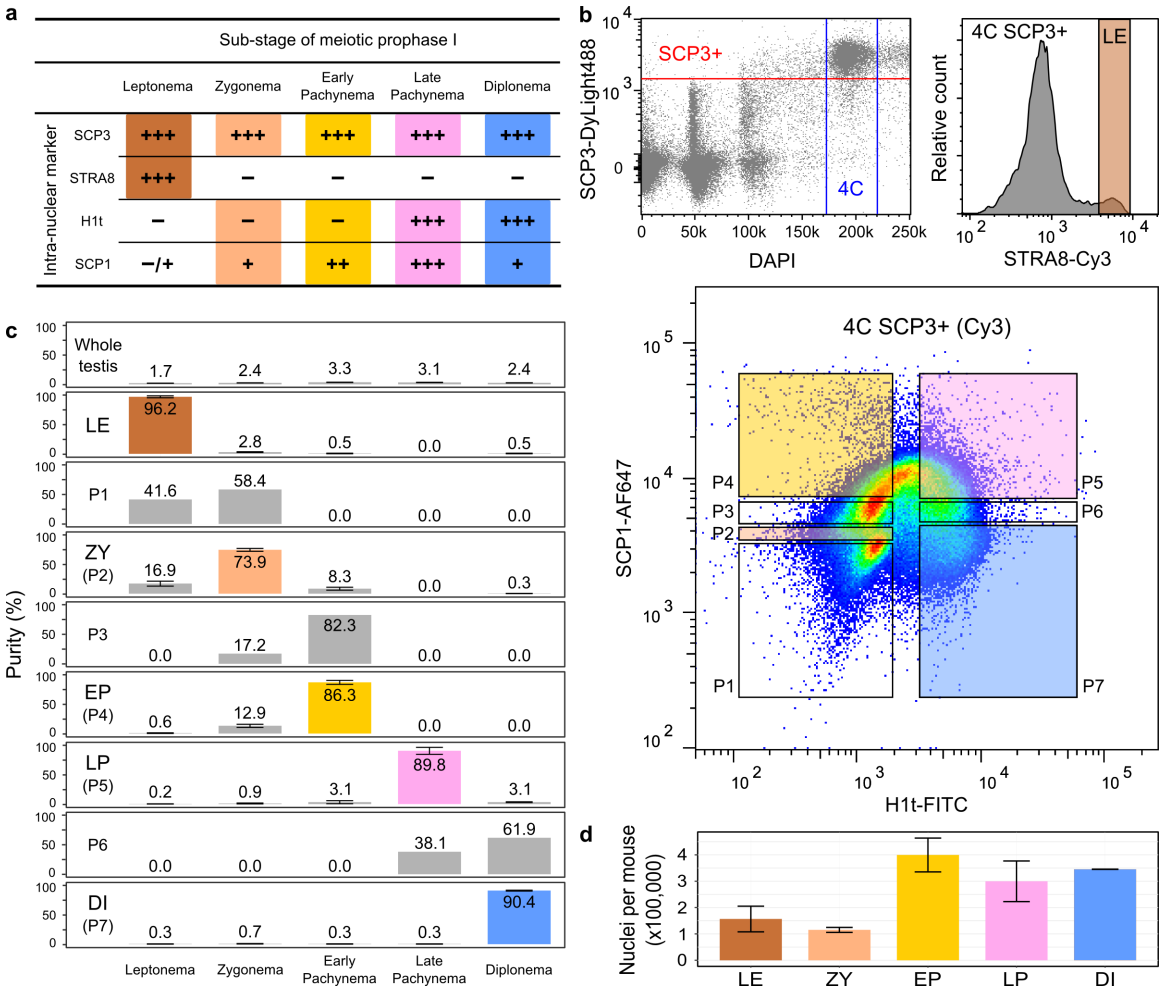


Fig. 2 Experimental design for isolating stage-specific spermatocyte nuclei. **a** Signal strength of intra-nuclear markers across meiotic sub-stages observed in immunofluorescence staining. Signal strength is classified as absent (–), very weak (–/+), weak (+), medium (++), or strong (+++). Combinatorial markers for isolating stage-specific nuclei are highlighted. **b** Flow cytometric strategies for isolating populations of stage-specific nuclei. Meiotic 4C nuclei are gated by DAPI and SCP3 signals (top left). Stage-specific nuclei are sorted into populations based on combinatorial signals of intra-nuclear markers in two separate sorts; one using antibodies against STRA8 for leptonema (top right), and the other one using antibodies against H1t and SCP1 for seven populations (P1 to P7) (bottom). **c** Distributions and purities of each specific type of nuclei in whole-testis and in

447 sorted populations. The five selected populations of leptonema, zygonema, early
448 pachynema, late pachynema, and diplonema are highlighted as LE, ZY, EP, LP,
449 and DI, respectively. Purities (mean with standard error) of these populations are
450 derived from two or three independent sorts. Source data are provided as a Source
451 Data file. **d** Numbers of nuclei in each sub-population collected in an adult mouse.
452 Data (mean with standard error) are derived from two independent sorts.

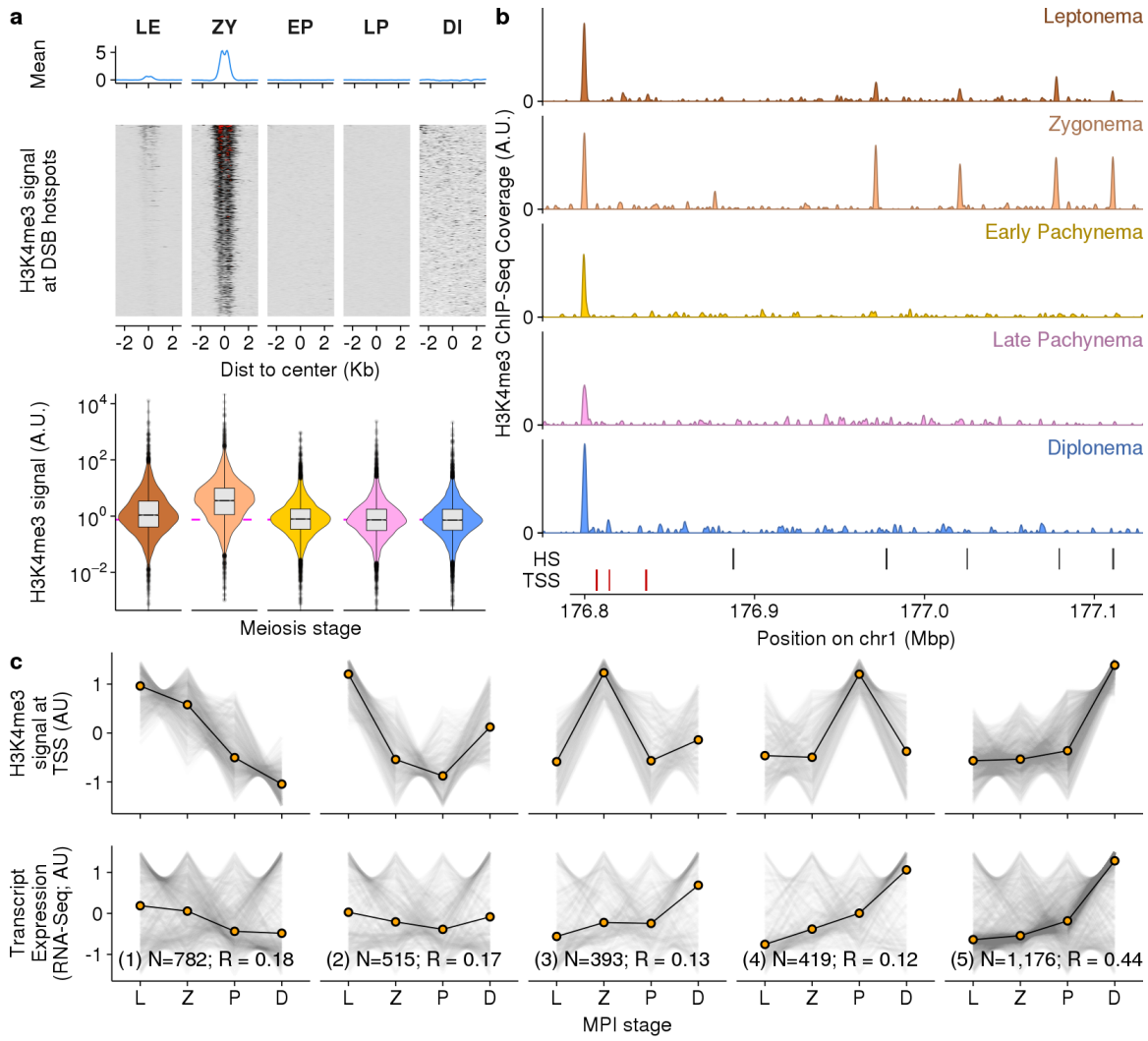


Fig. 3 H3K4me3 dynamics across meiotic sub-stages.

a H3K4me3 dynamics at DSB hotspots across five meiotic sub-stages. H3K4me3 signals at DSB hotspots are normalized by the spike-in H4K4me3 mononucleosomes. **b** H3K4me3 dynamics at different genomic elements in five sorted sub-populations of meiotic nuclei. **c** Stage-specific H3K4me3 at TSSs recapitulate meiotic gene expression patterns. Genes with only one isoform are selected for this analysis. H3K4me3 profiles at TSSs are grouped into five clusters (top), while the respective mRNA levels are shown in the bottom panel. L = Leptonema, Z = Zygonema, P = Early and Late Pachynema, D = Diplonema.

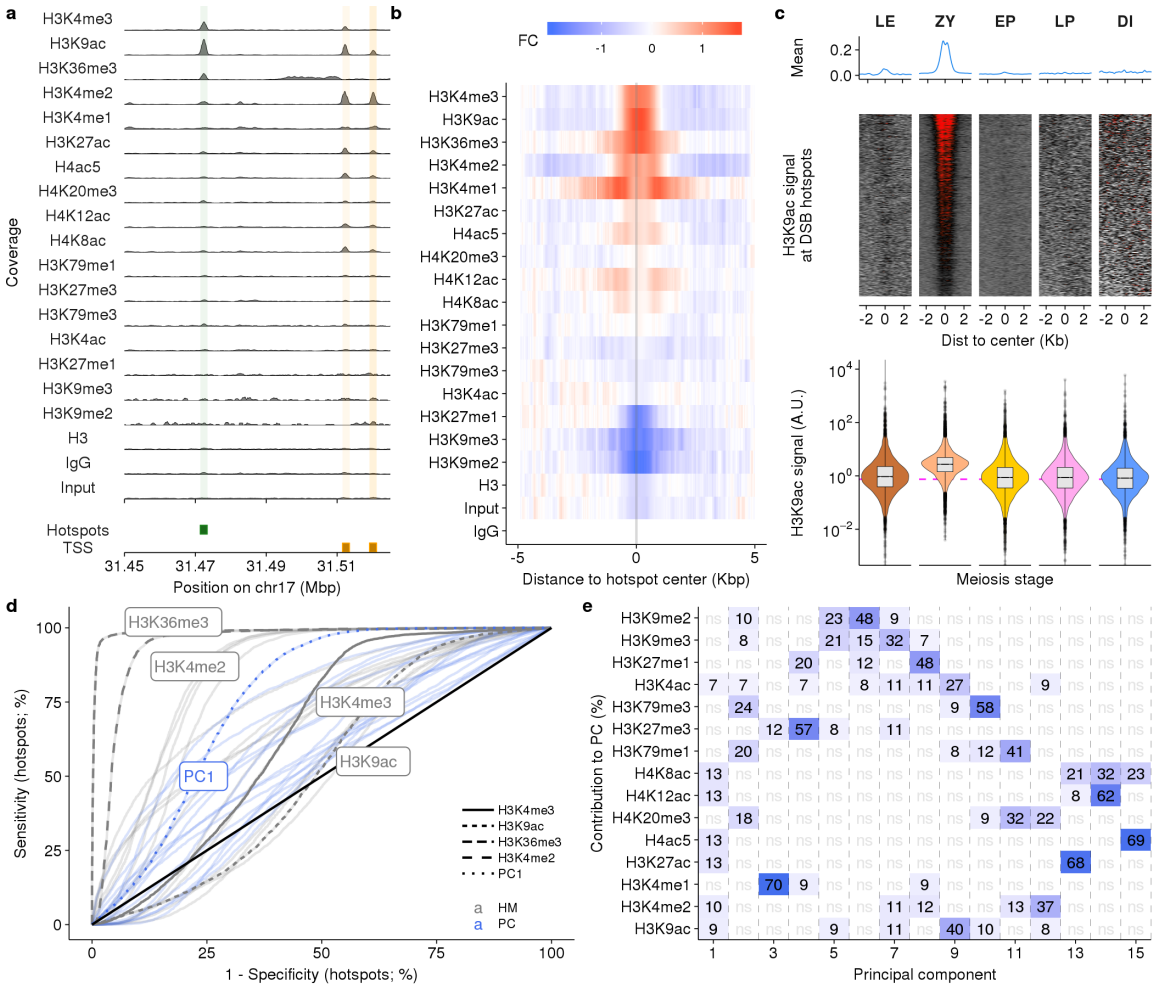


Fig. 4 Histone modifications at DSB hotspots.

a Distributions of histone marks along a genomic region. Locations of DSB hotspots are shaded in green, TSS in orange. **b** Histone marks are enriched at DSB hotspots. Red depicts enrichment relative to IgG, blue is depletion. **c** H3K9ac dynamics at DSB hotspots across five meiotic sub-stages. **d-e** Principal component analysis using histone marks can distinguish DSB hotspots from PRDM9-independent H3K4me3 sites. **d** Histone modifications can distinguish DSB hotspots from other H3K4me3 sites. An ROC curve for each histone mark (gray) or PC (blue) is shown. Selected histone modifications and PC1 are highlighted. **e** Contributions of histone marks to each PC. Only marks that contribute more than expected are shown. Deeper blue indicates a stronger contribution (numbers show percent contribution).

478 # Methods

479

480 **Mouse strains**

481 C57BL/6J (B6) mice were either obtained from The Jackson Laboratory (Stock no.
482 000664) or bred in-house. All experiments were done on adult mice (≥ 8 weeks of
483 age).

484

485 **Antibodies**

486 The following antibodies were used for primary immunofluorescence staining: anti-
487 SCP3 (D-1) (Santa Cruz, sc-74569), anti-STRA8 (Abcam, ab49602), anti-H1t (a
488 gift from Mary Ann Handel and a custom-made antibody), and anti-SCP1 [Biotin]
489 (Novus Biologicals, NB300-229B).

490

491 The following antibodies were used for secondary immunofluorescence staining:
492 goat anti-mouse IgG conjugated with Cy3 (Jackson ImmunoResearch, 115-167-
493 003), goat anti-rabbit IgG conjugated with FITC (Jackson ImmunoResearch, 115-
494 097-003), goat anti-mouse IgG conjugated with DyLight488 (Abcam, ab98757),
495 goat anti-rabbit IgG conjugated with Cy3 (Jackson ImmunoResearch, 111-165-
496 003), goat anti-guinea pig IgG conjugated with Alexa Fluor 488 (Thermo Fisher
497 Scientific, A-11073), streptavidin conjugated with Alexa Fluor 647 (Thermo Fisher
498 Scientific, S21374).

499

500 The following antibodies were used for ChIP: anti-H3K4me3 (Abcam, ab8580),
501 anti-H3K4me3 (EpiCypher, 13-0028), anti-H3K9ac (Active Motif, 39918), anti-
502 H3K4me2 (Active Motif, 39914), anti-H3K36me3 (Active Motif, 61102), anti-H4ac5
503 (Millipore, 06-946), anti-H3K4me1 (Abcam, ab8895), anti-H3K27ac (Abcam,
504 ab177178), anti-H4K8ac (Abcam, ab15823), anti-H4K12ac (Active Motif, 39928),
505 anti-H3K27me3 (Millipore, 07-449), anti-H4K20me3 (Millipore, 07-463), anti-
506 H3K79me1 (Abcam, ab2886), anti-H3K4ac (Abcam, ab176799), anti-H3K79me3
507 (Abcam, ab2621), anti-H3K9me2 (Abcam, ab1220), anti-H3K9me3 (Active Motif,

508 39766), anti-H3K27me1 (Millipore, 07-448), anti-H3 (Abcam, ab1791), normal
509 rabbit IgG (Millipore, 12-370).
510
511 **Nuclei preparation**
512 Testes from adult mice were de-capsulated and fixed with 1% formaldehyde for 10
513 min followed by 5 min of quenching at RT. Fixed tissues were homogenized and
514 filtered through a 70- or 100-μm cell strainer. After washing with chilled 1 x PBS,
515 nuclei were extracted using nucleus extraction buffer (15 mM Tris-HCl pH 7.4, 0.34
516 M sucrose, 15 mM NaCl, 60 mM KCl, 0.2 mM EDTA, 0.2 mM EGTA) on ice for 5
517 min and were homogenized with 20 strokes with loose pestle followed by 10
518 strokes with tight pestle. Nuclei were filtered through a 40-μm cell strainer and
519 resuspended in chilled PBTB buffer (1 x PBS with 0.1% Triton X-100, 5% BSA and
520 protease inhibitor).
521
522 **Immunofluorescence staining**
523 Nuclei were incubated in 10% of normal serum at RT for 10 min. This blocking step
524 helps reduce non-specific binding of antibodies. Nuclei were then labelled with
525 different combinations of primary antibodies (1 μg of antibodies to 11 million
526 events) in 10% of normal serum at 20°C for 40 min. Nuclei were washed,
527 resuspended in PBTB, blocked with serum, and then labelled with secondary
528 antibodies (1:250 dilution) in 10% of normal serum at 20°C for 30 min. Controls for
529 each secondary antibody were also prepared for setting up the threshold of
530 background fluorescent signal for FACS. Nuclei were washed, resuspended in
531 PBTB, and stored at 4°C until sorting.
532
533 **Counting stage-specific nuclei from whole-testis**
534 Nuclei isolated from whole-testis were stained with antibodies against SCP3, H1t
535 and SCP1, and subsequently with secondary antibodies as described.
536 Immunofluorescence-labelled nuclei were spread on microslides, stained with
537 mounting medium with DAPI (Vetashield), sealed with cover slips, and viewed
538 using fluorescence microscopy. Each of the five types of MPI nuclei (SCP3+) were

539 counted; these were characterized by the patterning of the SCP3 and H1t proteins
540 (Fig. 1). A total of 678 nuclei were counted from two independent experiments. To
541 extrapolate counts to the percentage of cells of each stage in whole testis (top
542 panel in Fig. 2c), we use the following formula:
543
544 WT_{stage} (%) = $N_{stage}/678 * (WT_{4C}/WT_{all}) * 100$
545
546 N_{stage} = number of nuclei of each stage
547 WT_{4C} = number of total 4C (SCP3+) nuclei from sorting
548 WT_{all} = number of total singlet nuclei (1C to 4C) from sorting
549
550 **Quantification of immunofluorescence signal intensity**
551 Nuclei isolated from whole-testis were stained with antibodies against SCP3 and
552 STRA8 or with antibodies against SCP3, H1t and SCP1, subsequently with
553 secondary antibodies, and spread as previously described. Immunofluorescence
554 signals of each marker were quantified using Volocity 6.2.1 (PerkinElmer). 4C
555 nuclei were identified using signals from the DAPI channel by excluding objects
556 with size < 40 μm^2. Touching objects were separated using the parameter of 30–50
557 μm^2. Stage-specific nuclei in MPI were characterized by the presence of SCP3
558 signal and by the patterning of synaptonemal complex[14]. Signals from all channels
559 were recorded and exported for signal quantification.
560
561 **Isolation of meiotic sub-populations with FACS**
562 Nuclei were filtered through a 40-μm cell strainer, and stained with DAPI for 30 min
563 or longer at RT before sorting. All sorting experiments were performed on either a
564 BD FACSAria II or a BD FACSAria Fusion flow cytometer at a flow rate of ~20,000
565 events/sec. Singlets were gated using both forward scatter and side scatter. Nuclei
566 from primary spermatocytes (4C) were gated based on DNA content deduced from
567 the DAPI signal, and sorted into populations of interest based on fluorochrome
568 intensity into collection tubes containing PBTB. Sorted nuclei were collected by
569 centrifugation. Nuclei were examined under microscope and counted with a

570 hemocytometer. Finally, excess buffer was removed and nuclei pellets were stored
571 at −80°C.
572
573 **Assessment of purity of sorted populations**
574 An aliquot of nuclei suspension from each sorted population was spread on
575 microslides, further stained with mounting medium with DAPI (Vectashield), and
576 sealed with coverslips for purity examination. About 150 nuclei from each sorted
577 population were checked at 400X magnification using immunofluorescence
578 microscopy. The staging was primarily accessed by the patterning of SCP3[14] and
579 further confirmed by signals of STRA8 or of H1t and SCP1 accordingly.
580
581 **Chromatin extraction**
582 Chromatin extracted for ChIP-Seq was either sheared by sonication or Micrococcal
583 nuclease (MNase) digestion. Chromatin for all H3K4me3/H3K9ac-ChIP-Seq for
584 studying histone dynamics in stage-specific populations was sheared by
585 sonication, whereas ChIP-Seq for identifying histone marks at hotspots were
586 performed on MNase-digested chromatin. Experimental procedures are listed
587 below in detail.
588
589 **Shearing chromatin by sonication:**
590 Frozen nuclei pellets were thawed at RT for 10 min. Nuclei were lysed with Lysis
591 buffer 1 (0.25% Triton X-100, 10 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl pH 8)
592 and incubated at RT for 10 min. Nuclei pellets were subsequently washed with
593 Lysis buffer 2 (200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl pH 8)
594 and lysed with RIPA buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 0.5 mM EGTA,
595 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 140 mM NaCl plus
596 protease inhibitor). Chromatin was sheared into ~50−300 bp fragments by
597 sonication using Bioruptor (diagenode). The amount of chromatin DNA was
598 measured using a Qubit dsDNA HS Assay Kit (Thermo Fisher).
599
600 **MNase digestion:**

601 Frozen nuclei pellets were thawed at RT for 10 min. Nuclei pellets were
602 resuspended with MNase buffer (50 mM Tris-HCl pH 8, 1 mM $CaCl_2$, 4 mM $MgCl_2$,
603 4% NP-40 plus protease inhibitor). MNase digestion was performed in a
604 concentration of 3U MNase (USB, Affymetrix) per one million nuclei at 37°C for 5
605 min. The reaction was stopped by adding a final concentration of 10 mM EDTA
606 and incubated at 4°C for 5 min. Soluble chromatin was collected and diluted with
607 RIPA buffer.
608
609 **Chromatin Immunoprecipitation**
610 For stage-specific H3K4me3-ChIP-Seq, modified mononucleosomes (SNAP-ChIP
611 K-MetStat Panel, EpiCypher) were added at a spike-in volume of 1 μl per ~1.5 μg
612 chromatin DNA. Chromatin was immunoprecipitated with 0.8–5 μg antibodies (or
613 5 μl unpurified serum) in 0.5–1 ml RIPA buffer at 4°C overnight (see
614 Supplementary Table 1 for details). The immuno-complexes were captured using
615 20–75 μl Dynabeads Protein G (30 mg/ml, Novex) at 4°C for 2 hr. The beads were
616 washed with buffers previously described[27]. ChIPed DNA was eluted using a IPure
617 kit v2 (diagenode).
618
619 **Sequencing library construction**
620 ChIP-seq libraries were constructed with a KAPA Hyper Prep Kit (Kapa
621 Biosystems) following steps for generating 1 μg of library DNA. DNA libraries were
622 cleaned up with an Agencourt AMPure XP system (Beckman Coulter). The DNA
623 concentration and fragment size of these libraries were measured with a Qubit
624 dsDNA HS Assay Kit (Thermo Fisher) and an Agilent High Sensitivity DNA Kit
625 (Agilent), respectively.
626
627 **High throughput DNA sequencing**
628 DNA sequencing was performed on the Illumina HiSeq 2500 or HiSeq X.
629 Sequencing tags were aligned to the mouse mm10 reference genome using BWA
630 mem 0.7.12[49]. For ChIP-Seq samples containing the K-MetStat spike-in panel, we

631 aligned to a custom mm10 genome to which we added the sequences of the DNA
632 bound by each K-MetStat nucleosome.
633
634 **Inter-sample normalization and peak strength estimation using spike-in**
635 **nucleosomes**
636 We retained only sequencing reads with an alignment q-score >= 20, that were
637 mapped in a "proper pair" (for paired-end data), that passed vendor QC and that
638 were mapped as a primary alignment. For paired-end data, we only retained reads
639 from fragments with an insert size between 100 bp and 200 bp. Normalization
640 using the K-MetStat spike-in was performed as described previously [29], with slight
641 modification. Briefly, each spike-in nucleosome is bound to a unique DNA
642 sequence. Thus, the number of reads recovered for each nucleosome is a proxy
643 for the frequency of that nucleosome in the library. We first converted paired-end
644 reads to fragments, then calculated the enrichment of fragments mapping to each
645 spike-in sequence after ChIP:
646
647 E_{si} = (barcode fragments in ChIP / barcode fragments in input)
648
649 Next, we calculated H3K4me3 ChIP-Seq fragment coverage in all 25 bp non-
650 overlapping windows in the genome. The per-locus enrichment was then
651 calculated as :
652
653 E_{locus} = (fragment coverage in ChIP) / (fragment coverage in Input)
654
655 The histone modification density was then calculated as :
656
657 HMD (%) = 100 * (E_{locus} / $E_{H3K4me3_A}$)
658
659 $E_{H3K4me3_A}$ is the E_{si} value for the H3K4me3_A spike_in. For our final estimate of
660 H3K4me3 signal, we normalized the HMD by the global average HMD. The ChIP-

661 Seq signal at peaks and other genomic intervals was calculated as the sum of the
662 signal in all overlapping 25 bp windows.
663
664 **Inter-sample normalization and peak strength estimation using "stable"**
665 **promoters**
666 The K-MetStat panel cannot be used to normalize data from H3K9ac ChIP-Seq.
667 Thus, we used an alternative strategy using stable promoters. NCIS[50] was used to
668 estimate the contribution of background to each sequencing library. We then
669 calculated the NCIS normalized ChIP-Seq fragment coverage in all 25 bp windows
670 in the genome (c = ChIP coverage - (Input coverage * NCIS factor)). The ChIP-
671 Seq signal at the central ± 1 Kbp region around GENCODE transcript 5' ends
672 (TSS) was quantified as the sum of the coverage in all 25 bp widows overlapping
673 the site. TSS coverage was estimated in each of the five meiotic populations. TSS
674 that overlapped a DSB hotspot, TSSs with a negative strength after correction and
675 TSS with the strongest (1%) and weakest (15%) H3K4me3 signal at each stage
676 were discarded. The $\log_2$ ratio of NCIS-corrected H3K4me3 was calculated at each
677 TSS between all pairs of stages. TSSs with an absolute $\log_2$ (fold-change) ≤ 1.8
678 between all stages were retained. This resulted in 362 TSSs for H3K4me3 and
679 187 TSSs for H3K9ac. The median signal across these intervals was calculated
680 for each ChIP and used as a normalization factor. The strength of each 25 bp
681 genome interval was calculated as the NCIS-corrected coverage divided by the
682 normalization factor. The ChIP-Seq signal at peaks and other genomic intervals
683 was calculated as the sum of the signal in all overlapping 25 bp windows. Note
684 that for H3K4me3 ChIP-Seq, this internal normalization gives comparable results
685 to those using the K-MetStat spike-in panel (Supplementary Fig. 4).
686
687 **Correlating H3K4me3 profiles at TSSs with mRNA expression levels**
688 We used RNA-Seq data from spermatocyte populations obtained by classical cell-
689 sorting to quantify gene expression[32]. Gene expression at GENCODE vM20
690 transcripts was quantified for each RNA-Seq dataset using kallisto 0.45.0 (kallisto
691 quant --single --single --single-overhang --fragment-length=250 --sd=100 --

692 seed=42). Only transcripts with maximal expression > 0.2 tags per million (TPM),
693 with at least a 1.25-fold change in expression and with non-zero expression and
694 H3K4me3 signals at all stages were considered for subsequent analysis. To
695 minimize the confounding effects of multiple isoforms, only TSS with a single
696 transcript in this list were retained.
697
698 The temporal expression profile for each transcript was described as a four unit
699 vector (Leptonema, Zygonema, Pachynema, Diplonema). Similarly, a temporal
700 profile of H3K4me3 at the TSS was generated. We averaged the Early Pachnema
701 and Late Pachynema H3K4me3 signals to allow direct comparison with the RNA-
702 Seq data.
703
704 The overall correlation between H3K4me3 and gene expression was calculated by
705 concatenating all gene expression vectors and comparing with a concatenated
706 vector of H3K4me3. To calculate the expected random correlation, we shuffled the
707 order of each 4-unit transcript H3K4me3 vector before concatenation. 10,000
708 iterations of this process were performed.
709
710 k-means clustering was used to group similar H3K4me3 temporal profiles and to
711 compare each set to gene expression. We used an implementation of the gap-
712 statistic to determine the optimal number of clusters (R; factoextra package).
713 Within each set of clusters the method described in the previous paragraph was
714 employed to determine the correlation between H3K4me3 and expression
715 temporal profiles and to derive an empirical p-value for this correlation. Clustering
716 by gene expression instead of by H3K4me3 signal gives similar correlations,
717 though the number of clusters can change (data not shown).
718
719 **Peak calling**
720 Peaks for H3K4me3 ChIP-Seq were called using MACS2 (version 2.1.2)[51] with
721 default parameters except (-q 0.1 --broad) and with a stage-matched input DNA
722 sample as a control. Peaks overlapping DSB hotspots or gene promoters were

723 ascertained using bedtools (version v2.27.1)[52] after removing blacklisted regions
724 described in[53]. DSB hotspots were defined and reported in previous studies[28].
725 Transcription start sites (TSSs) were defined as the ± 0.5 Kbp region around
726 GENCODE v20 transcripts[54].
727
728 **Unbiased clustering**
729 ChIP-Seq peaks from all five populations were merged and peak strength was
730 calculated (as described above) for each merged peak in each population. The
731 H3K4me3 profile for each peak was described as a five unit vector (LE, ZY, EP,
732 LP, DI) and scaled by subtracting the mean and dividing by the standard deviation.
733 k-means clustering was used to cluster. Five clusters was determined as optimal
734 using the gap-statistic.
735
736 **Multiple linear regression**
737 Hotspots overlapping a TSS or overlapping a site used for DSB formation in
738 *Prdm9*$^{-/-}$ mice were discarded. Only hotspots with read coverage > 0 for all histone
739 marks were used for regression analyses because we performed regression on
740 the log10 transformed coverage values. The leaps package in R was used to
741 perform an all-subsets regression using the seven histone marks enriched at DSB
742 hotspots.
743
744 **Principal Component Analysis**
745 All H3K4me3 peaks from LE, ZY, EP and LP were used for analysis. Each interval
746 was resized to ± 250 bp around the center. Sequencing reads for each histone
747 mark were counted at each feature. Input DNA reads were also counted and
748 subtracted from the count for each mark, following NCIS[50] correction. DSB
749 hotspots and GENCODE TSSs were expanded to ± 1,500 bp to determine
750 H3K4me3 peaks that overlapped hotspots and TSSs, respectively. Peaks that
751 overlapped both a hotspot and TSS were discarded as the potential compound
752 signal would confound these analyses. Only autosomal peaks were used.
753 H3K36me3 and H3K4me3 were excluded. The R prcomp command was used for

754 Principal Component Analysis. Variables were scaled to have unit variance and
755 zero centered.
756
757 ROC curves were built by ranking intervals by the dependent variable. Intervals
758 were ranked either from high to low or from low to high and the ROC with the higher
759 area under the curve was used.
760
761 **Code availability**
762 The custom pipeline used for these analyses is deposited at zenodo
763 (https://doi.org/10.5281/zenodo.2651204)[55].

764 **Data availability**

765 The authors declare that all data generated or analysed during this study are
766 included in this published article (and its supplementary information files). The
767 sequencing data reported in this paper are archived at the Gene Expression
768 Omnibus (www.ncbi.nlm.nih.gov/geo) as accession no. GSE121760. The source
769 data underlying Figs 1 and 2c are provided as a Source Data file.

770 ## References

771 1. Baudat, F., Imai, Y. & de Massy, B. Meiotic recombination in mammals: localization
772 and regulation. *Nat. Rev. Genet.* **14**, 794–806 (2013).
773 2. Handel, M. A., Eppig, J. J. & Schimenti, J. C. Applying 'Gold Standards' to In-Vitro-
774 Derived Germ Cells. *Cell* **157**, 1257–1261 (2014).
775 3. Bellve, A. R., Cavicchia, J. C. & Millette, C. F. Spermatogenic cells of the prepuberal
776 mouse: isolation and morphological characterization. *J. Cell Biol.* (1977).
777 4. Goetz, P., Chandley, A. C. & Speed, R. M. Morphological and temporal sequence of
778 meiotic prophase development at puberty in the male mouse. *J. Cell Sci.* **65**, 249–
779 263 (1984).
780 5. Kluin, P. M., Kramer, M. F. & de Rooij, D. G. Spermatogenesis in the immature
781 mouse proceeds faster than in the adult. *Int. J. Androl.* **5**, 282–294 (1982).
782 6. Zelazowski, M. J. *et al.* Age-Dependent Alterations in Meiotic Recombination Cause
783 Chromosome Segregation Errors in Spermatocytes. *Cell* **171**, 601–614.e13 (2017).
784 7. Grive, K. J. *et al.* Dynamic transcriptome profiles within spermatogonial and
785 spermatocyte populations during postnatal testis maturation revealed by single-cell
786 sequencing. *PLoS Genet.* **15**, e1007810 (2019).
787 8. Hogarth, C. A. *et al.* Turning a spermatogenic wave into a tsunami: synchronizing
788 murine spermatogenesis using WIN 18,446. *Biol. Reprod.* **88**, 40–41 (2013).
789 9. Bastos, H. *et al.* Flow cytometric characterization of viable meiotic and postmeiotic
790 cells by Hoechst 33342 in mouse spermatogenesis. *Cytometry A* **65**, 40–49 (2005).
791 10. Barchi, M., Geremia, R., Magliozzi, R. & Bianchi, E. Isolation and analyses of
792 enriched populations of male mouse germ cells by sedimentation velocity: the
793 centrifugal elutriation. *Methods Mol. Biol.* **558**, 299–321 (2009).
794 11. Gaysinskaya, V., Soh, I. Y., Heijden, G. W. & Bortvin, A. Optimized flow cytometry

795 isolation of murine spermatocytes. *Cytometry A* **85**, 556–565 (2014).
796 12. Bonn, S. *et al.* Cell type–specific chromatin immunoprecipitation from multicellular
797 complex samples using BiTS-ChIP. *Nat. Protoc.* **7**, 978 (2012).
798 13. Meuwissen, R. L. *et al.* A coiled-coil related protein specific for synapsed regions of
799 meiotic prophase chromosomes. *EMBO J.* **11**, 5091–5100 (1992).
800 14. Dobson, M. J., Pearlman, R. E., Karaiskakis, A., Spyropoulos, B. & Moens, P. B.
801 Synaptonemal complex proteins: occurrence, epitope mapping and chromosome
802 disjunction. *J. Cell Sci.* **107 (Pt 10)**, 2749–2760 (1994).
803 15. Inselman, A., Eaker, S. & Handel, M. A. Temporal expression of cell cycle-related
804 proteins during spermatogenesis: establishing a timeline for onset of the meiotic
805 divisions. *Cytogenet. Genome Res.* **103**, 277–284 (2003).
806 16. Mark, M. *et al.* STRA8-deficient spermatocytes initiate, but fail to complete, meiosis
807 and undergo premature chromosome condensation. *J. Cell Sci.* **121**, 3233–3242
808 (2008).
809 17. Powers, N. R. *et al.* The Meiotic Recombination Activator PRDM9 Trimethylates
810 Both H3K36 and H3K4 at Recombination Hotspots In Vivo. *PLoS Genet.* **12**,
811 e1006146 (2016).
812 18. Grey, C. *et al.* In vivo binding of PRDM9 reveals interactions with noncanonical
813 genomic sites. *Genome Res.* **27**, 580–590 (2017).
814 19. Margolin, G., Khil, P. P., Kim, J., Bellani, M. A. & Camerini-Otero, R. D. Integrated
815 transcriptome analysis of mouse spermatogenesis. *BMC Genomics* **15**, 39 (2014).
816 20. Sun, F. *et al.* Nuclear localization of PRDM9 and its role in meiotic chromatin
817 modifications and homologous synapsis. *Chromosoma* **124**, 397–415 (2015).
818 21. Moens, P. B. *et al.* The time course and chromosomal localization of recombination-
819 related proteins at meiosis in the mouse are compatible with models that can

820 resolve the early DNA-DNA interactions without reciprocal recombination. *J. Cell*
821 *Sci.* **115**, 1611–1622 (2002).
822 22. Cole, F. *et al.* Homeostatic control of recombination is implemented progressively in
823 mouse meiosis. *Nat. Cell Biol.* **14**, 424–430 (2012).
824 23. Moens, P. B. Histones H1 and H4 of surface-spread meiotic chromosomes.
825 *Chromosoma* **104**, 169–174 (1995).
826 24. Adli, M. & Bernstein, B. E. Whole-genome chromatin profiling from limited numbers
827 of cells using nano-ChIP-seq. *Nat. Protoc.* **6**, 1656–1668 (2011).
828 25. Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone
829 modifications in human and mouse. *Cell* **120**, 169–181 (2005).
830 26. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the
831 human genome. *Nature* **489**, 57–74 (2012).
832 27. Smagulova, F. *et al.* Genome-wide analysis reveals novel molecular features of
833 mouse recombination hotspots. *Nature* **472**, 375–378 (2011).
834 28. Brick, K., Smagulova, F., Khil, P., Camerini-Otero, R. D. & Petukhova, G. V. Genetic
835 recombination is directed away from functional genomic elements in mice. *Nature*
836 **485**, 642–645 (2012).
837 29. Grzybowski, A. T., Chen, Z. & Ruthenburg, A. J. Calibrating ChIP-Seq with
838 Nucleosomal Internal Standards to Measure Histone Modification Density Genome
839 Wide. *Mol. Cell* **58**, 886–899 (2015).
840 30. Mimitou, E. P., Yamada, S. & Keeney, S. A global view of meiotic double-strand
841 break end resection. *Science* **355**, 40–45 (2017).
842 31. Shah, R. N. *et al.* Examining the Roles of H3K4 Methylation States with
843 Systematically Characterized Antibodies. *Mol. Cell* **72**, 162–177.e7 (2018).
844 32. Gaysinskaya, V. *et al.* Transient reduction of DNA methylation at the onset of

845 meiosis in male mice. *Epigenetics Chromatin* **11**, 15 (2018).
846 33. Tibshirani, R., Walther, G. & Hastie, T. Estimating the number of clusters in a data
847 set via the gap statistic. *J. R. Stat. Soc. Series B Stat. Methodol.* **63**, 411–423
848 (2001).
849 34. Vastenhouw, N. L. *et al.* Chromatin signature of embryonic pluripotency is
850 established during genome activation. *Nature* **464**, 922–926 (2010).
851 35. Diagouraga, B. *et al.* PRDM9 Methyltransferase Activity Is Essential for Meiotic DNA
852 Double-Strand Break Formation at Its Binding Sites. *Mol. Cell* **69**, 853–865.e6
853 (2018).
854 36. Buard, J., Barthès, P., Grey, C. & de Massy, B. Distinct histone modifications define
855 initiation and repair of meiotic recombination in the mouse. *EMBO J.* **28**, 2616–2624
856 (2009).
857 37. Walker, M. *et al.* Affinity-seq detects genome-wide PRDM9 binding sites and
858 reveals the impact of prior chromatin modifications on mammalian recombination
859 hotspot usage. *Epigenetics Chromatin* **8**, 31 (2015).
860 38. Getun, I. V. *et al.* Functional Roles of Acetylated Histone Marks at Mouse Meiotic
861 Recombination Hot Spots. *Mol. Cell. Biol.* **37**, (2017).
862 39. Merker, J. D. *et al.* The histone methylase Set2p and the histone deacetylase
863 Rpd3p repress meiotic recombination at the HIS4 meiotic recombination hotspot in
864 Saccharomyces cerevisiae. *DNA Repair* **7**, 1298–1308 (2008).
865 40. Yamada, S., Ohta, K. & Yamada, T. Acetylated Histone H3K9 is associated with
866 meiotic recombination hotspots, and plays a role in recombination redundantly with
867 other factors including the H3K4 methylase Set1 in fission yeast. *Nucleic Acids Res.*
868 **41**, 3504–3517 (2013).
869 41. Lange, J. *et al.* The Landscape of Mouse Meiotic Double-Strand Break Formation,

870 Processing, and Repair. *Cell* **167**, 695–708.e16 (2016).
871 42. Yamada, S. *et al.* Genomic and chromatin features shaping meiotic double-strand
872 break formation and repair in mice. *Cell Cycle* **16**, 1870–1884 (2017).
873 43. Thacker, D., Mohibullah, N., Zhu, X. & Keeney, S. Homologue engagement controls
874 meiotic DNA break number and distribution. *Nature* **510**, 241–246 (2014).
875 44. Wojtasz, L. *et al.* Mouse HORMAD1 and HORMAD2, two conserved meiotic
876 chromosomal proteins, are depleted from synapsed chromosome axes with the help
877 of TRIP13 AAA-ATPase. *PLoS Genet.* **5**, e1000702 (2009).
878 45. Kauppi, L. *et al.* Numerical constraints and feedback control of double-strand breaks
879 in mouse meiosis. *Genes Dev.* **27**, 873–886 (2013).
880 46. Page, J. *et al.* Inactivation or non-reactivation: what accounts better for the silence
881 of sex chromosomes during mammalian male meiosis? *Chromosoma* **121**, 307–326
882 (2012).
883 47. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian
884 organogenesis. *Nature* **566**, 496–502 (2019).
885 48. Gates, L. A., Foulds, C. E. & O'Malley, B. W. Histone Marks in the 'Driver's Seat':
886 Functional Roles in Steering the Transcription Cycle. *Trends Biochem. Sci.* **42**, 977–
887 989 (2017).
888 49. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-
889 MEM. *arXiv [q-bio.GN]* (2013).
890 50. Liang, K. & Keleş, S. Normalization of ChIP-seq data with control. *BMC*
891 *Bioinformatics* **13**, 199 (2012).
892 51. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137
893 (2008).
894 52. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing

895 genomic features. *Bioinformatics* **26**, 841–842 (2010).

896 53. Smagulova, F., Brick, K., Pu, Y., Camerini-Otero, R. D. & Petukhova, G. V. Erratum:
897 The evolutionary turnover of recombination hot spots contributes to speciation in
898 mice. *Genes Dev.* **30**, 871 (2016).

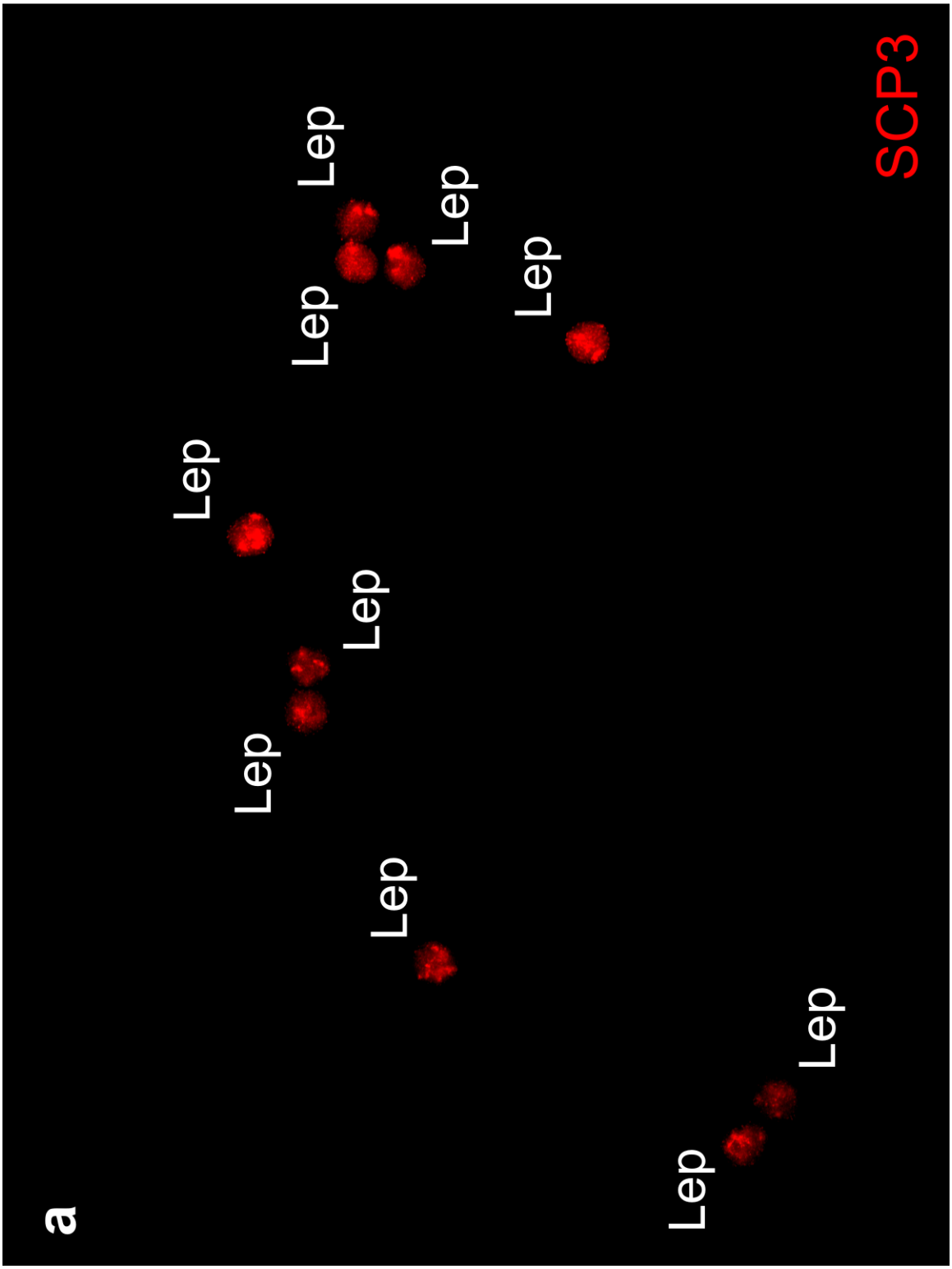
899 54. Mudge, J. M. & Harrow, J. Creating reference gene annotation for the mouse
900 C57BL6/J genome assembly. *Mamm. Genome* **26**, 366–378 (2015).

901 55. kevbrick. *kevbrick/lametal2019: Analytic pipeline for Lam et al. (Nature*
902 *Communications revision)*. (2019). doi:10.5281/zenodo.2656746

903 56. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible
904 platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–91
905 (2014).

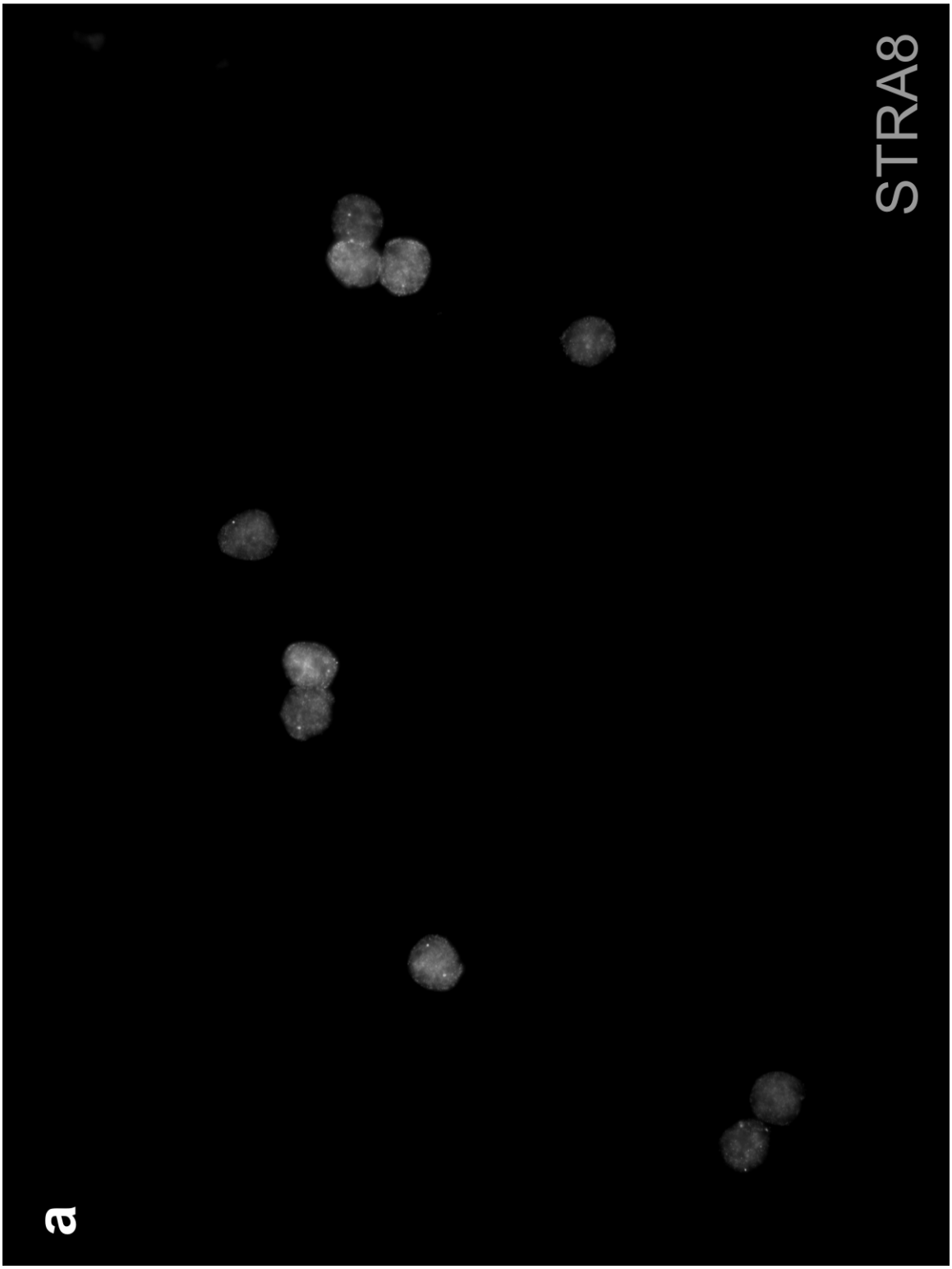
906

907 **Supplementary Figures**

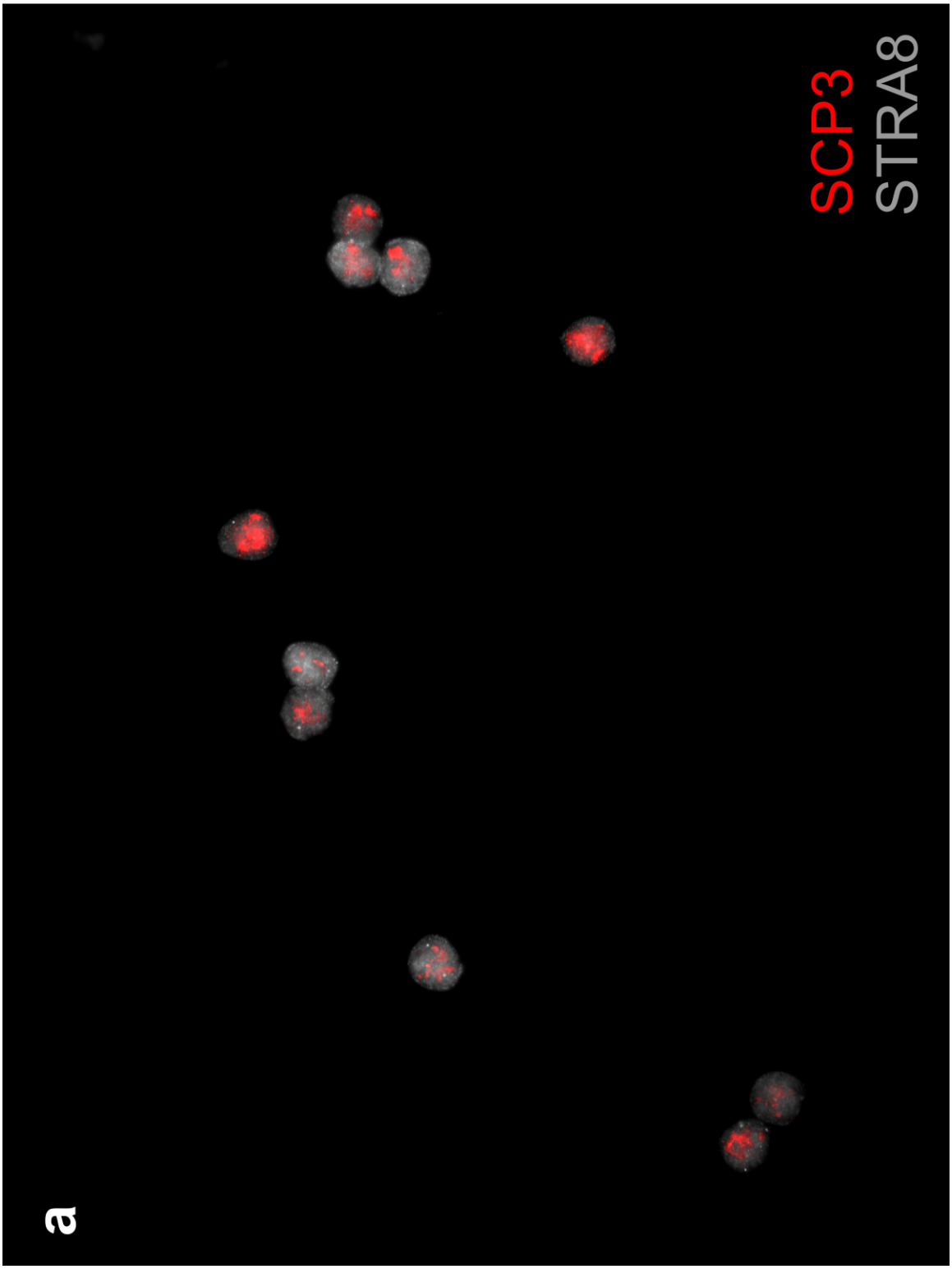


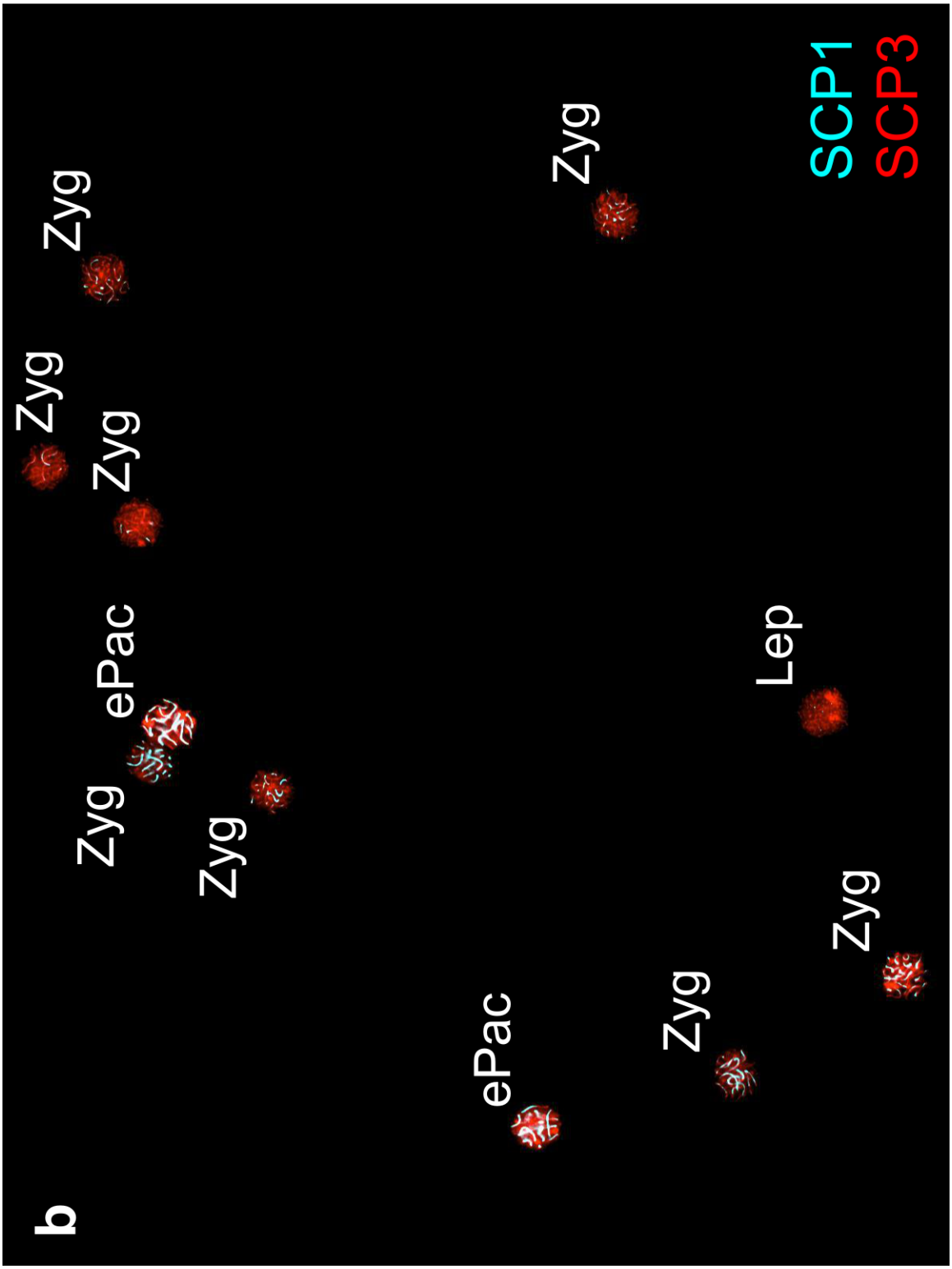
908

909

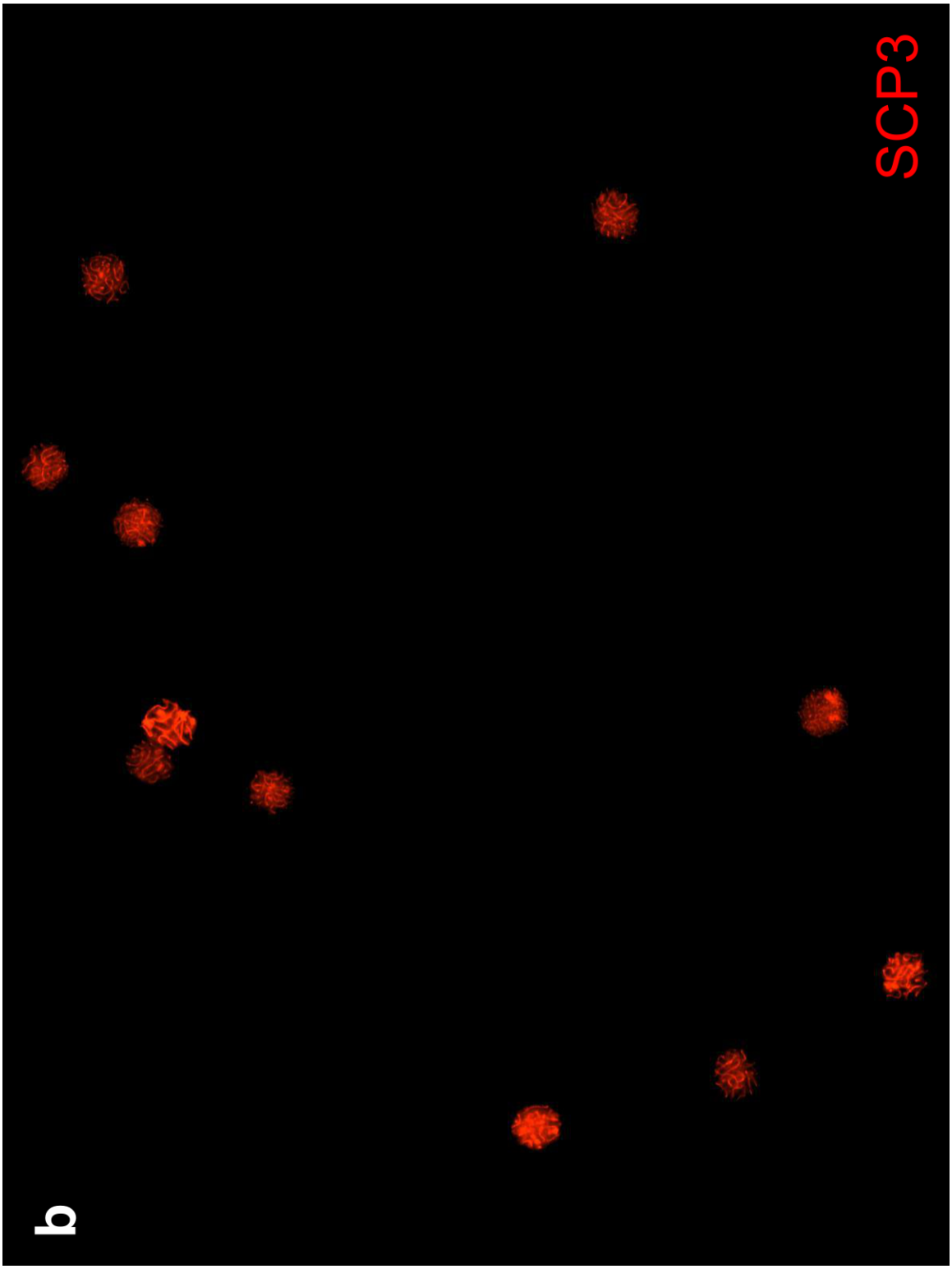


910

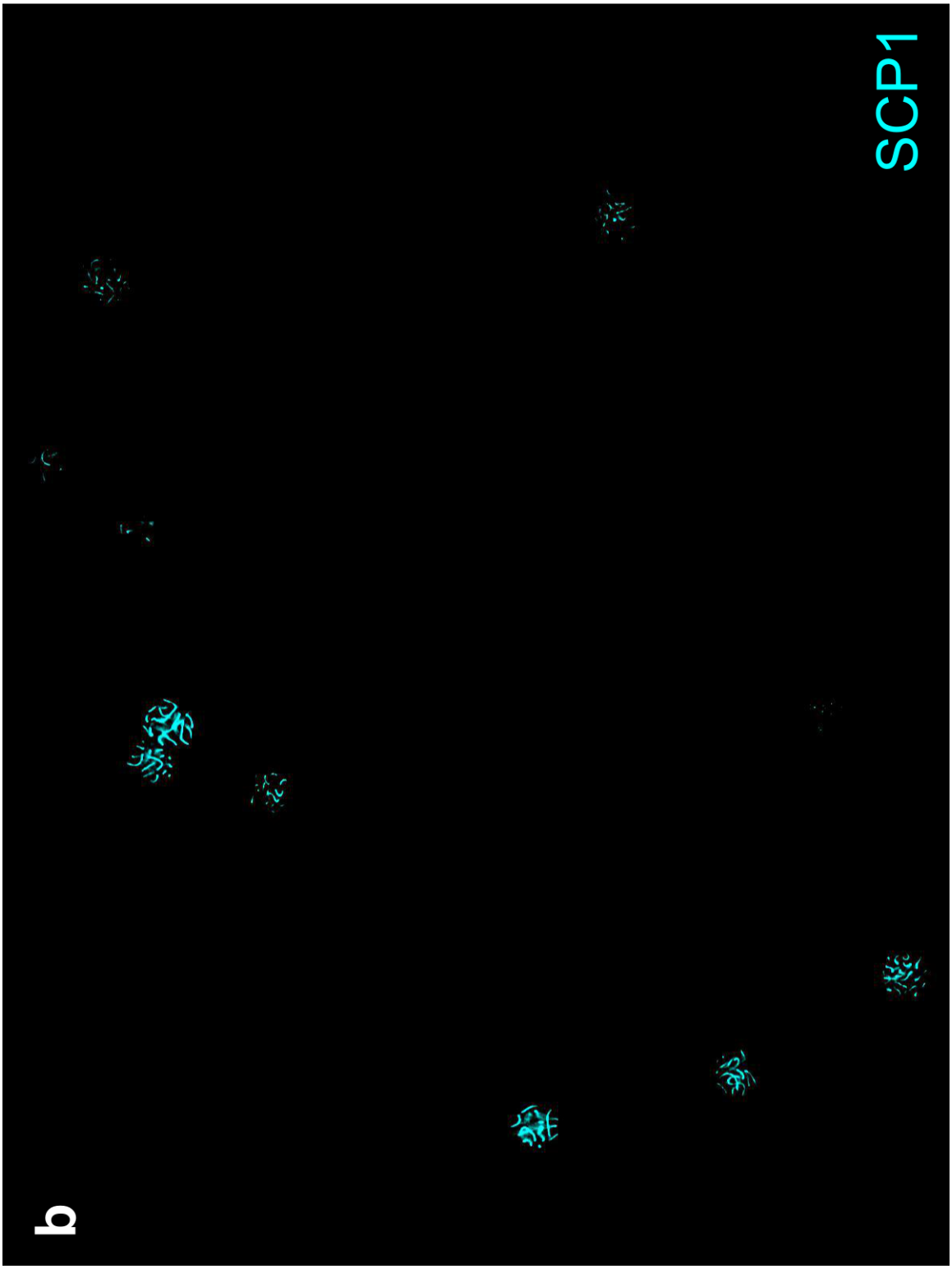




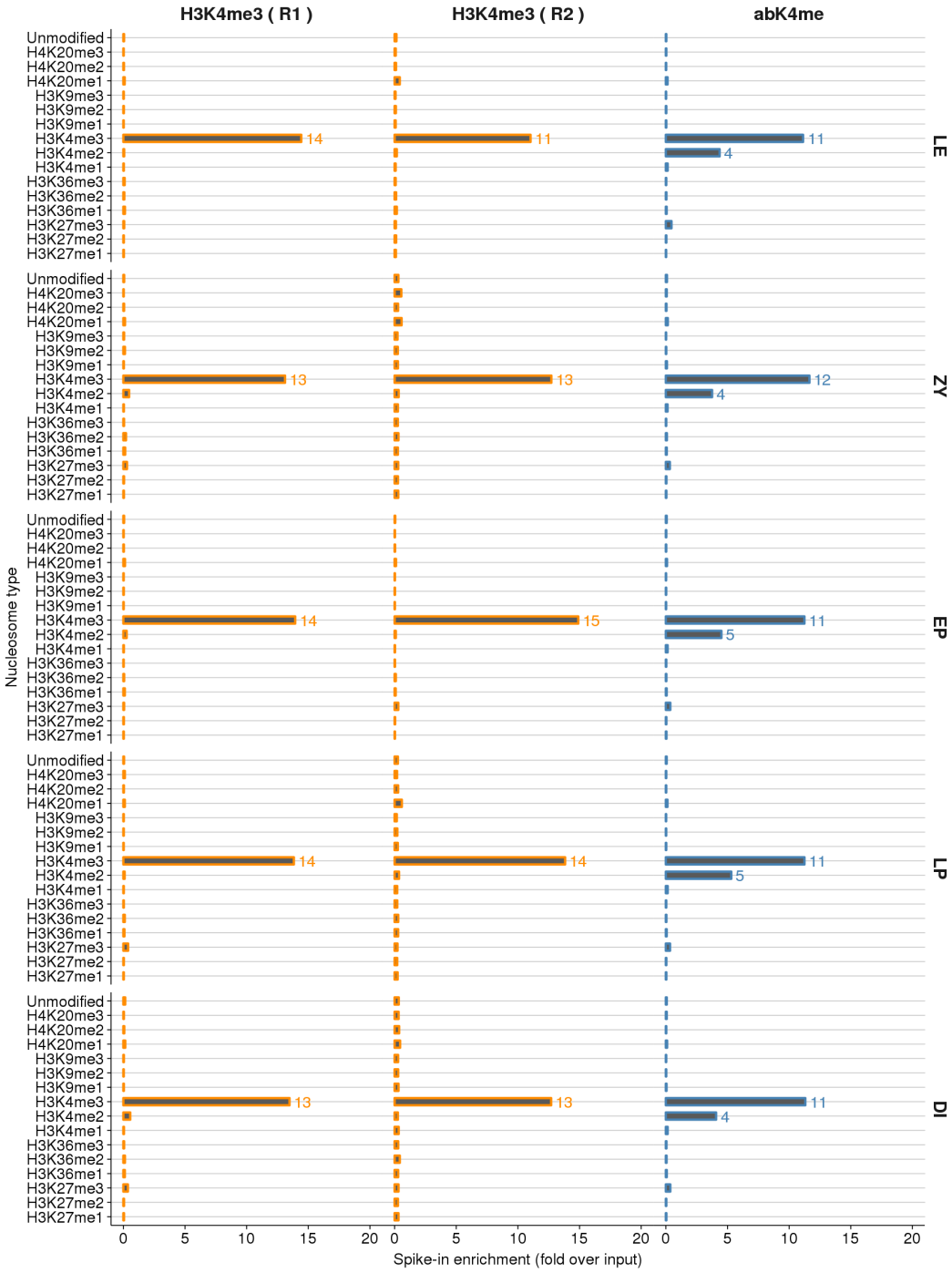
911



912


SCP1
b

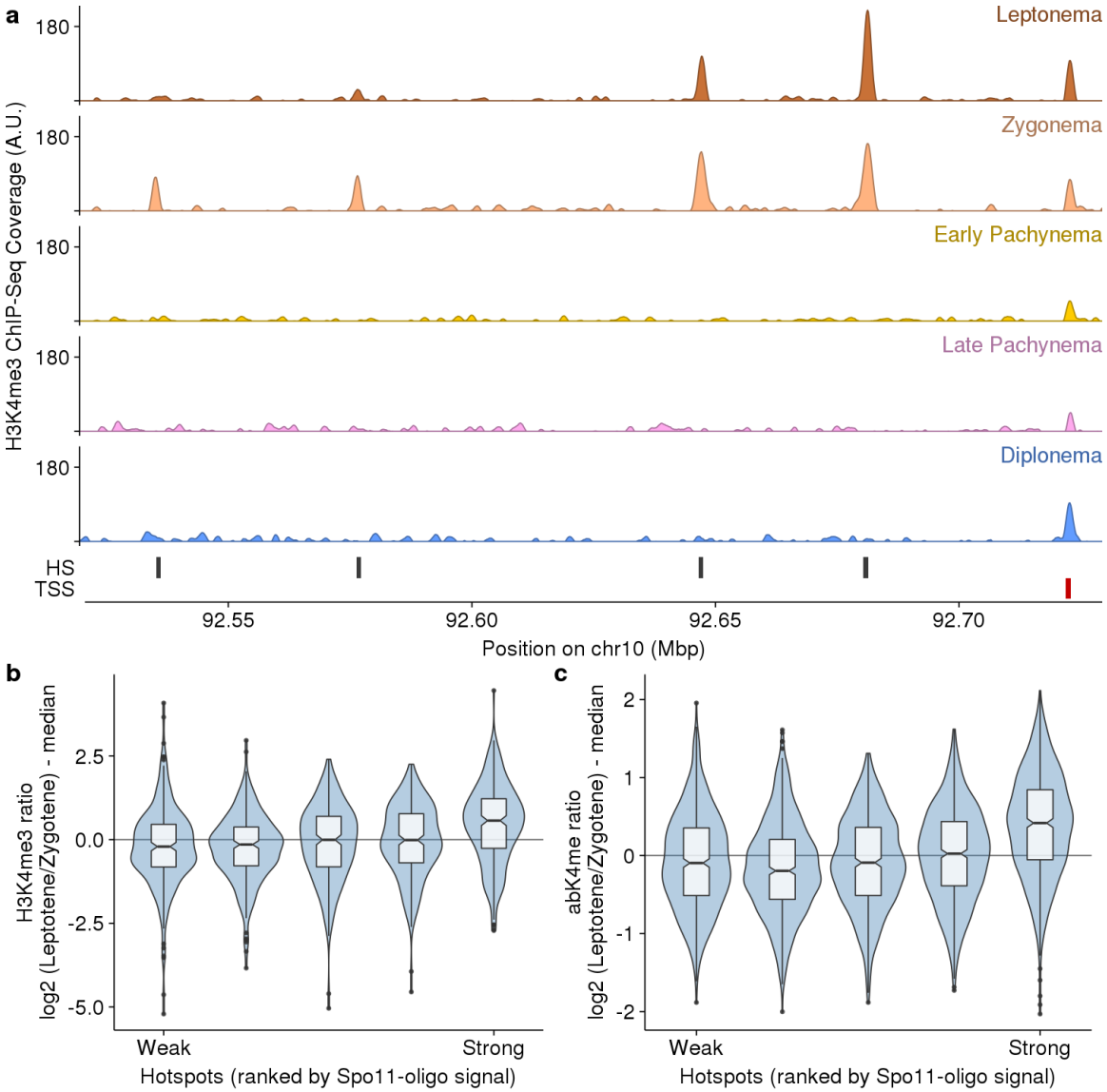
917 **Supplementary Fig. 1** Microscopic examination of sorted nuclei after FACS.
918 Whole-field immunofluorescence images of sorted nuclei in the populations of **a**
919 leptonema and **b** zygonema. Lep=leptonema; Zyg=zygonema; ePac=early
920 pachynema.
921
922



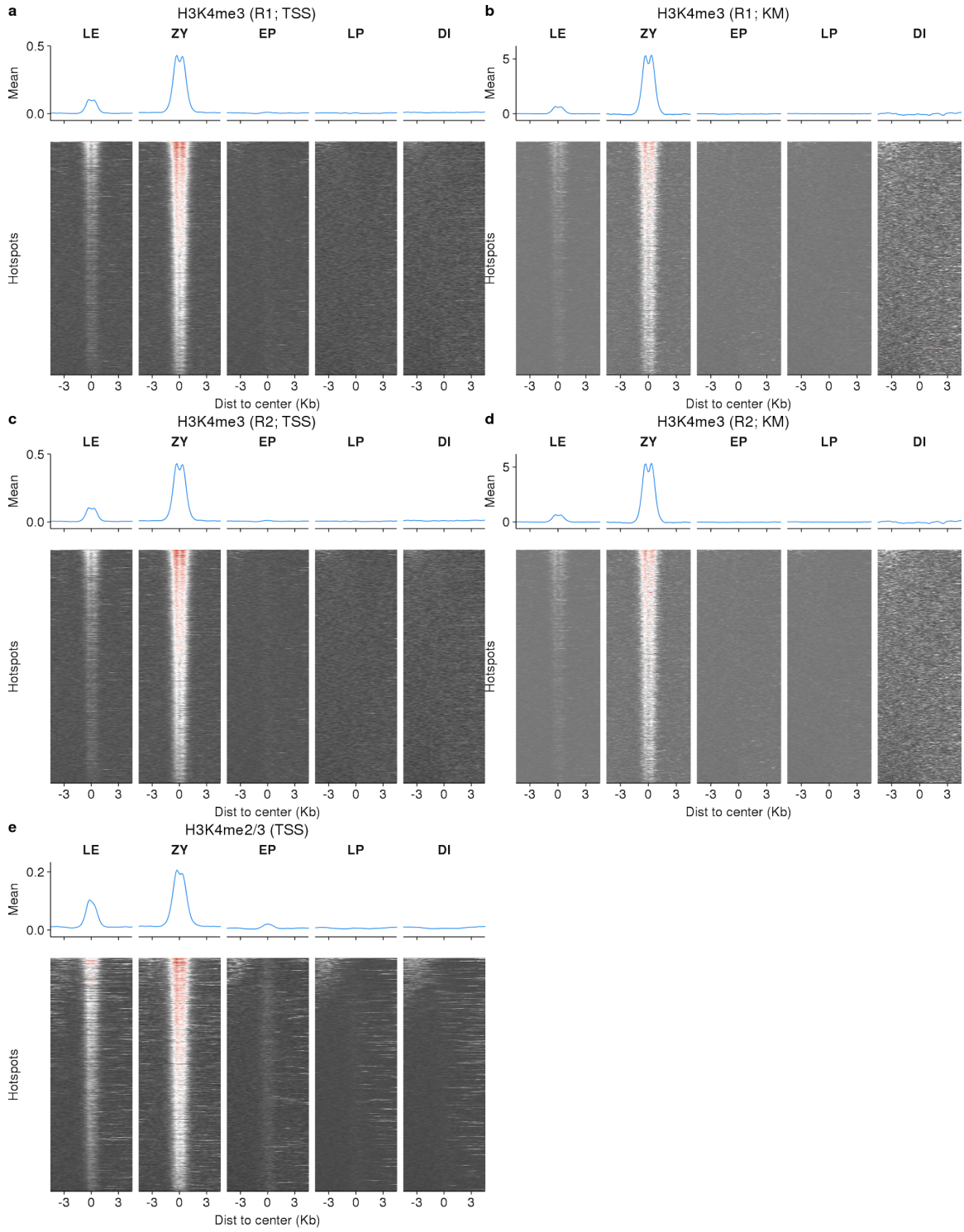
Supplementary Fig. 2 H3K4me3 antibodies have varying levels of substrate specificity.

We performed ChIP-Seq across five MPI stages using two H3K4me3 antibodies; EpiCypher 13-0028 (H3K4me3; two replicates) and Abcam ab8580 (abK4me). The EpiCypher K-MetStat panel is a collection of nucleosomes with different histone modifications that was spiked into each H3K4me3 ChIP-Seq experiment to validate the specificity of the antibody and to allow cross sample normalization

931 for ChIP-efficiency (see methods). The relative enrichment of sequences bound
932 to each modified nucleosome are shown here, compared to the relative
933 frequency in the sample before ChIP was performed. The Epicypher antibody is
934 highly specific to H3K4me3, however, as previously demonstrated[31], the Abcam
935 antibody shows substantial cross-reactivity with H3K4me2. Note that although
936 the enrichment for H3K4me2 is relatively low compared to H3K4me3, the
937 relative abundance of H3K4me2 in the genome compared to H3K4me3[31] makes
938 it difficult to assess the quantitative impact of this cross-reactivity in ChIP-Seq
939 experiments.



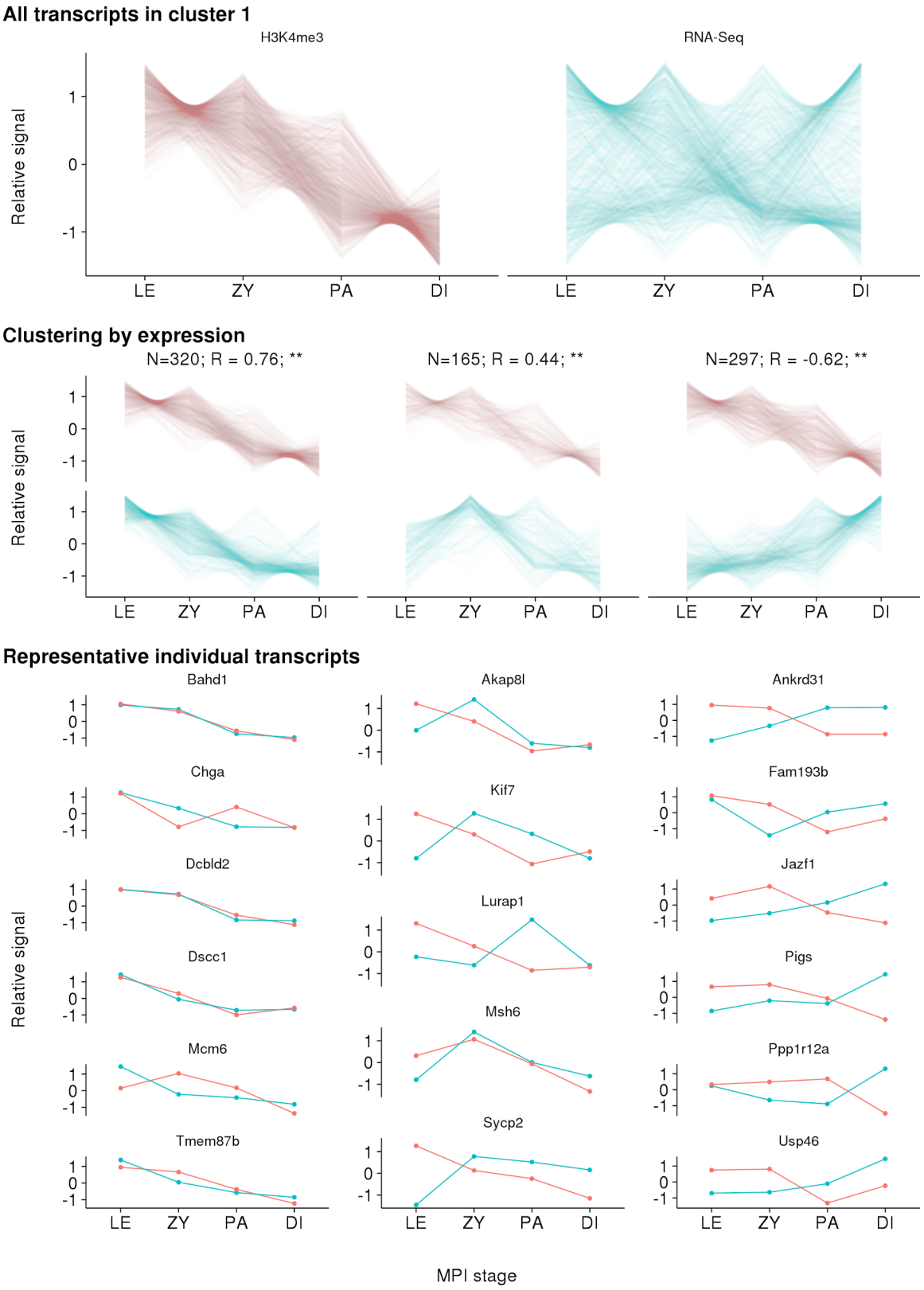
Supplementary Fig. 3 H3K4me3 is relatively elevated at stronger DSB hotspots in leptonema. **a** Snapshot of H3K4me3 at four DSB hotspots. The DSB hotspots are progressively stronger from left to right. At the three weaker hotspots (left) the H3K4me3 signal is maximal at zygonema. At the strongest hotspot (right) the H3K4me3 signal is strongest at leptonema. **b-c** Systematic trend towards stronger leptonema H3K4me3 at stronger DSB hotspots. **b** The H3K4me3 ratio between leptonema and zygonema is shown. **c** A more pronounced trend is seen using the Abcam ab8580 H3K4me3 antibody (abK4me).



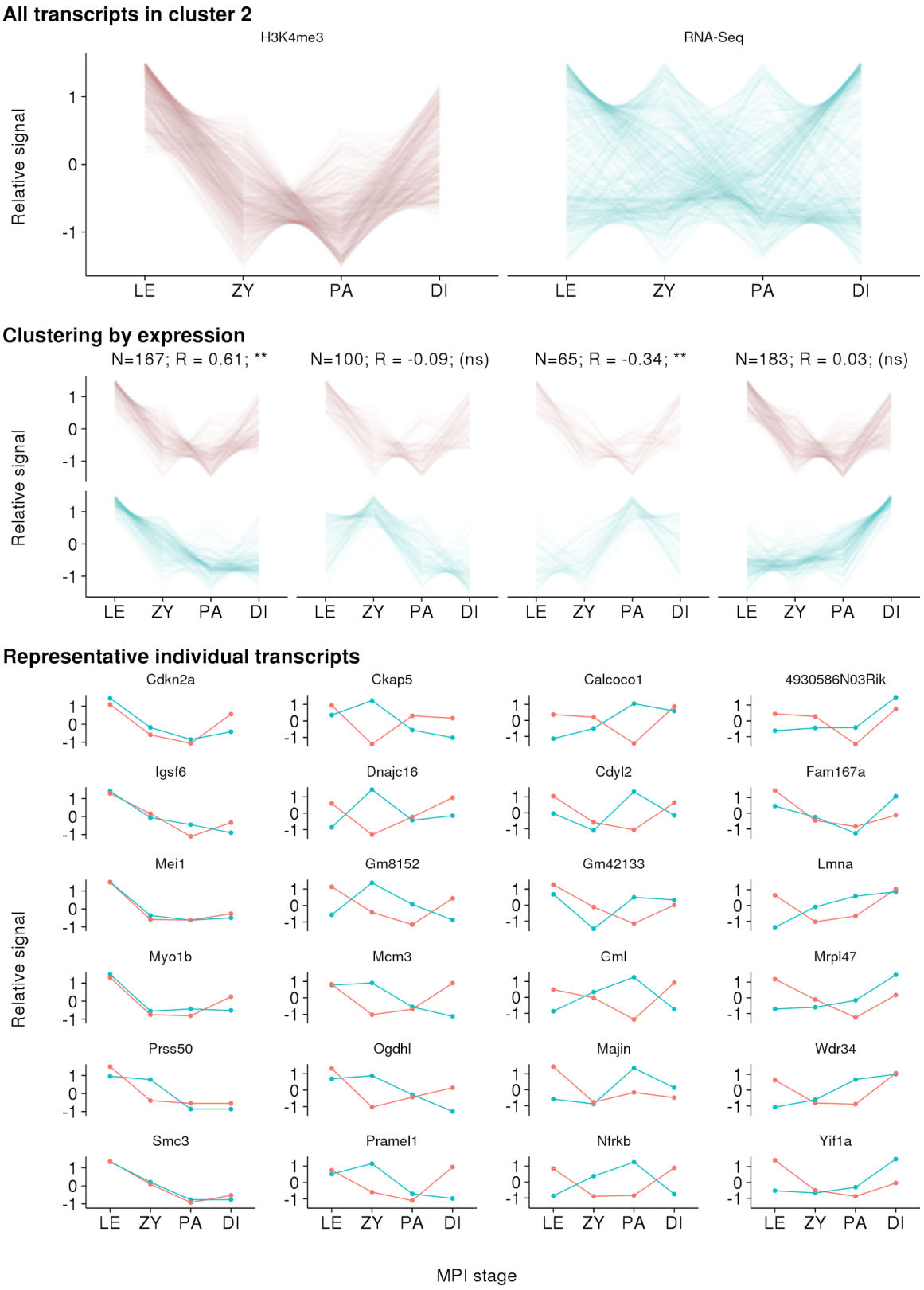
Supplementary Fig. 4 H3K4me3 dynamics are similar using different normalization approaches.

Coverage heatmaps at DSB hotspots through MPI. **a,c,e** Coverage is normalized using NCIS and stable TSS signals (see methods). **b,d** Coverage is normalized

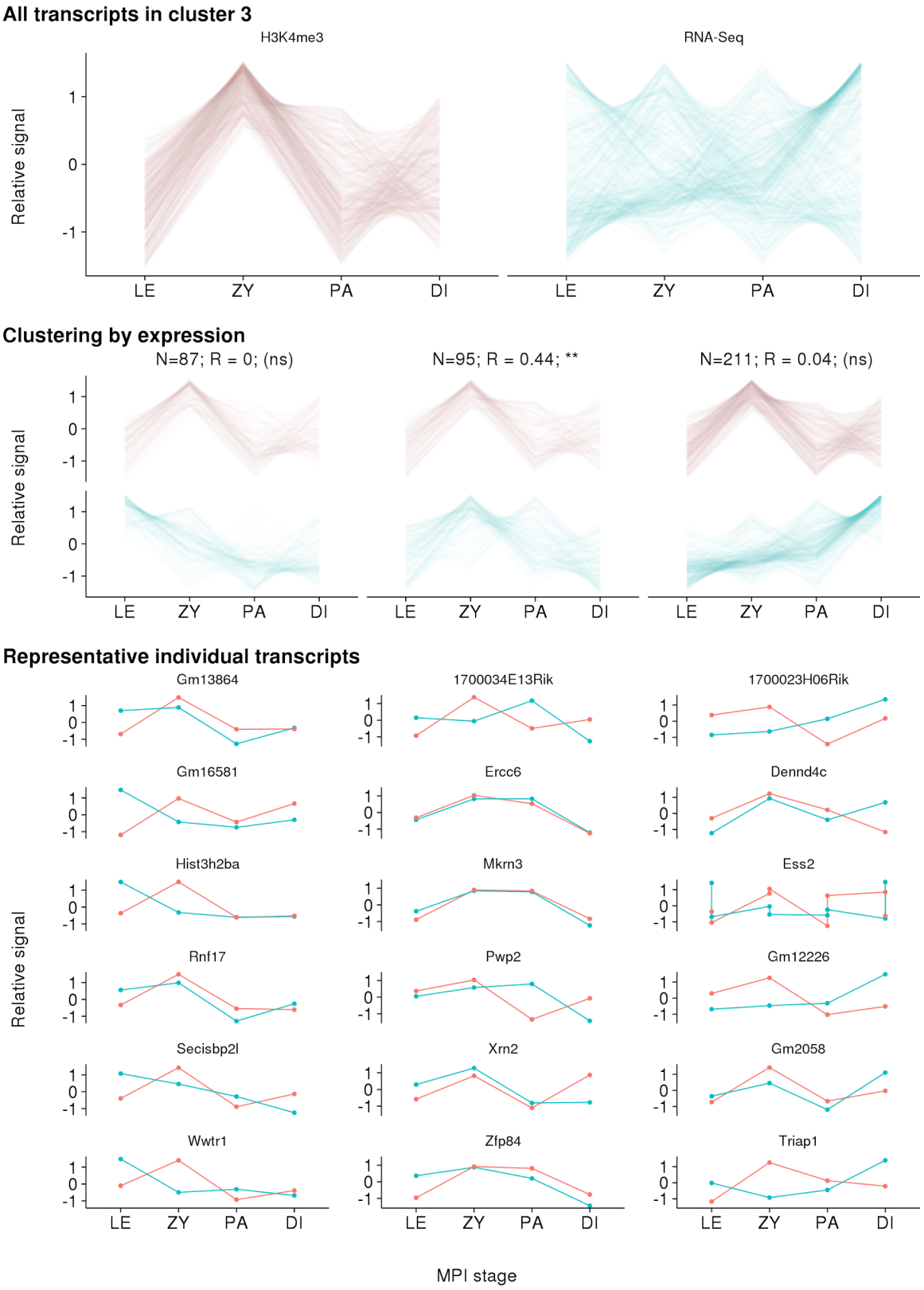
955 using the K-MetStat spike-in. The spike-in normalization cannot be used for the
956 H3K4me2/3 sample because of antibody cross-reactivity (Abcam ab8580
957 antibody). Both normalization methods yield similar results for H3K4me3 replicate
958 1 (R1; **a,b**) and replicate 2 (R2; **c,d**). A slight increase in the leptonema signal and
959 a marginal signal at early pachynema are seen with the H3K4me2/3 ChIP-Seq
960 (**e**).
961


All transcripts in cluster 1
H3K4me3
RNA-Seq
Relative signal
1
0
-1
LE
ZY
PA
DI
Clustering by expression
N=320; R = 0.76; **
N=165; R = 0.44; **
N=297; R = -0.62; **
Representative individual transcripts
Bahd1
Akap8l
Ankrd31
Chga
Kif7
Fam193b
Dcbld2
Lurap1
Jazf1
Dscc1
Pigs
Msh6
Mcm6
Ppp1r12a
Sycp2
Tmem87b
Usp46
MPI stage

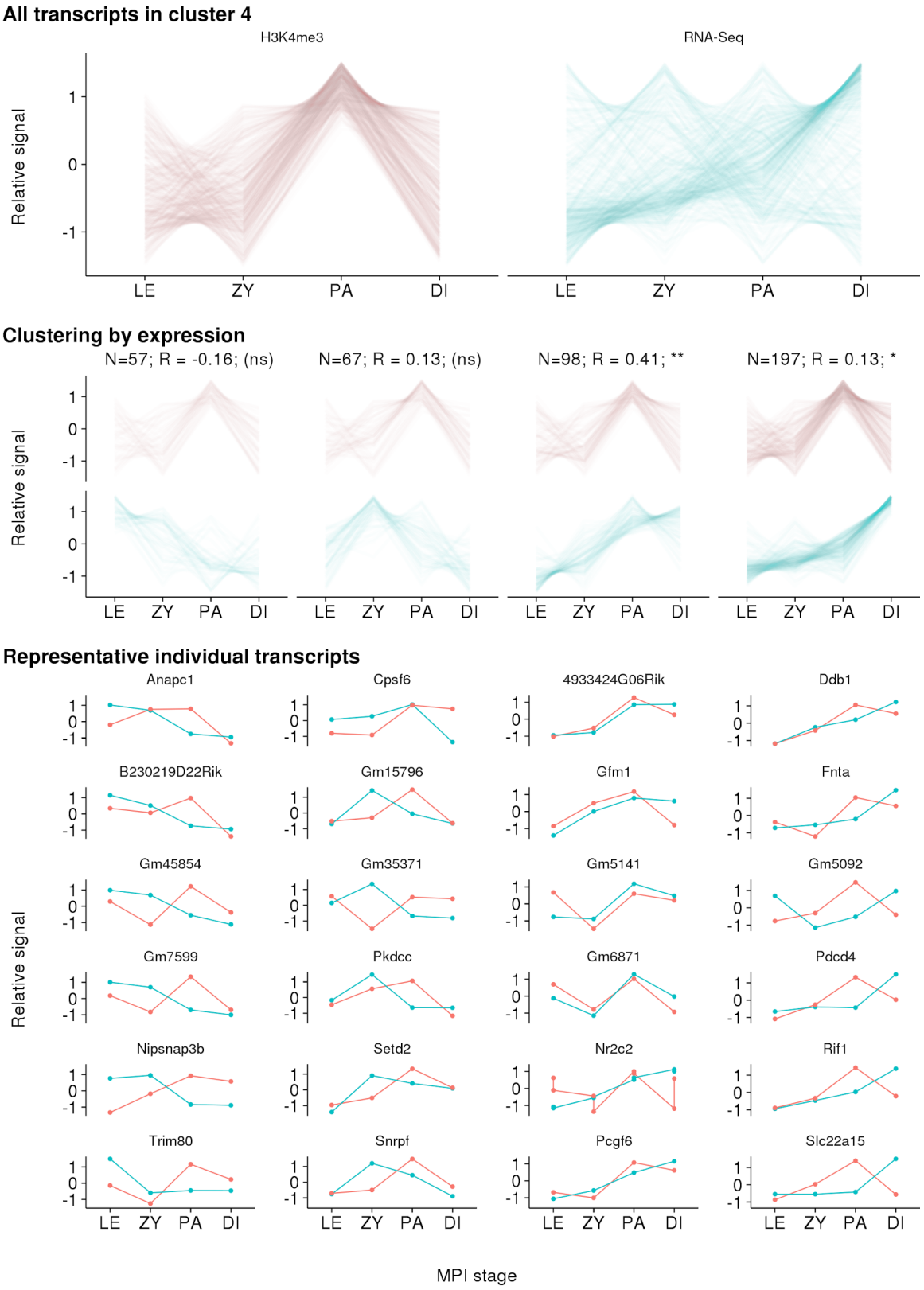
962


All transcripts in cluster 2
H3K4me3
RNA-Seq
Relative signal
1
0
-1
LE
ZY
PA
DI
Clustering by expression
N=167; R = 0.61; **
N=100; R = -0.09; (ns)
N=65; R = -0.34; **
N=183; R = 0.03; (ns)
Representative individual transcripts
Cdkn2a
Ckap5
Calcoco1
4930586N03Rik
Igsf6
Dnajc16
Cdyl2
Fam167a
Mei1
Gm8152
Gm42133
Lmna
Myo1b
Mcm3
Gml
Mrpl47
Prss50
Ogdhl
Majin
Wdr34
Smc3
Pramel1
Nfrkb
Yif1a
MPI stage

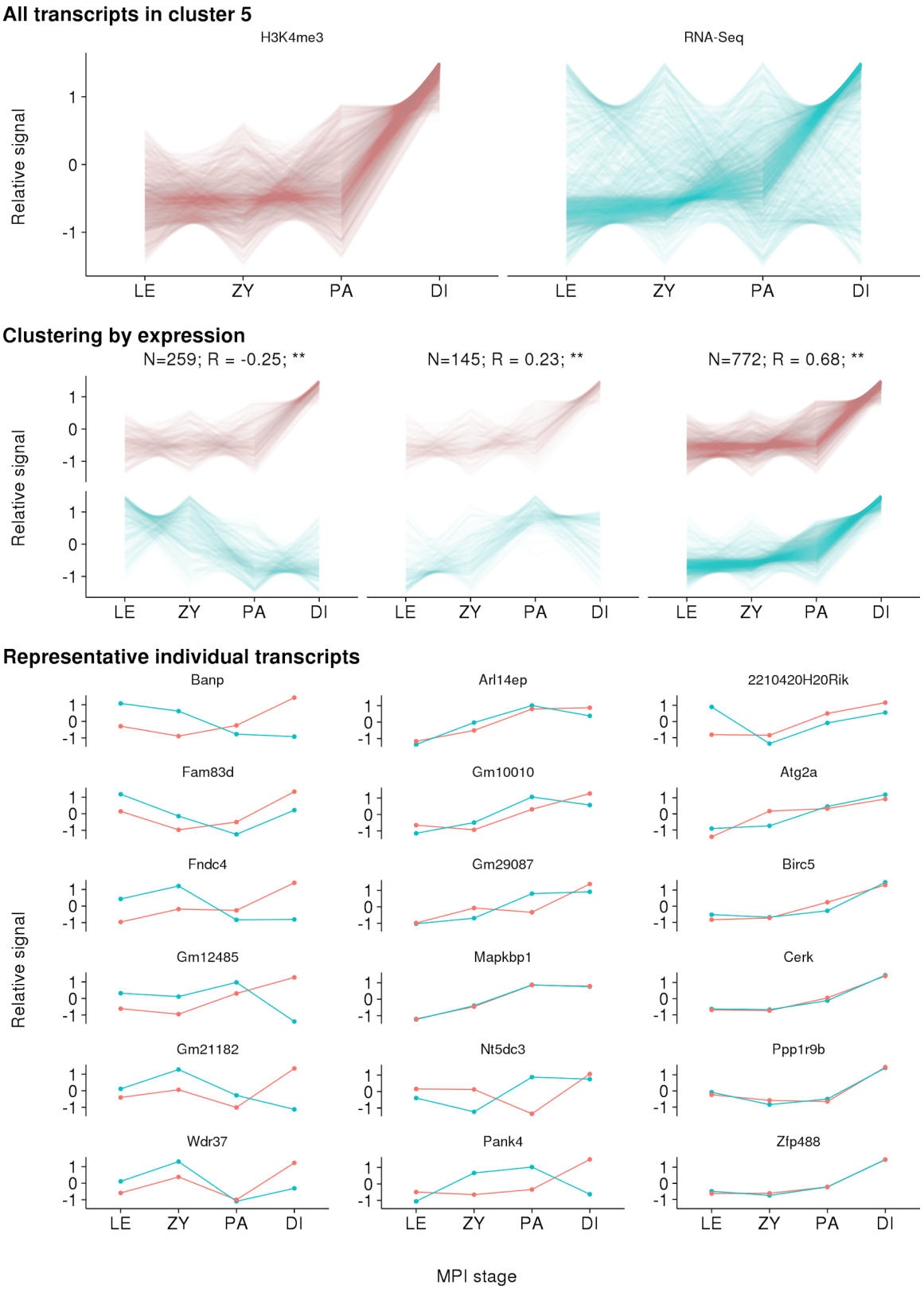
963


All transcripts in cluster 3
H3K4me3
RNA-Seq
Relative signal
1
0
-1
LE
ZY
PA
DI
Clustering by expression
N=87; R = 0; (ns)
N=95; R = 0.44; **
N=211; R = 0.04; (ns)
Representative individual transcripts
Gm13864
1700034E13Rik
1700023H06Rik
Gm16581
Ercc6
Dennd4c
Hist3h2ba
Mkrn3
Ess2
Rnf17
Pwp2
Gm12226
Secisbp2l
Xrn2
Gm2058
Wwtr1
Zfp84
Triap1
MPI stage

964

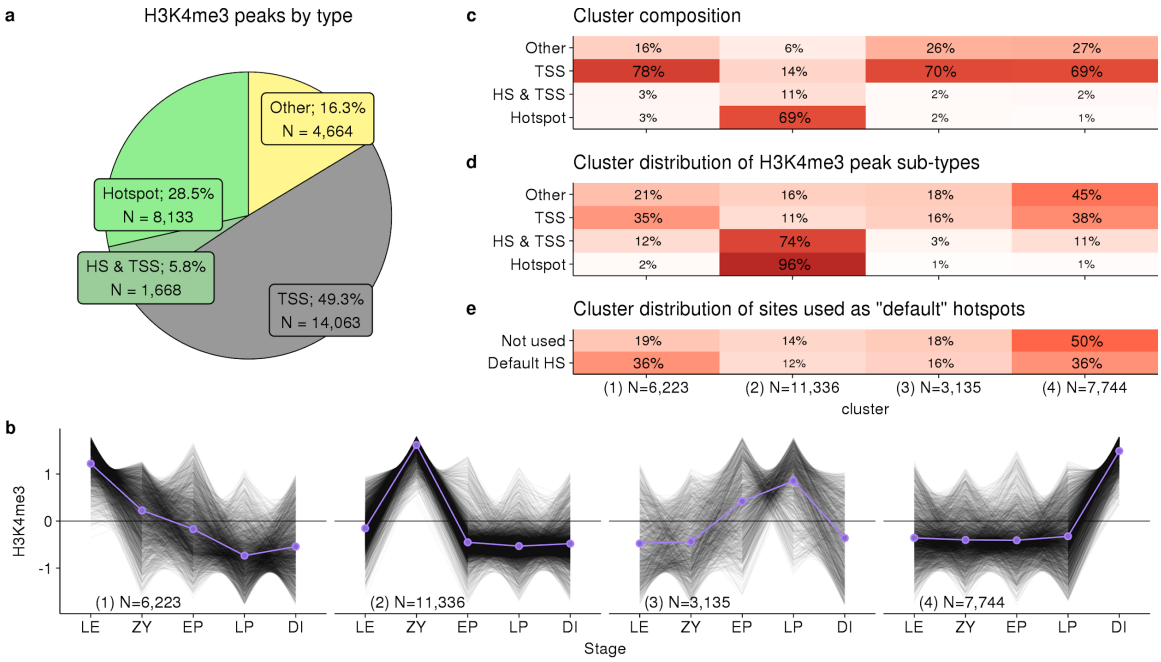

All transcripts in cluster 4
H3K4me3
RNA-Seq
Relative signal
LE
ZY
PA
DI
Clustering by expression
N=57; R = -0.16; (ns)
N=67; R = 0.13; (ns)
N=98; R = 0.41; **
N=197; R = 0.13; *
Representative individual transcripts
Anapc1
Cpsf6
4933424G06Rik
Ddb1
B230219D22Rik
Gm15796
Gfm1
Fnta
Gm45854
Gm35371
Gm5141
Gm5092
Gm7599
Pkdcc
Gm6871
Pdcd4
Nipsnap3b
Setd2
Nr2c2
Rif1
Trim80
Snrpf
Pcgf6
Slc22a15
MPI stage

965

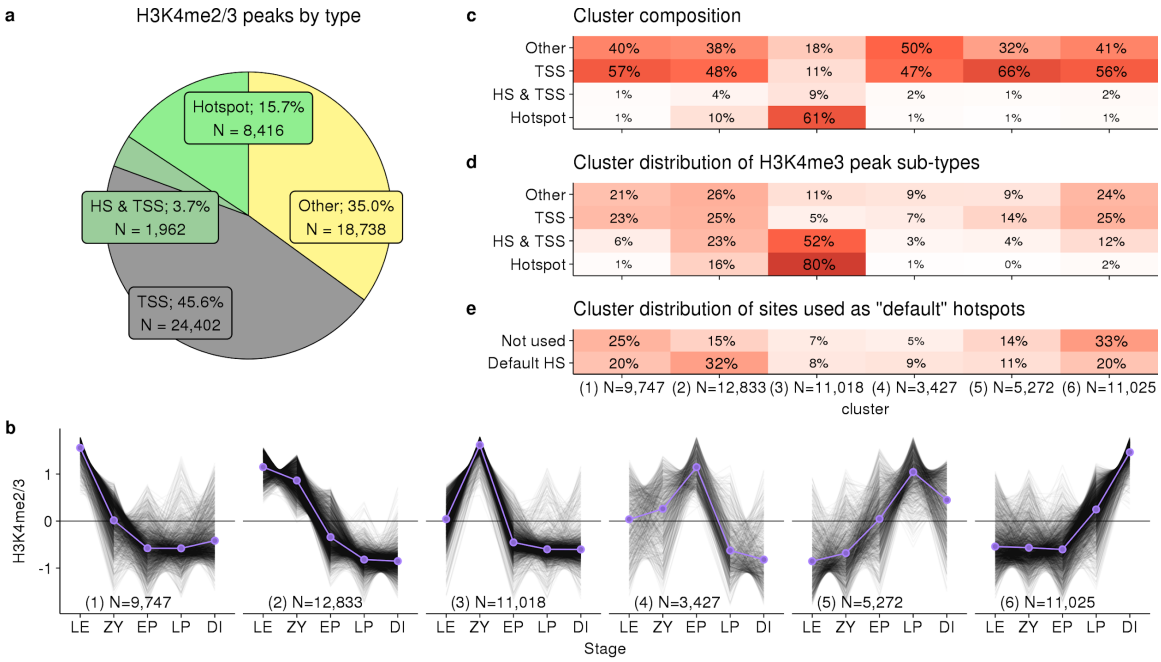

All transcripts in cluster 5
H3K4me3
RNA-Seq
Relative signal
1
0
-1
LE
ZY
PA
DI
Clustering by expression
N=259; R = -0.25; **
N=145; R = 0.23; **
N=772; R = 0.68; **
Representative individual transcripts
Banp
Arl14ep
2210420H20Rik
Fam83d
Gm10010
Atg2a
Fndc4
Gm29087
Birc5
Gm12485
Mapkbp1
Cerk
Gm21182
Nt5dc3
Ppp1r9b
Wdr37
Pank4
Zfp488
MPI stage

966

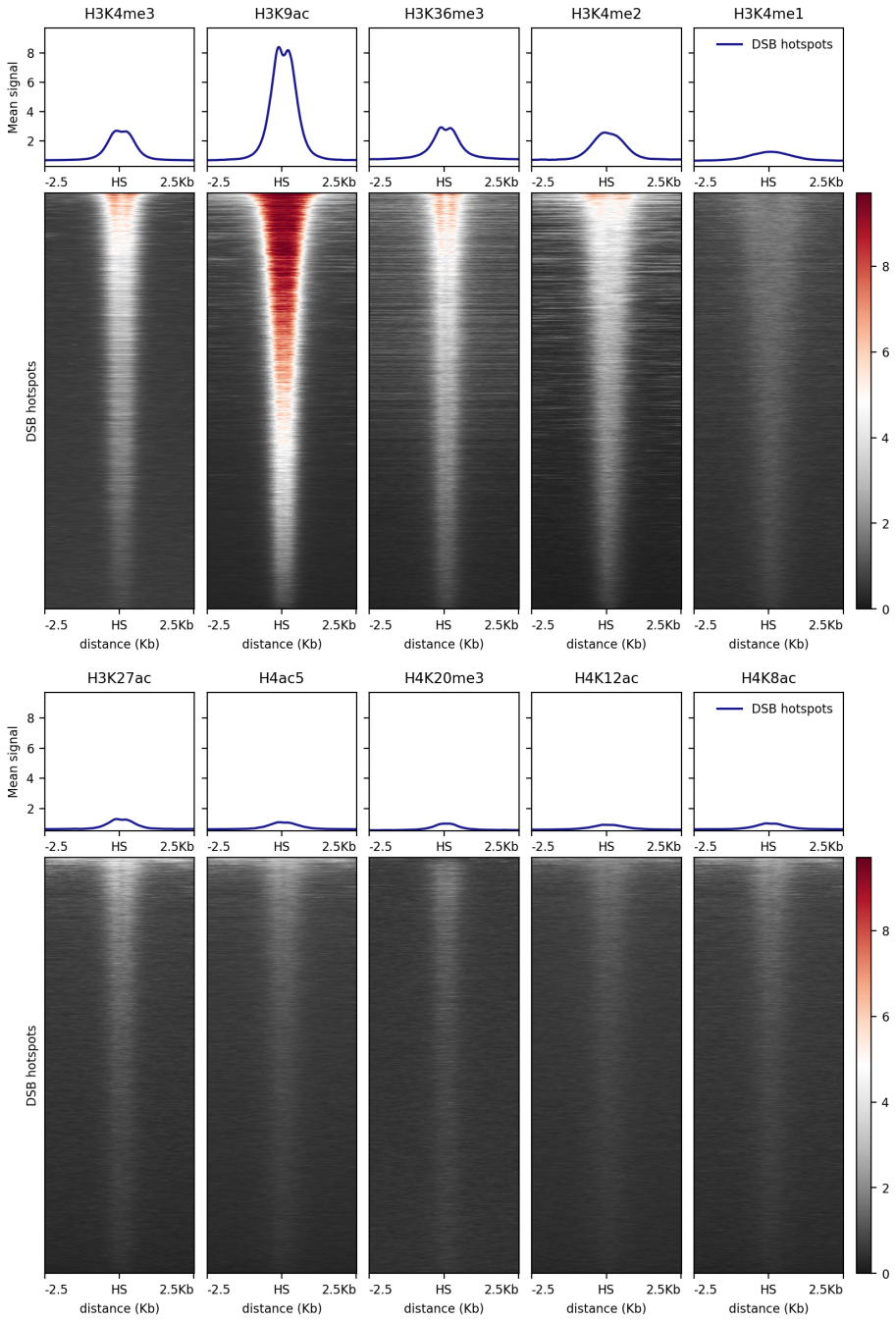
967 **Supplementary Fig. 5 H3K4me3 dynamics correlate with gene expression**
968 **profiles from male juvenile mice.**
969 Each of the plots represents a single cluster of transcripts from Fig. 3 (Cluster 1-5
970 are from left to right on Fig. 3c, respectively). Throughout, red depicts H3K4me3
971 ChIP-Seq signal and blue depicts gene expression from RNA-Seq. Uppermost
972 panels show all transcripts in the cluster; lines represent individual transcripts.
973 Transcripts were next sub-clustered by gene expression profile (middle panels; k-
974 means clustering; optimal k chosen using gap statistic). Clusters in which profiles
975 correlate more or less than expected are highlighted (* = $P < 0.01$, ** = $P < 0.001$;
976 1,000 bootstraps of shuffling stages for each gene). Six representative transcripts
977 from each cluster are shown in the lower panel below each cluster. Transcripts of
978 genes known to play a role in MPI were preferentially selected; if < 6 such genes
979 were present in a cluster, the remainder were randomly selected.



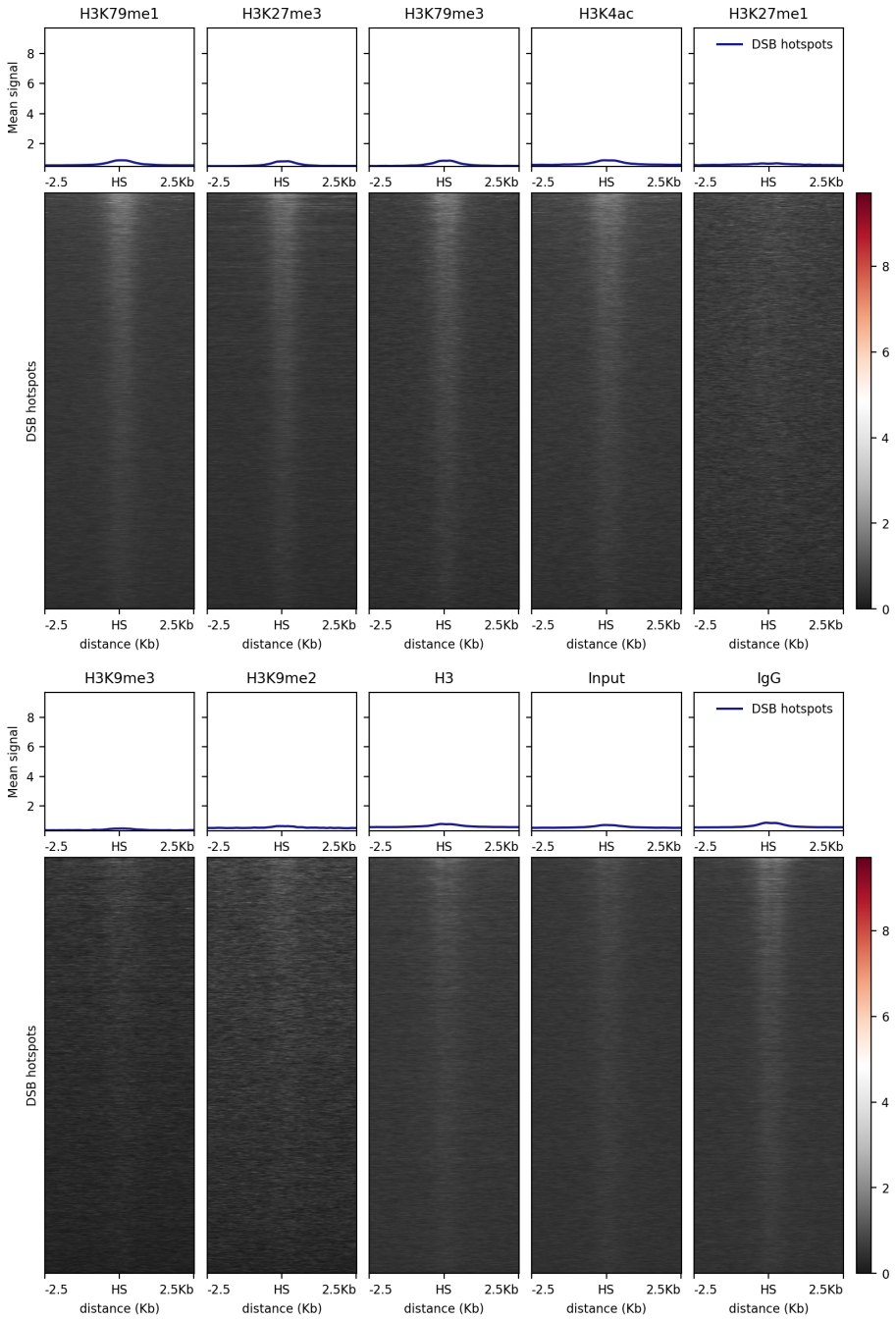
Supplementary Fig. 6 Extensive dynamics revealed by unbiased clustering of H3K4me3 peaks. a 16% of all H3K4me3 peaks occur at sites not defined at DSB hotspots or transcription start sites (TSS). These are classified as “Other”. 5.8% of H3K4me3 peaks coincide with both a DSB hotspot and a GENCODE TSS (HS & TSS). **b** Unbiased k-means clustering of MPI H3K4me3 profiles. Cluster number is indicated in parentheses in the bottom left corner. Each black line represents the MPI profile of an individual H3K4me3 peak. 10,000 randomly chosen profiles are shown. Purple lines depict the mean profile for each cluster. Profiles are normalized by the mean and standard deviation. The optimum number of clusters was identified using the gap statistic (optimum k = 5; range tested = 2 < k < 24). Cluster number is shown on each panel in parentheses. **c** Cluster composition shown as the percentage of each type of H3K4me3 peak within each cluster. Cluster 2 (N=11,336) is the only cluster composed primarily of peaks at DSB hotspots. **d** The percentage of peaks of each type across clusters is shown. Notably, 96% of DSB hotspots occur in cluster 2. 74% of HS & TSS peaks also occur in this cluster, suggesting that H3K4me3 at many of these sites is PRDM9-mediated. **e** In mice lacking PRDM9, PRDM9-independent H3K4me3 peaks are used for targeting the DSB machinery. These sites are termed “default hotspots” and are defined as H3K4me3 peaks that coincide with a DSB hotspot in *Prdm9*$^{-/-}$ mice[28]. Default hotspots form more frequently than expected at H3K4me3 peaks that are present in early MPI (cluster 1; N = 6,223). Peaks with a late dynamic (cluster 4; N=7,744) are used less than expected. The “Not used” peaks are all H3K4me3 peaks that do not coincide with either a PRDM9-defined or a default hotspot.



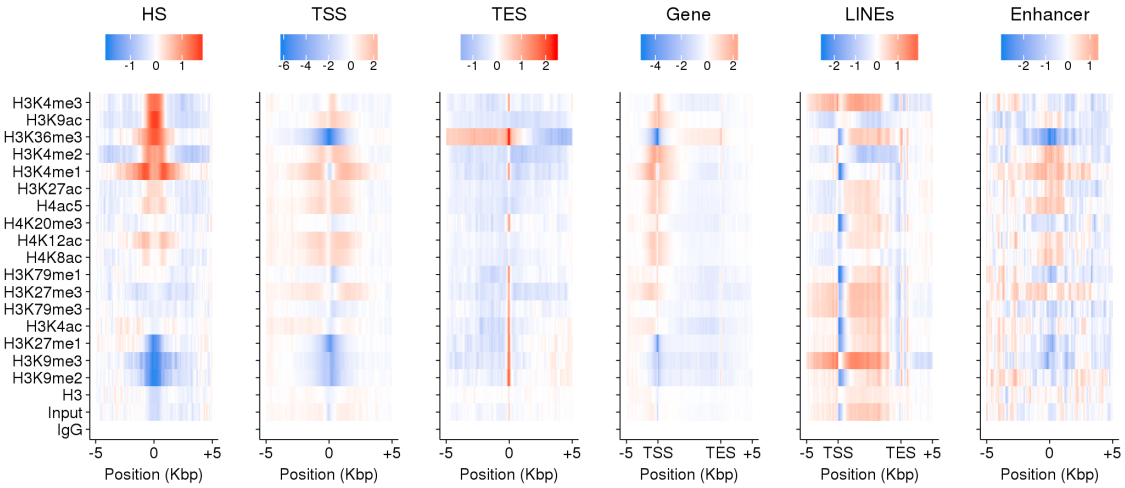
Supplementary Fig. 7 Extensive dynamics revealed by unbiased clustering of H3K4me2/3 peaks. a 35% of all H3K4me2/3 peaks (using Abcam ab8580 antibody) occur at sites not defined at DSB hotspots or transcription start sites (TSS). These are classified as “Other”. 3.7% of H3K4me3 peaks coincide with both a DSB hotspot and a GENCODE TSS (HS & TSS). **b** Unbiased k-means clustering of MPI H3K4me2/3 profiles. Cluster number is indicated in parentheses in the bottom left corner. Each black line represents the MPI profile of an individual H3K4me2/3 peak. 10,000 randomly chosen profiles are shown. Purple lines depict the mean profile for each cluster. Profiles are normalized by the mean and standard deviation. The optimum number of clusters was identified using the gap statistic (optimum k = 6; range tested = 2 < k < 24). Cluster number is shown on each panel in parentheses. **c** Cluster composition shown as the percentage of each type of H3K4me2/3 peak within each cluster. Cluster 3 is the only cluster composed primarily of peaks at DSB hotspots. **d** The percentage of peaks of each type across clusters is shown. Notably, 80% of DSB hotspots occur in cluster 3. 52% of HS & TSS peaks also occur in this cluster, suggesting that H3K4me3 at many of these sites is PRDM9-mediated. **e** In mice lacking PRDM9, PRDM9-independent H3K4me3 peaks are used for targeting the DSB machinery. These sites are termed “default hotspots” and are defined as H3K4me3 peaks that coincide with a DSB hotspot in *Prdm9*$^{-/-}$ mice[28]. Default hotspots form more frequently than expected at H3K4me3 peaks that are present in early MPI (cluster 2). Peaks with a late dynamic (cluster 5, 6) are used less than expected. “Not used” peaks are all H3K4me3 peaks that do not coincide with either a PRDM9-defined or a default hotspot.


H3K4me3
H3K9ac
H3K36me3
H3K4me2
H3K4me1
DSB hotspots
Mean signal
DSB hotspots
-2.5
HS
2.5Kb
distance (Kb)
H3K27ac
H4ac5
H4K20me3
H4K12ac
H4K8ac

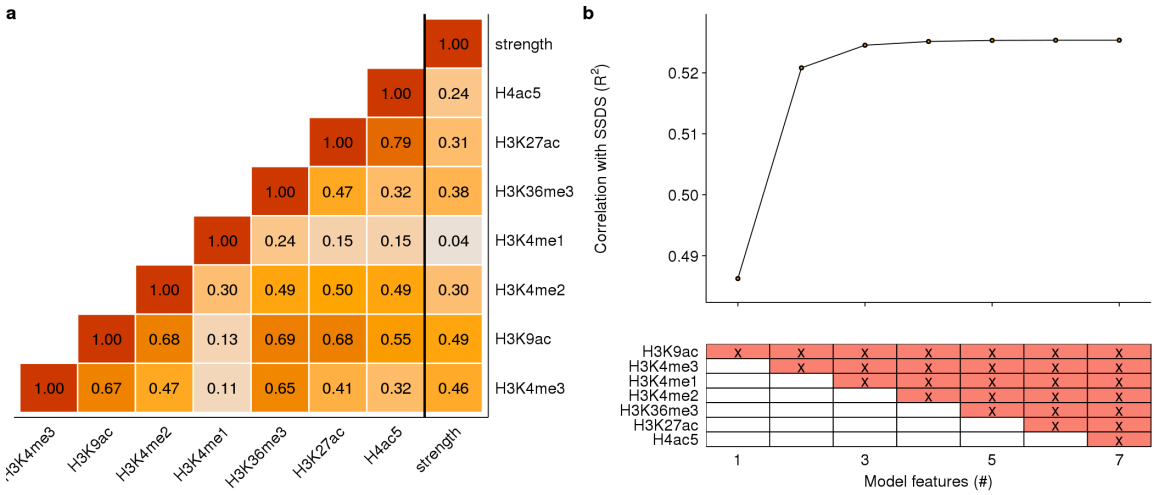
1030



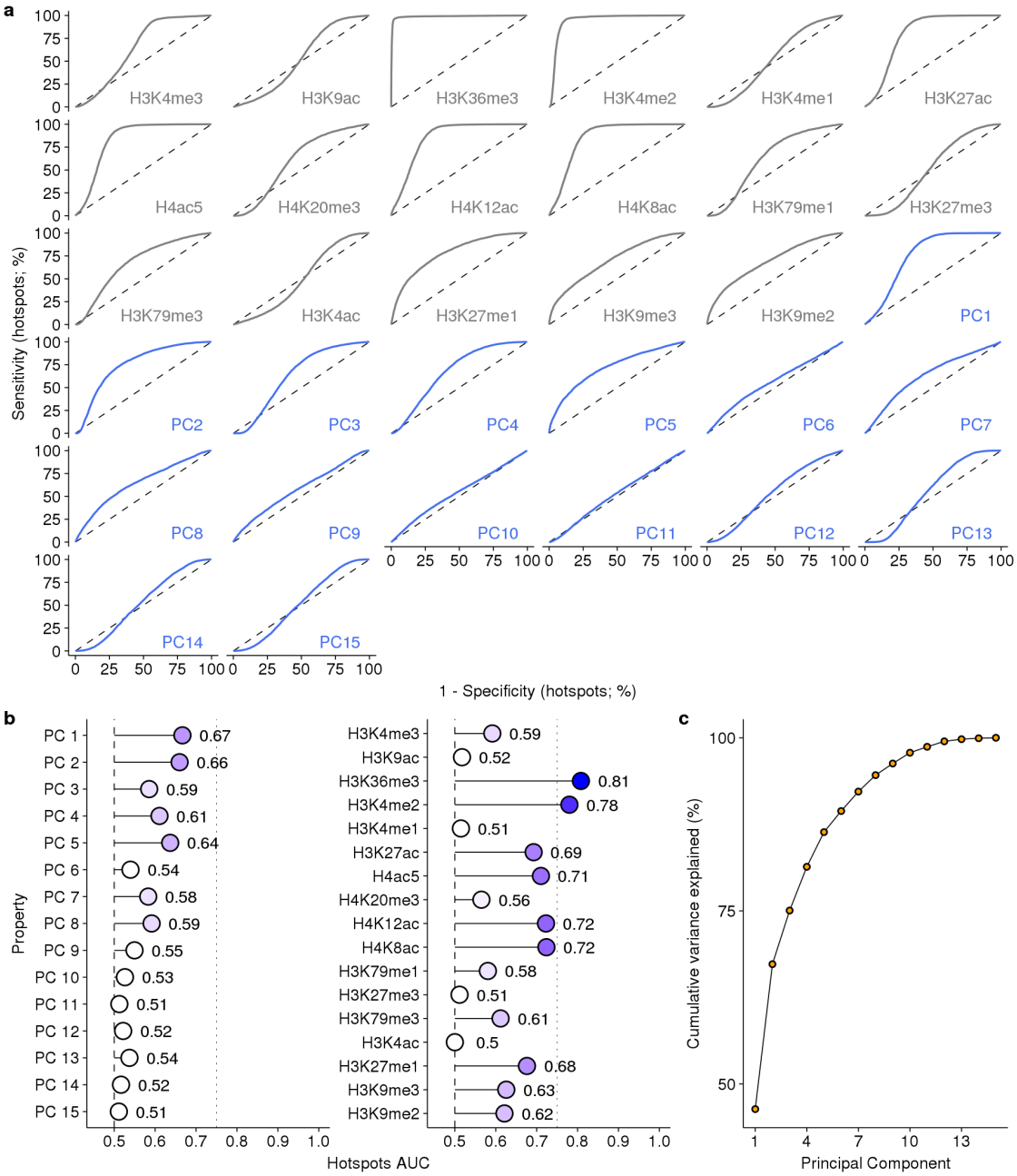
Supplementary Fig. 8 Heatmaps of histone modification ChIP-Seq at DSB hotspots. Coverage data were averaged in 150 bp windows and converted to reads per Kbp per million (RPKM) to facilitate cross-comparison. Heatmaps were generated using deeptools[56]. It is important to note that some enrichment at DSB hotspots is seen when using a non-specific antibody (IgG; last panel).



Supplementary Fig. 9 Enrichment of histone modifications at functional genomic elements. Histone modification coverage was plotted around functional sites in the genome. Enrichment is shown as mean signal - mean signal in IgG ChIP-Seq (red=high; blue=low). L1 LINEs were obtained from the repeatmasker database. Enhancers were obtained from the UCSC table browser RefSeq functional elements table.



Supplementary Fig. 10 Correlations between histone modifications and DSB hotspot strength. a Pearson correlation coefficients of log-transformed read counts and hotspot strength (strength). Only histone marks enriched at DSB hotspots were considered. Hotspots with above-zero coverage for all histone marks and hotspot strength were used (N = 11,253). **b** Multiple linear regression with histone marks only marginally improves correlation with DSB hotspots. The lower panel depicts the histone marks used (red; x) for models with each number of features. The order of addition was determined using all-subsets regression with the leaps package in R. Small increases in R^2 may simply arise because additional samples reduce stochastic noise in strength estimates.



Supplementary Fig. 11: Histone marks distinguish hotspots from other H3K4me3 sites.

a ROC curves for individual histone marks and principal components (PCs) for discriminating hotspots from other H3K4me3 sites in the genome. Dashed line indicates no predictive ability. **b** The area under the ROC curves (AUC). **c** Most of the variance in histone mark signals is captured by few PCs.