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 simple method for isolating pure sub-populations of meiocytes 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 simple, scalable framework for interrogating the complexities of mammalian meiosis.

Main

Meiosis is the specialized cell division required to form gametes. Despite its importance, 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*[1]. *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. As a result, technical challenges preclude the isolation and study of specific meiotic sub-stages. Here, we introduce a method to isolate and purify stage-specific meiotic nuclei. We leveraged this method to interrogate the histone modification landscape in meiotic prophase I (MPI) in unprecedented detail.

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[2,3]. 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[4,5] may preclude generalization to adults. Chemically induced synchronization of meiosis has recently been developed[6], 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[7–9]. 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 simple, yet highly specific strategy to isolate pure sub-populations of meiotic nuclei. This approach uses a variant of FACS[10] 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[11–14] 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 DNA double-strand breaks (DSBs) are targeted to sites defined by DNA binding and subsequent H3K4/K36-trimethylation by the PRDM9 protein[15,16]. PRDM9 gene expression[17], nuclear protein levels of PRDM9[18] and microscopy-based estimates of DSB formation[19,20] 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 likely to be functional elements, such as enhancers and cryptic promoters. These are the first genome-wide analyses of the dynamic chromatin landscape in meiosis. Studies of individual DSB hotspots have implied that H3K4/K36me3 alone do not fully describe the chromatin landscape at these sites[21]. Indeed, we found that the histone code at DSB hotspots is more complex than has been previously appreciated, with multiple histone tail modifications exhibiting enrichment. This

95 includes histone acetylation marks unlikely to have been deposited by the DSB-
96 defining histone methyltransferase PRDM9.
97
98 Together, these data demonstrate the simplicity, utility and feasibility of our
99 approach. This method opens the door to a far better understanding of the
100 molecular mechanisms underlying meiosis and other complex developmental
101 processes.
102

103 ## Results

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

105 We first devised a discriminative panel of intra-nuclear proteins to distinguish
106 between five, classically-defined MPI sub-stages; leptonema, zygonema, early
107 pachynema, late pachynema and diplonema (Fig. 1). Nuclei were extracted from
108 formaldehyde-fixed testes from adult male mice, then incubated with antibodies
109 against the desired proteins (see methods). Nuclei were also stained with DAPI
110 to allow measurement of DNA content. The combinatorial signals from these
111 fluorescent markers were subsequently used to isolate specific sub-populations
112 of spermatocyte nuclei using FACS (Fig. 2A and B; see methods).
113
114 We used the meiosis-specific protein SCP3 to distinguish post-replicative meiotic
115 prophase nuclei (4C) from fully replicated mitotic nuclei (also 4C). We then
116 combined SCP3 with other markers for FACS. Our combination of antibodies
117 required two separate sorts; one to isolate leptotene nuclei and a second to
118 isolate the other four sub-stages (Fig. 2A and B). In male mice, STRA8 regulates
119 the entry to meiotic S phase, and is expressed from primitive spermatogonia to
120 leptotene spermatocytes[14]. We therefore used STRA8 to isolate leptotene nuclei
121 from later stages. To isolate the four other meiotic stages, we used the presence
122 or absence of H1t and quantitative differences in SCP1 intensity (Fig. 1, Fig. 2A).
123 H1t (testis-specific histone H1) is present from mid pachynema through MPI[22],
124 whereas SCP1 is a marker of progressive chromosome synapsis from zygotene

125 to pachytene stage and the SCP1 signal is diminished in diplonema when
126 chromosomes desynapse[12] (Fig. 1). Thus, we collected seven populations,
127 assessed the purity of each population and determined the gating thresholds for
128 the four meiotic sub-stages (Fig. 2B and C).
129
130 The purity of each sorted population was assessed using immunofluorescence
131 microscopy to stage the nuclei according to the known localization of markers
132 (Fig. 1, Fig. 2A and C, Supplementary Fig. 1, see methods)[11–14]. The purity of the
133 sorted leptotene population was 94.6% - 97.9% (n = 2), a >20-fold increase in
134 purity compared to the percentage of leptotene cells in whole testis (4.1% - 5.5%
135 (n = 2), Fig. 2C). The purity of the other four populations ranged from 73.9 -
136 90.4%, each more than 10-fold enriched relative to the starting population from
137 whole testis (Fig. 2C). The least pure sample was that of zygotene (71.3% -
138 78.7% (n = 3)), likely because we rely entirely on quantitative differences in
139 SCP1 to distinguish between adjacent stages (Fig. 1, Fig. 2B and C). Narrowing
140 of the gate can increase purity but at the cost of yield (data not shown).
141 Quantitative thresholding issues also explain the 9.0% - 16.9% (n = 3) of
142 zygotene nuclei found in the sorted early pachytene population (Fig. 1, Fig. 2B
143 and C).
144
145 From a single adult mouse, we obtained between 115,000 and 400,000 nuclei for
146 each meiotic sub-stage (Fig. 2D). This range broadly reflects differences in the
147 proportion of each sub-stage in the adult testis, where leptotene and zygotene
148 nuclei are less abundant. Thus, the different sub-stages are isolated in sufficient
149 numbers for experiments using standard genomics protocols[23].
150
151 **Stage-specific ChIP-Seq reveals histone modification dynamics in meiosis**
152 We next used ChIP-Seq to examine the dynamics of H3K4me3 in MPI.
153 H3K4me3 is of key biological interest in meiosis because it regulates multiple
154 independent dynamic processes during MPI; H3K4me3 is found at gene
155 promoters and its presence correlates with active transcription[24,25], whereas

156 PRDM9-mediated H3K4me3 marks the future sites of meiotic DSBs genome-
157 wide[26,27]. We therefore performed H3K4me3 ChIP-Seq in the five
158 aforementioned sorted populations to assess the temporal dynamics of this
159 histone mark in MPI.
160
161 To establish the requirements for the amount of chromatin needed for ChIP-Seq,
162 we examined the ChIP-Seq quality of samples using different amounts of starting
163 nuclei (5,000–264,000). Between 7,517 and 47,804 H3K4me3 peaks were
164 detected in each sub-stage (Table 1). We merged these peaks to yield a super-
165 set of 58,862 H3K4me3 peaks present in at least one sub-stage. The Signal
166 Percentage of Tags (SPoT; a measure of the signal-to-noise ratio of ChIP-Seq
167 samples[28]), ranged from 13% to 50%, far exceeding ENCODE recommendations
168 of > 1%[29]. As expected, samples derived from fewer nuclei had a lower SPoT
169 and detected fewer H3K4me3 peaks (Table 1). Importantly, approximately
170 40,000–60,000 nuclei are sufficient to begin reaching saturation levels of peak
171 detection (Fig. 3A). Thus, sorted populations of meiotic nuclei are amenable to
172 ChIP-Seq with little, if any loss of sample quality.
173
174 To allow for quantitative cross-comparison of H3K4me3 signals from samples of
175 differing quality, we first normalized H3K4me3 peak strength in each sample by
176 the signal at 99 gene promoters at which the H3K4me3 signal was relatively
177 invariant across MPI (<1.2-fold change in all pairwise comparisons between
178 stages, see methods for details). We observed a highly dynamic H3K4me3 signal
179 at hotspots in MPI (Fig. 3B and C); the H3K4me3 signal first appears at
180 leptonema, and is maximal at zygonema, decreases in early pachynema and is
181 gone by late pachynema (Fig. 3C). The maximal signal at zygonema was
182 somewhat surprising because DSBs are formed and nuclear PRDM9 protein
183 levels are maximal at leptonema[18].
184
185 Intriguingly, at hotspots that are used more frequently for DSB formation,
186 H3K4me3 is relatively weak at zygonema compared to leptonema

187 (Supplementary Fig. 2A and B). This implies that DSB formation or repair
188 removes H3K4me3-modified nucleosomes, similar to findings in *Saccharomyces*
189 *cerevisae*[30]. PRDM9-mediated H3K4me3 decreases after zygonema, however,
190 the bias toward strong hotspots is not observed (Supplementary Fig. 2C). This
191 indicates that the H3K4me3 reduction may be caused by mechanisms other than
192 nucleosome eviction resulting from DSB formation or repair. Instead, PRDM9
193 may mark more sites than are used for DSB formation, and the decrease from
194 zygonema to early pachynema reflects systematic removal of excess marks. This
195 conclusion is reinforced by the strong H3K4me3 signal at early pachynema,
196 when DSB formation is unlikely[31,32].
197
198 The H3K4me3 level at gene promoters is correlated with gene expression in
199 mitotic cells[25]. We therefore investigated if the temporal dynamics of H3K4me3 at
200 transcription start sites (TSSs) is predictive of meiotic gene expression patterns.
201
202 Stage-specific gene expression in MPI has been deduced from classical cell
203 sorting of meiotic cells followed by RNA-Seq[9,33]. We found that the H3K4me3
204 profiles at TSSs were positively correlated with gene expression through MPI
205 (Fig. 3D, $R = 0.32$; Spearman test). This correlation was highly significant, far
206 higher than the maximum correlation observed in randomized comparisons
207 (empirical $P < 0.0001$; $N = 10{,}000$ bootstraps, $R_{max} = 0.03$, see methods). To
208 explore the relationship between H3K4me3 and gene expression in more detail,
209 we clustered transcripts based on the temporal H3K4me3 pattern (Fig. 3D;
210 optimal clusters = 3; k-means clustering using gap statistic[34]). There was
211 significant correlation with gene expression in all clusters ($P < 0.0001$; $N =$
212 $10{,}000$ bootstraps, range; $0.14 \leq R \leq 0.43$) (Fig. 3D). Each cluster is composed
213 of a large fraction of transcripts where H3K4me3 and gene expression correlate
214 very well. Nonetheless, there are also many transcripts that exhibit non-
215 correlated H3K4me3 and expression. This is particularly evident in cluster 1,
216 where many genes have higher mRNA levels late in MPI than we would predict

217 from H3K4me3 data (Supplementary Fig. 3). This may be explained by mRNA
218 accumulation through MPI, or by H3K4me3 marking poised, but not yet active
219 promoters[35]. Genes in cluster 3 are predictably less affected by these
220 confounding effects and the correlation is far higher for this cluster. Together,
221 these data demonstrate that ChIP-Seq of H3K4me3 in stage-specific nuclei
222 reveals a dynamic that parallels that of gene expression. These data also reveal
223 an intricate and poorly understood interplay between H3K4me3 and gene
224 expression during MPI.
225
226 **H3K4me3 dynamics at other genomic elements**
227 Dynamic H3K4-trimethylation is observed at both gene promoters and DSB
228 hotspots in MPI. However, almost half of the H3K4me3 peaks detected
229 (28,951/59,191) are outside of these sites (Supplementary Fig. 4A). These
230 unannotated H3K4me3 peaks may represent enhancers, cryptic promoters or
231 other functional elements. To assess the H3K4me3 dynamics of these peaks, we
232 derived an MPI profile for each H3K4me3 peak across the five MPI stages, then
233 performed unbiased k-means clustering. This approach was validated by the
234 presence of a large cluster that peaks at zygonema and that mostly contains
235 PRDM9-defined DSB hotspots (cluster 2; Supplementary Fig. 4B-D; 92% of
236 hotspots occur in this cluster). A subset of unannotated peaks (17%) also occur
237 in this cluster (cluster 2; Supplementary Fig. 4D) and may represent sites of
238 PRDM9 binding that were not detected in hotspot mapping experiments.
239 Alternatively, we cannot rule out that a cluster of cryptic functional sites simply
240 exhibit a similar dynamic. The remaining unannotated H3K4me3 peaks exhibit an
241 MPI dynamic that broadly mirrors that of H3K4me3 at TSS (Supplementary Fig.
242 4B, D). It seems likely therefore that functional elements at these sites play a role
243 in regulating meiotic progression. These dynamic, but unannotated sites
244 represent a completely unexplored aspect of the regulation of mammalian
245 meiosis.
246

247 Clustering also allowed us to study H3K4me3 dynamics at sites involved in other
248 important processes during MPI. In particular, we examined H3K4me3 sites that
249 are used for DSB targeting in mice that lack functional PRDM9[26,27]. H3K4me3
250 peaks used as "default" hotspot locations were predominantly those with peak
251 signal in early MPI (Supplementary Fig. 4E).
252
253 **Identifying other histone marks at hotspots**
254 At DSB hotspots defined by PRDM9, H3K4me3 is necessary for DSB
255 formation[36]. Nonetheless, H3K4me3 alone is not sufficient to define DSB sites,
256 as this histone mark is also present at other functional sites, such as gene
257 promoters. In mice, H3K36me3 is the only other histone mark described
258 genome-wide at DSB hotspots[15], while other histone marks may also be involved
259 in defining DSB sites[21,37,38]. The H3K36me3 ChIP-Seq signal at hotspots was
260 weak compared to H3K4me3 ChIP-Seq [15], thus, having the correct population to
261 interrogate appears important when exploring histone marks at hotspots. We
262 therefore used our method to isolate target nuclei from adult mice for a
263 systematic survey of multiple histone modifications.
264
265 We performed ChIP-Seq using 17 antibodies against histone methylation and
266 acetylation marks on isolated nuclei (Fig. 4A and B). These marks were either
267 previously analyzed at individual hotspots in mice[21,38] or reported to be enriched
268 at DSB sites in other organisms[39,40]. In this exploratory phase, we did not perform
269 ChIP-Seq in all five sub-populations, but instead, collected the SCP3+ and H1t–
270 population (leptonema to early pachynema) in which the H3K4me3 signal at
271 hotspots was apparent. As a positive control, H3K4me3 ChIP-Seq yielded a
272 comparable SPoT (3% at hotspots) to that in the stage-specific experiments
273 (Table 1).
274
275 H4K8ac, H4K12ac, H4K20me3, H3K4ac, H3K79me1, H3K79me3, H3K27me1,
276 H3K9me2, H3K9me3 and H3K27me3 were not enriched at hotspots relative to
277 controls (Fig. 4A, B; Supplementary Fig. 5). Most of these experiments showed

278 enrichment at other functional genomic regions (Supplementary Fig. 6), lending
279 confidence that lack of enrichment at hotspots is not the result of inefficient ChIP.
280 Several experiments (H3K9me2, H3K27me1 and H3K79me1) did not show
281 enrichment at other functional sites in the genome, therefore the lack of hotspot
282 signal may simply result from poor ChIP-quality.
283
284 H3K4me1, H3K27ac and H4ac5 showed marginal enrichment at hotspots (Fig.
285 4A, B; Supplementary Fig. 5). The weak signals at hotspots are unlikely the
286 result of sub-optimal ChIP-quality, as strong signals are seen at gene promoters
287 and enhancers (Supplementary Fig. 6). Instead, these marks may be rapidly
288 turned-over or present only in a sub-population of early spermatocytes. Indeed, a
289 weak DSB-dependent H4ac5 signal was previously reported at two mouse
290 hotspots[21].
291
292 PRDM9 can trimethylate both H3K4 and H3K36[36] and as expected, H3K4me3
293 and H3K36me3 were both enriched at hotspots (Fig. 4A and B; Supplementary
294 Fig. 5). H3K4me2 and H3K9ac were also strongly enriched at DSB hotspots
295 genome-wide; this clearly resolves the ambiguity from previous studies where
296 these marks were only enriched at one individual hotspot[21]. H3K4me2 may be an
297 intermediate of PRDM9 H3K4-trimethylation, however since PRDM9 lacks a
298 histone acetyltransferase domain, it is unlikely that PRDM9 directly acetylates
299 H3K9. In stage-specific experiments, the dynamics of H3K9ac through MPI
300 resemble those of H3K4me3; signal is first observed at hotspots in leptonema,
301 maximal at zygonema and progressively diminishes through diplonema (Fig. 4C).
302 Thus, H3K9ac is a *bona-fide* marker of meiotic DSB hotspots genome-wide,
303 explicitly demonstrating for the first time that proteins other than PRDM9 modify
304 nucleosomes at the sites of DSB hotspots.
305
306 H3K4me3 is positively correlated with DSB frequency[27,41], but most variation in
307 hotspot strength is not accounted for by changes in H3K4me3 (Spearman R^2 =
308 0.48) (Supplementary Fig. 7A). H3K36me3 was previously reported to slightly

309 improve the correlation between H3K4me3 and DSB strength[42], therefore we
310 tried to better predict hotspot strength using a combination of histone marks that
311 are enriched at DSB hotspots. Multiple linear regression with all histone marks
312 only slightly improved the correlation with hotspot strength (max R^2 = 0.51;
313 Supplementary Fig. 7B; see methods) suggesting that these extra histone marks
314 (including H3K36me3) do not offer an independent readout of DSB usage as
315 compared to H3K4me3.
316
317 **A histone code distinguishes DSB hotspots from PRDM9-independent**
318 **H3K4me3 sites**
319 In the absence of PRDM9, DSB hotspots occur at sites of non-PRDM9-mediated
320 H3K4me3[27]. Nonetheless, in *wild-type* mice, PRDM9-defined H3K4me3 sites are
321 used. In agreement with previous results[15,16], we found that H3K36me3 is a
322 potent discriminator between DSB hotspots and other H3K4me3-marked sites in
323 the genome (Fig. 4D). Among the other histones, H3K4me2 and H3K9ac can
324 best distinguish hotspots (Fig. 4D; Supplementary Fig. 8). We excluded histone
325 marks made by PRDM9 (H3K4me3 and H3K36me3), then performed principal
326 component analysis to explore if any combination(s) of histone marks better
327 define DSB hotspots. The most discriminative PC (PC3) captured signal from
328 H3K4me1, H3K4me2 and H3K9ac (Fig. 4E), all marks that are enriched at DSB
329 hotspots, however it did not discriminate hotspots any better than H3K4me2
330 alone. These data suggest that H3K36me3, H3K4me2, H3K9ac and to a lesser
331 extent, H3K4me1, distinguish the chromatin at sites of PRDM9-marked DSB
332 hotspots from that at other functional sites where H3K4 is tri-methylated.
333

334 ## Discussion

335 In this study, we developed a method to isolate highly pure populations of meiotic
336 nuclei from whole testis. This method is simple, rapid, requires very little starting
337 material, and resolves a major hurdle to studying meiosis in mammals. We have
338 demonstrated that we can efficiently sort five populations of MPI nuclei, however,

339 in principle, nuclei at any specific stage can be purified if the requisite antibodies
340 are available. We obtained populations of up to 96% purity for a given meiotic
341 sub-stage, negating the need to artificially synchronize meiosis using chemicals
342 or the need to use juvenile mice to obtain enriched populations of meiocytes.
343 This opens the door to many detailed studies of meiosis and also to experiments
344 in challenging-to-breed knockout mice and importantly in humans, where other
345 synchronization strategies are impractical.
346
347 Importantly, fixation prior to sorting allows for genome-wide interrogation of
348 transient protein-DNA interactions in sorted populations. To this end, we profiled
349 H3K4me3 using ChIP-Seq in five MPI populations, capturing extensive dynamics
350 of this histone mark. The H3K4me3 marks made at DSB hotspots by PRDM9
351 appear in leptonema, are maximal in zygonema, then reduce in early pachynema
352 before being removed by late pachynema. Intriguingly, the presence of robust
353 H3K4me3 signal at early pachynema implies that PRDM9 trimethylates more
354 nucleosomes than are used during DSB formation. An alternate explanation
355 would require DSB formation in early pachynema. This is unlikely because
356 homolog synapsis shuts down the DSB machinery in yeast[43] and an analogous
357 system appears to operate in mammals[31,44]. These are the first data to
358 demonstrate excess H3K4me3 by PRDM9 in *wild-type* mice, however recent
359 work in mice co-expressing both endogenous and transgenic PRDM9 reached a
360 similar conclusion[36].
361
362 H3K4me3 at gene promoters is also dynamic through MPI. By comparing
363 H3K4me3 at promoters with gene expression, we found that while most transcript
364 levels are predicted by H3K4me3, a large subset of transcripts exhibit gene
365 expression that appears decoupled from H3K4me3 levels. Most such transcripts
366 exhibit a strong H3K4me3 signal at the TSS in early MPI, but expression is
367 maximal late in MPI; this may result from long-lived mRNA, from mRNA
368 accumulation, or from H3K4me3 marks at the sites of poised RNA polymerase[35].
369 Indeed, these transcripts are intriguing candidates for further understanding gene

370 regulation in MPI. Finally, whereas DSB hotspots and gene promoters constitute
371 about half of the dynamic H3K4me3 marks in MPI, the functions of the other half
372 of dynamic histone marks remain completely unknown. At least some of these
373 sites are likely enhancers, and together, they represent a large, heretofore
374 unstudied set of functional sites that may be modulating mouse meiosis.
375
376 To examine the chromatin landscape at DSB hotspots, we also performed the
377 first systematic genome-wide survey on a wide repertoire of histone marks. Our
378 ability to efficiently sort early MPI nuclei eliminated the need to perform these
379 experiments in juvenile mice in which early MPI nuclei were only relatively
380 enriched[15,45]. Aside from H3K4me3[27] and H3K36me3[15,16], H3K9ac, H3K4me2,
381 H3K4me1, H3K27ac and H4ac5 were enriched at DSB hotspots genome-wide.
382 Incorporating the extra information about these histone marks only marginally
383 increases our ability to predict hotspot strength, implying that even the expanded
384 chromatin landscape at hotspots remains a poor predictor of DSB formation. The
385 genome is replete with H3K4me3 marks that are used for DSB formation in the
386 absence of PRDM9, however, enigmatically, these sites are not used if PRDM9
387 is functional. We demonstrated that both H3K36me3[15,40,42], and multiple histone
388 marks distinguish between these sites. Although this unique histone "code" may
389 help the DSB machinery to distinguish PRDM9-defined H3K4me3 from other
390 H3K4me3 marks, it is equally possible that direct interactions mediated by
391 PRDM9 itself are important for this distinction. It seems unlikely that PRDM9 is
392 responsible for all of the histone tail modifications at hotspots, therefore other
393 histone remodelers are likely acting at the sites of PRDM9-marked
394 H3K4/K36me3. H3K9ac and H3K4me3 are strength correlated, have a similar
395 temporal dynamic and are of a similar magnitude, implying that they occur at a
396 comparable frequency at hotspots. In mitotic cells, H3K9 acetylation is actively
397 promoted at H3K4 trimethylated nucleosomes[46]. It is therefore likely that H3K9ac
398 is a constitutive response to H3K4 trimethylation. Nonetheless, histone
399 acetylation may be functionally important *per se* in the context of DSB formation;

for example, in fission yeast, H3K9 to H3A9 mutation eliminates H3K9ac and consequently reduces DSB formation[40].

Overall, we demonstrated that our simple, robust method can explore previously hidden dynamics of mammalian meiotic recombination. Importantly, our method is widely applicable to other organisms, tissues and cell types, paving the way for further understanding of the temporal dynamics of other developmental processes.

Data availability

The sequencing data reported in this paper are archived at the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo) as accession no. GSE121760.

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., F.P. and R.D.C-O. conceived the study and designed the experiments. K-W.G.L. and G.C. 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.

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

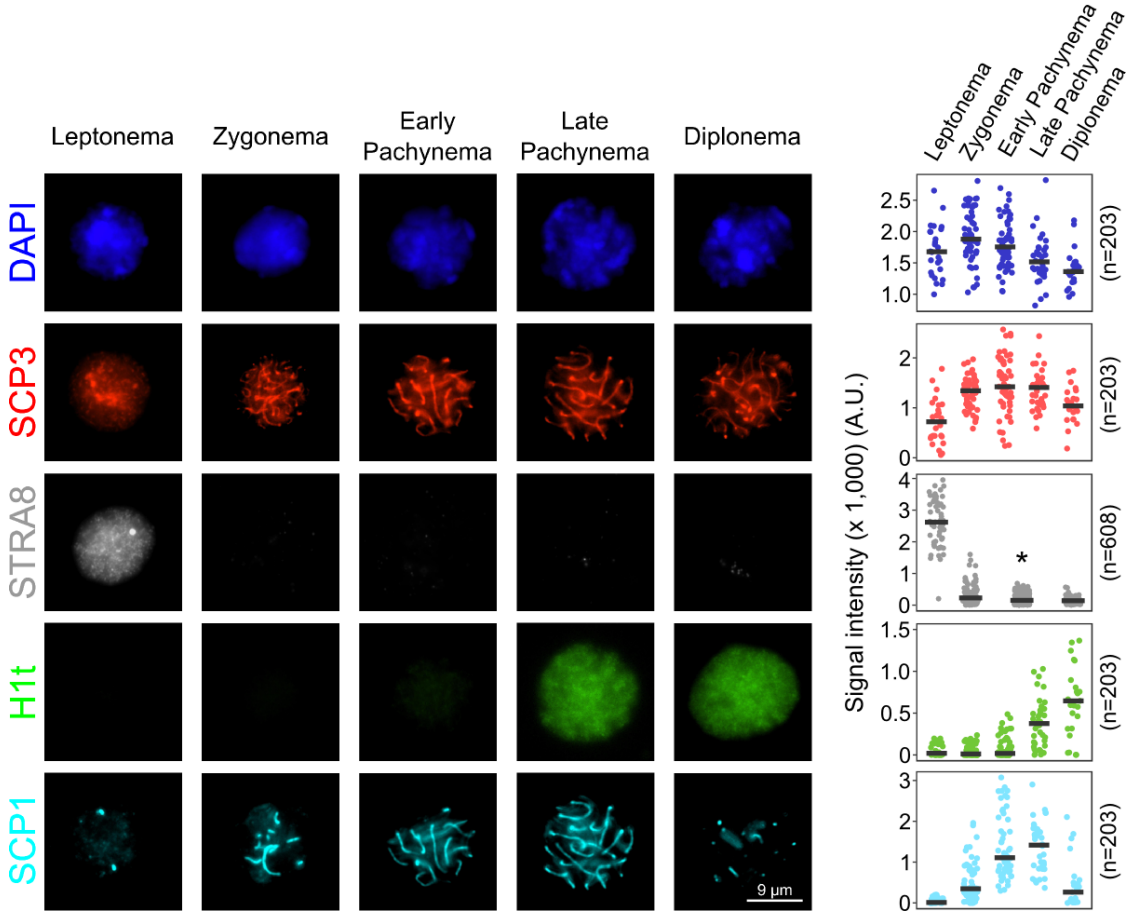
428

Sub-stage	Replicate	DNA input (ng)	Nuclei (#)	SPoT (HS) (%)	SPoT (All) (%)	Peaks (#)
Leptonema	1	81	7,000	2	32	29,004
	2	1,306	109,000	4	45	47,804
Zygonema	1	342	28,000	2	15	12,367
	2	3,174	264,000	4	28	35,014
Early Pachynema	1	339	28,000	2	13	7,517
	2	1,693	141,000	4	34	36,125
Late Pachynema	1	126	10,000	1	23	11,577
	2	1,067	89,000	2	49	31,630
Diplonema	1	60	5,000	1	28	16,526
	2	1,196	100,000	2	50	25,605

429

430

431 **Fig. 1**



432

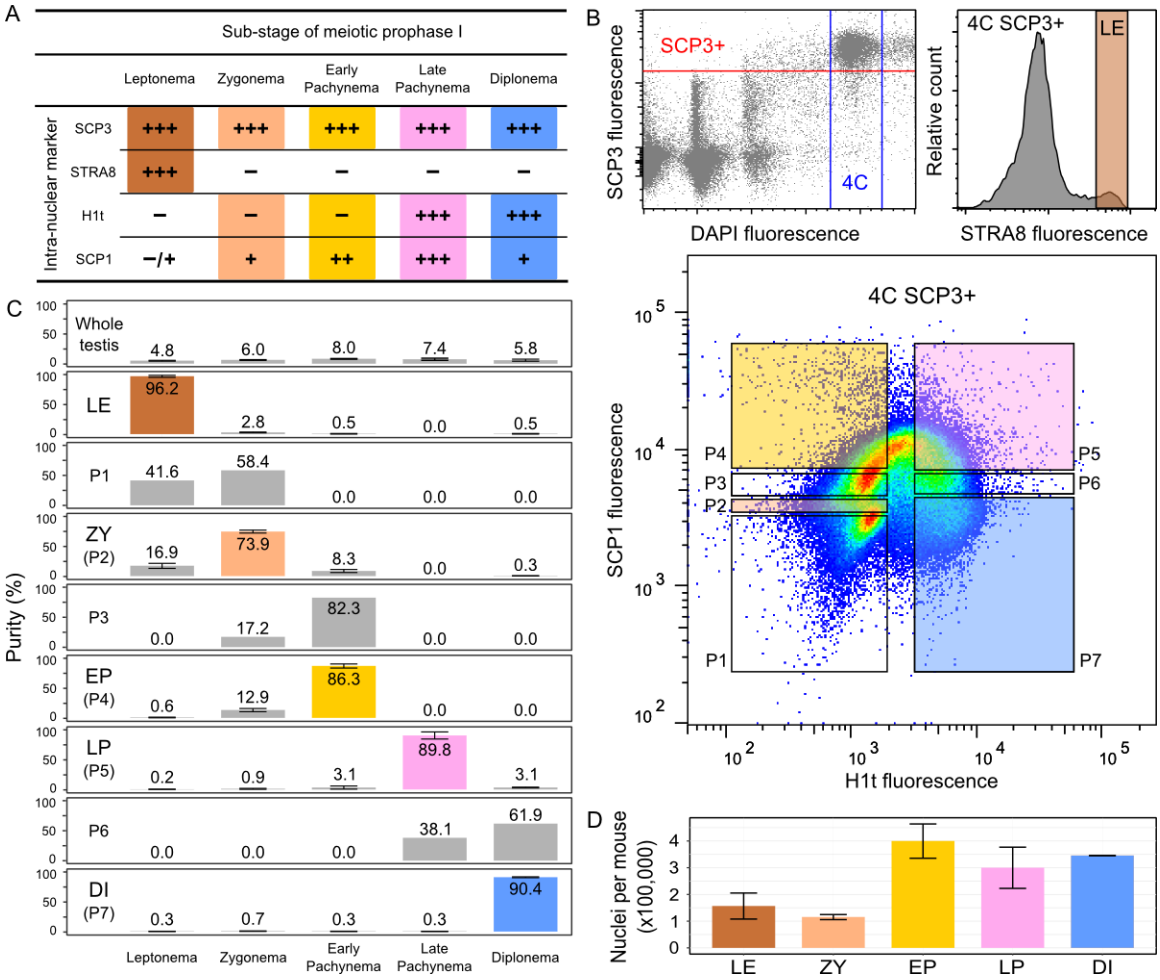
433

434 **Fig 1: Immunofluorescence staining of spermatocyte nuclei.**

435 Immunofluorescence images and signal quantification of stage-specific
436 spermatocyte nuclei through meiosis prophase I. Details for signal quantification
437 are described in Methods. Microscopic images are selected from two
438 independent experiments in which two different combinations of primary
439 antibodies are used; one using SCP3, H1t and SCP1, another one using SCP3
440 and STRA8. *Early and late pachytene nuclei cannot be unambiguously
441 differentiated in the absence of H1t staining, and are therefore merged for
442 counting and signal quantification.

443

444 **Fig. 2**



445

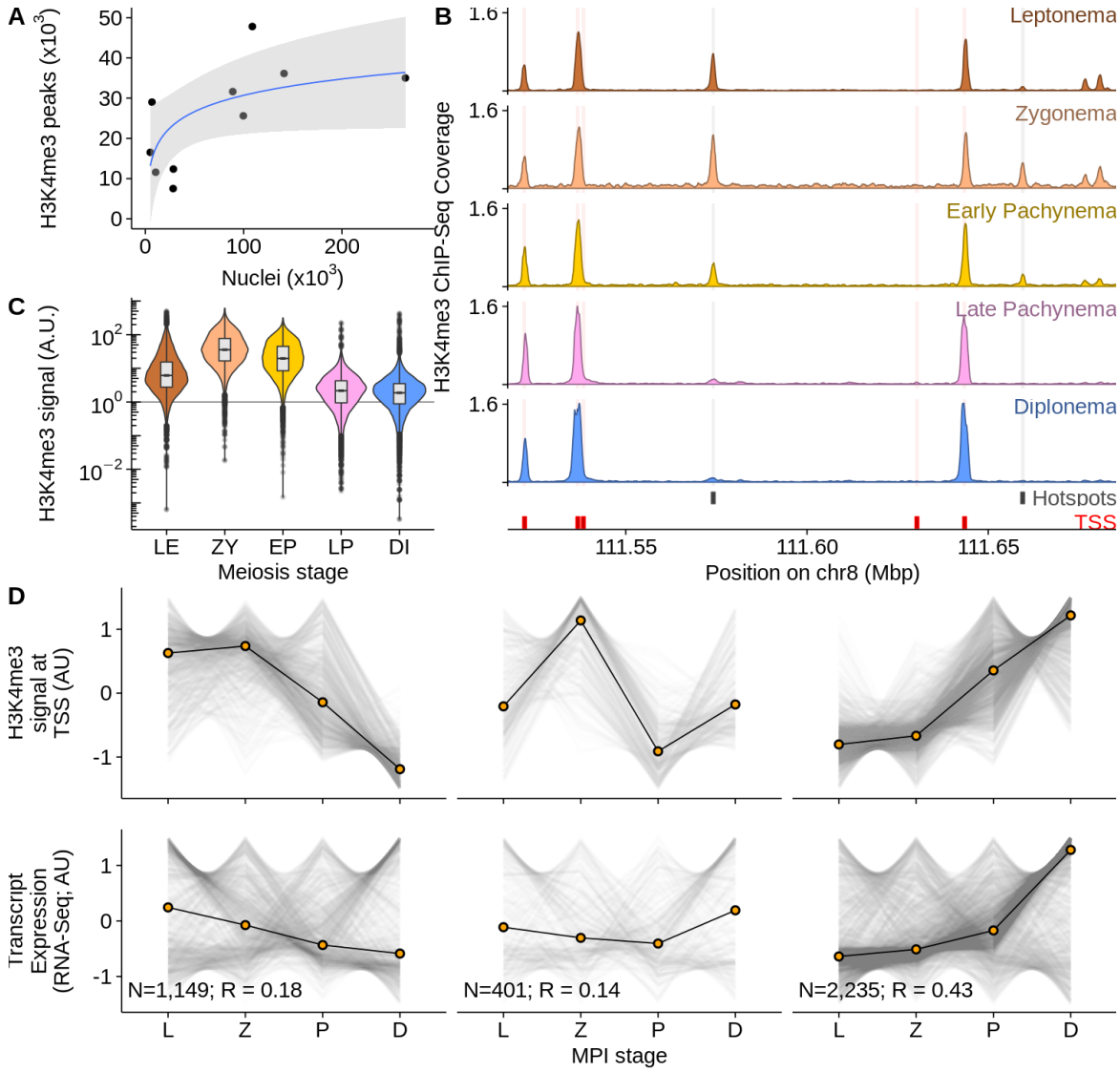
446

447 **Fig 2: Experimental design for isolating stage-specific spermatocyte nuclei.**

448 (A) Signal strength of intra-nuclear markers across meiotic sub-stages observed
449 in immunofluorescence staining. Signal strength is classified as absent (–), very
450 weak (–/+), weak (+), medium (++), or strong (+++). Combinatorial markers for
451 isolating stage-specific nuclei are highlighted. (B) Flow cytometric strategies for
452 isolating populations of stage-specific nuclei. Meiotic 4C nuclei are gated by
453 DAPI and SCP3 signals (top left). Stage-specific nuclei are sorted into
454 populations based on combinatorial signals of intra-nuclear markers in two
455 separate sorts; one using antibodies against STRA8 for leptonema (top right),
456 and the other one using antibodies against H1t and SCP1 for seven populations

457 (P1 to P7) (bottom). (C) Distributions and purities of each specific type of nuclei
458 in whole-testis and in sorted populations. The five selected populations of
459 leptonema, zygonema, early pachynema, late pachynema, and diplonema are
460 highlighted as LE, ZY, EP, LP, and DI, respectively. Purities (mean with standard
461 error) of these populations are derived from two or three independent sorts. (D)
462 Numbers of nuclei in each sub-population collected in an adult mouse. Data
463 (mean with standard error) are derived from two independent sorts.

464 **Fig. 3**



465

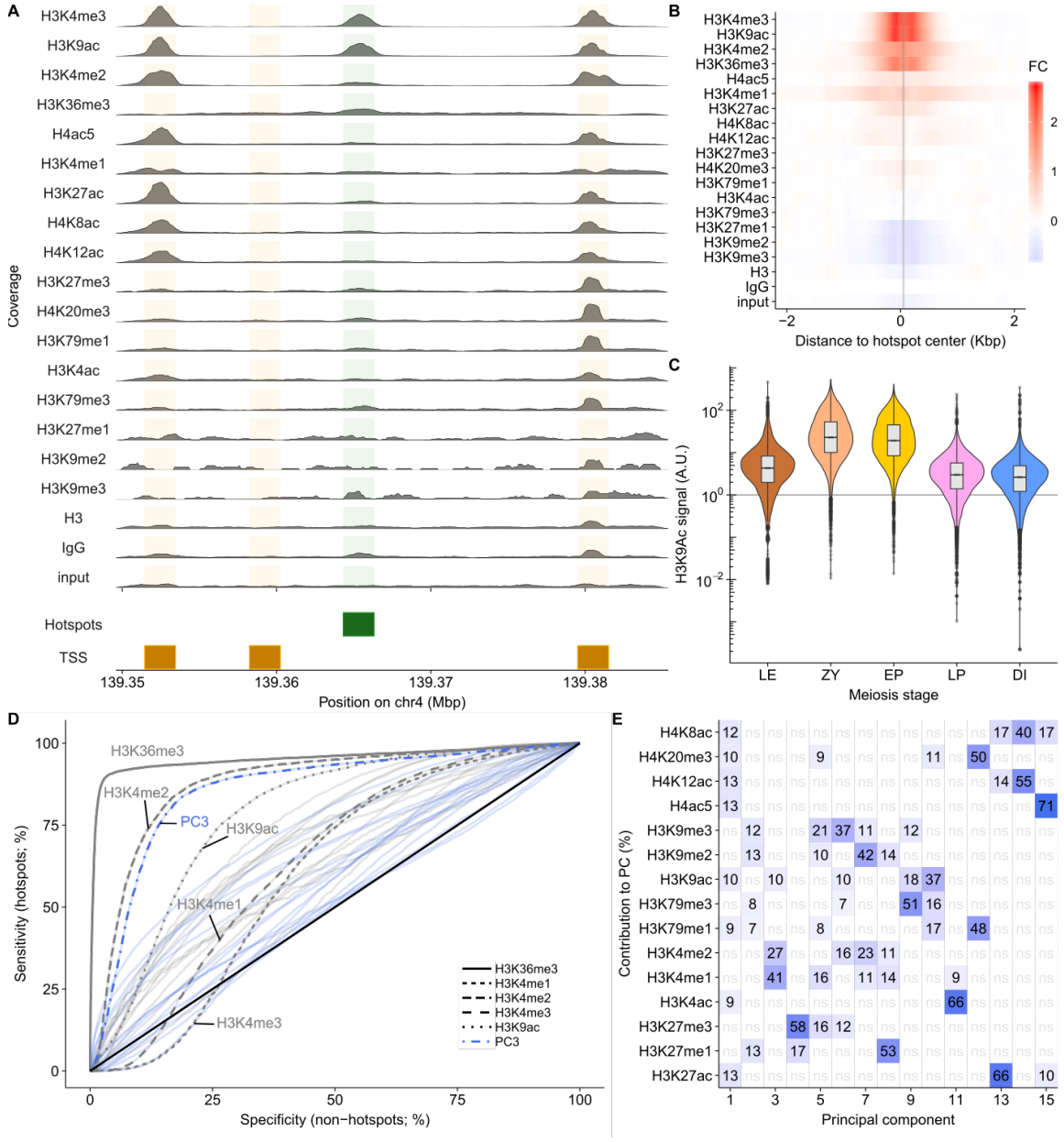
466

467 **Fig 3: H3K4me3 dynamics across meiotic sub-stages.**

468 (A) Numbers of starting nuclei and H3K4me3 peaks detected in stage-specific
469 H3K4me3-ChIP-Seq. (B) H3K4me3 dynamics at different genomic elements in
470 five sorted sub-populations of meiotic nuclei. (C) H3K4me3 dynamics at DSB
471 hotspots across five meiotic sub-stages. Normalized H3K4me3 signals at DSB
472 hotspots are relative to those at 99 selected TSSs where H3K4me3 signals are
473 constant through meiosis. (D) Stage-specific H3K4me3 at TSSs recapitulate
474 meiotic gene expression patterns. Genes with only one isoform are selected for

475 this analysis. H3K4me3 profiles at TSSs are grouped into three clusters (top),
476 while the respective mRNA levels are shown in the bottom panel. L =
477 Leptonema, Z = Zygonema, P = Early and Late Pachynema, D = Diplonema.

478 **Fig. 4**



479

480

481 **Fig 4: Histone modifications at DSB hotspots.**

482 (**A**) Distributions of histone marks along a genomic region. Locations of DSB
483 hotspots are shaded in green, TSS in orange. (**B**) Histone marks are enriched at
484 DSB hotspots. Red depicts enrichment relative to IgG, blue is depletion. (**C**)
485 H3K9ac dynamics at DSB hotspots across five meiotic sub-stages. **(D-E)**
486 Principal component analysis using histone marks can distinguish DSB hotspots

487 from PRDM9-independent H3K4me3 sites. **(D)** Histone modifications can
488 distinguish DSB hotspots from other H3K4me3 sites. An ROC curve for each
489 histone mark (grey) or PC (blue) is shown. Selected histone modifications and
490 PC3 are highlighted. **(E)** Contributions of histone marks to each PC. Only marks
491 that contribute more than expected are shown. Deeper blue indicates a stronger
492 contribution (numbers show percent contribution).
493
494
495

496 # Methods

497

498 **Mouse strains**

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

502

503 **Antibodies**

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

508

509 The following antibodies were used for secondary immunofluorescence staining:
510 goat anti-mouse IgG conjugated with Cy3 (Jackson ImmunoResearch, 115-167-
511 003), goat anti-rabbit IgG conjugated with FITC (Jackson ImmunoResearch, 115-
512 097-003), goat anti-mouse IgG conjugated with DyLight488 (Abcam, ab98757),
513 goat anti-rabbit IgG conjugated with Cy3 (Jackson ImmunoResearch, 111-165-
514 003), goat anti-guinea pig IgG conjugated with Alexa Fluor 488 (Thermo Fisher
515 Scientific, A-11073), streptavidin conjugated with Alexa Fluor 647 (Thermo Fisher
516 Scientific, S21374).

517

518 The following antibodies were used for ChIP: anti-H3K4me3 (Abcam, ab8580),
519 anti-H3K9ac (Active Motif, 39918), anti-H3K4me2 (Active Motif, 39914), anti-
520 H3K36me3 (Active Motif, 61102), anti-H4ac5 (Millipore, 06-946), anti-H3K4me1
521 (Abcam, ab8895), anti-H3K27ac (Abcam, ab177178), anti-H4K8ac (Abcam,
522 ab15823), anti-H4K12ac (Active Motif, 39928), anti-H3K27me3 (Millipore, 07-
523 449), anti-H4K20me3 (Millipore, 07-463), anti-H3K79me1 (Abcam, ab2886), anti-
524 H3K4ac (Abcam, ab176799), anti-H3K79me3 (Abcam, ab2621), anti-H3K9me2
525 (Abcam, ab1220), anti-H3K9me3 (Active Motif, 39766), anti-H3K27me1

526 (Millipore, 07-448), anti-H3 (Abcam, ab1791), normal rabbit IgG (Millipore, 12-
527 370).
528
529 **Nuclei preparation**
530 Testes from adult mice were de-capsulated and fixed with 1% formaldehyde for
531 10 min followed by 5 min of quenching at RT. Fixed tissues were homogenized
532 and filtered through a 70- or 100-μm cell strainer. After washing with chilled 1 x
533 PBS, nuclei were extracted using nucleus extraction buffer (15 mM Tris-HCl pH
534 7.4, 0.34 M sucrose, 15 mM NaCl, 60 mM KCl, 0.2 mM EDTA, 0.2 mM EGTA) on
535 ice for 5 min and were homogenized with 20 strokes with loose pestle followed
536 by 10 strokes with tight pestle. Nuclei were filtered through a 40-μm cell strainer
537 and resuspended in chilled PBTB buffer (1 x PBS with 0.1% Triton X-100, 5%
538 BSA and protease inhibitor).
539
540 **Immunofluorescence staining**
541 Nuclei were incubated in 10% of normal serum at RT for 10 min. This blocking
542 step helps reduce non-specific binding of antibodies. Nuclei were then labelled
543 with different combinations of primary antibodies (1 μg of antibodies to 11 million
544 events) in 10% of normal serum at 20°C for 40 min. Nuclei were washed,
545 resuspended in PBTB, blocked with serum, and then labelled with secondary
546 antibodies (1:250 dilution) in 10% of normal serum at 20°C for 30 min. Controls
547 for each secondary antibody were also prepared for setting up the threshold of
548 background fluorescent signal for FACS. Nuclei were washed, resuspended in
549 PBTB, and stored at 4°C until sorting.
550
551 **Quantification of immunofluorescence signal intensity**
552 Nuclei isolated from whole-testis were stained with antibodies against SCP3 and
553 STRA8 or with antibodies against SCP3, H1t and SCP1, and subsequently with
554 secondary antibodies as described. Immunofluorescence-labelled nuclei were
555 spread on microslides, stained with mounting medium with DAPI (Vetashield),
556 sealed with cover slips, and viewed using fluorescence microscopy.

557 Immunofluorescence signals of each marker were quantified using Volocity 6.2.1
558 (PerkinElmer). 4C nuclei were identified using signals from the DAPI channel by
559 excluding objects with size < 40 μm^2. Touching objects were separated using the
560 parameter of 30–50 μm^2. Stage-specific nuclei in MPI were characterized by the
561 presence of SCP3 signal and by the patterning of synaptonemal complex[12].
562 Signals from all channels were recorded and exported for signal quantification.
563
564 **Isolation of meiotic sub-populations with FACS**
565 Nuclei were filtered through a 40-μm cell strainer, and stained with DAPI for
566 30 min or longer at RT before sorting. All sorting experiments were performed on
567 either a BD FACSAria II or a BD FACSAria Fusion flow cytometer at a flow rate
568 of ~20,000 events/sec. Singlets were gated using both forward scatter and side
569 scatter. Nuclei from primary spermatocytes (4C) were gated based on DNA
570 content deduced from the DAPI signal, and sorted into populations of interest
571 based on fluorochrome intensity into collection tubes containing PBTB. Sorted
572 nuclei were collected by centrifugation. Nuclei were examined under microscope
573 and counted with a hemocytometer. Finally, excess buffer was removed and
574 nuclei pellets were stored at –80°C.
575
576 **Assessment of purity of sorted populations**
577 An aliquot of nuclei suspension from each sorted population was spread on
578 microslides, further stained with mounting medium with DAPI (Vectashield), and
579 sealed with coverslips for purity examination. An average of 208 nuclei from each
580 sorted population were checked at 400X magnification using
581 immunofluorescence microscopy. The staging was primarily accessed by the
582 patterning of SCP3[12] and further confirmed by signals of STRA8 or of H1t and
583 SCP1 accordingly.
584
585 **Chromatin extraction**
586 Chromatin extracted for ChIP-Seq was either sheared by sonication or
587 Micrococcal nuclease (MNase) digestion. Chromatin for all H3K4me3-ChIP-Seq

588 for studying histone dynamics in stage-specific populations was sheared by
589 sonication, whereas ChIP-Seq for identifying histone marks at hotspots were
590 performed on MNase-digested chromatin. Experimental procedures are listed
591 below in detail.
592
593 **Shearing chromatin by sonication:**
594 Frozen nuclei pellets were thawed at RT for 10 min. Nuclei were lysed with Lysis
595 buffer 1 (0.25% Triton X-100, 10 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl pH
596 8) and incubated at RT for 10 min. Nuclei pellets were subsequently washed with
597 Lysis buffer 2 (200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris-HCl pH 8)
598 and lysed with RIPA buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 0.5 mM EGTA,
599 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 140 mM NaCl plus
600 protease inhibitor). Chromatin was sheared into ~500–1000 bp fragments by
601 sonication using Bioruptor (diagenode). Chromatin concentration was measured
602 using a Qubit dsDNA HS Assay Kit (Thermo Fisher).
603
604 **MNase digestion:**
605 Frozen nuclei pellets were thawed at RT for 10 min. Nuclei pellets were
606 resuspended with MNase buffer (50 mM Tris-HCl pH 8, 1 mM $CaCl_2$, 4 mM
607 $MgCl_2$, 4% NP-40 plus protease inhibitor). MNase digestion was performed in a
608 concentration of 3U MNase (USB, Affymetrix) per one million nuclei at 37°C for 5
609 min. The reaction was stopped by adding a final concentration of 10 mM EDTA
610 and incubated at 4°C for 5 min. Chromatin was pelleted and resuspended in
611 RIPA buffer. Chromatin concentration was measured.
612
613 **Chromatin Immunoprecipitation**
614 Chromatin was immunoprecipitated with 0.8–5 μg antibodies (or 5 μl unpurified
615 serum) in 0.5–1 ml RIPA buffer at 4°C overnight (see Supplementary Table 1 for
616 details). The immuno-complexes were captured using 20–75 μl Dynabeads
617 Protein G (30 mg/ml, Novex) at 4°C for 2 hr. The beads were washed with

618 buffers previously described[26]. ChIPed DNA was eluted using a IPure kit v2
619 (diagenode).
620
621 **Sequencing library construction**
622 ChIP-seq libraries were constructed with a KAPA Hyper Prep Kit (Kapa
623 Biosystems) following steps for generating 1 μg of library DNA. DNA libraries
624 were cleaned up with an Agencourt AMPure XP system (Beckman Coulter). The
625 DNA concentration and fragment size of these libraries were measured with a
626 Qubit dsDNA HS Assay Kit (Thermo Fisher) and an Agilent High Sensitivity DNA
627 Kit (Agilent), respectively.
628
629 **High throughput DNA sequencing**
630 DNA sequencing was performed on the Illumina HiSeq 2500 or HiSeq X.
631 Sequencing tags were aligned to the mouse mm10/NCBI38 reference genome
632 using BWA mem 0.7.12[47].
633
634 **Strength estimation**
635 NCIS[48] was used to estimate the contribution of background to each sequencing
636 library. The H3K4me3/H3K9ac strength was calculated for each interval as the
637 number of sequencing reads are less than the NCIS-corrected number of input
638 reads.
639
640 **Inter-sample normalization using "stable" promoters**
641 The H3K4me3/H3K9ac signal at the central ± 1 Kbp region around GENCODE
642 transcript 5' ends (TSS) was quantified (as described above) from ChIP-Seq
643 reads in each of the five meiotic populations. TSS that overlapped a DSB
644 hotspot, TSSs with a negative strength after correction and TSS with the
645 strongest (1%) and weakest (15%) H3K4me3 signal at each stage were
646 discarded. The log2 ratio of H3K4me3 was calculated at each TSS between all
647 pairs of stages. TSSs with an absolute log2 (fold-change) ≤ 1.2 between all

648 stages were retained. This resulted in 99 TSSs for H3K4me3 and 93 TSSs for
649 H3K9ac. The strength of each histone mark was then normalized by dividing by
650 the median signal at the respective set of stable TSSs.
651
652 **Correlating H3K4me3 profiles at TSSs with mRNA expression levels**
653 We used RNA-Seq data from spermatocyte populations obtained by classical
654 cell-sorting to quantify gene expression[33]. Gene expression at GENCODE vM18
655 transcripts was quantified for each RNA-Seq dataset using kallisto 0.42.1 (kallisto
656 quant --single -l 36 -b 100). Only transcripts with maximal expression > 1 tags per
657 million (TPM), with at least a 2-fold change in expression and with non-zero
658 expression and H3K4me3 signals at all stages were considered for subsequent
659 analysis. To minimize the confounding effects of multiple isoforms, only TSS with
660 a single transcript in this list were retained.
661
662 The temporal expression profile for each transcript was described as a four unit
663 vector (Leptonema, Zygonema, Pachynema, Diplonema). Similarly, a temporal
664 profile of H3K4me3 at the TSS was generated. We averaged the Early
665 Pachnema and Late Pachynema H3K4me3 signals to allow direct comparison
666 with the RNA-Seq data.
667
668 The overall correlation between H3K4me3 and gene expression was calculated
669 by concatenating all gene expression vectors and comparing with a
670 concatenated vector of H3K4me3. To calculate the expected random correlation,
671 we shuffled the order of each 4-unit transcript H3K4me3 vector before
672 concatenation. 10,000 iterations of this process were performed.
673
674 k-means clustering was used to group similar H3K4me3 temporal profiles and to
675 compare each set to gene expression. We used an implementation of the gap-
676 statistic to determine the optimal number of clusters (R; factoextra package). 3
677 clusters were determined as optimal. Within each set of clusters the method

678 described in the previous paragraph was employed to determine the correlation
679 between H3K4me3 and expression temporal profiles and to derive an empirical
680 p-value for this correlation. Clustering by gene expression instead of by
681 H3K4me3 signal gives similar correlations, though the number of clusters can
682 change (data not shown).
683
684 **Peak calling**
685 Peaks for H3K4me3 ChIP-Seq were called using MACS2 (version
686 2.1.0.20140616)[49] with default parameters and with a stage-matched input DNA
687 sample as a control. Peaks overlapping DSB hotspots or gene promoters were
688 ascertained using bedtools (version v2.22.1-4-g3ca83fb)[50] after removing
689 blacklisted regions described in[51]. DSB hotspots were defined and reported in
690 previous studies[27]. Transcription start sites (TSSs) were defined as the ± 0.5 Kbp
691 region around GENCODE v11 transcripts[52]. Overlapping TSSs were merged.
692
693 **Unbiased clustering of H3K4me3 peaks**
694 H3K4me3 ChIP-Seq peaks from all five populations were merged. H3K4me3
695 strength was calculated (as described above) for each merged peak in each
696 population. The H3K4me3 profile for each peak was described as a five unit
697 vector (LE, ZY, EP, LP, DI) and scaled by subtracting the mean and dividing by
698 the standard deviation. k-means clustering was used to cluster. Five clusters was
699 determined as optimal using the gap-statistic.
700
701 **Multiple linear regression**
702 Hotspots overlapping a TSS or overlapping a site used for DSB formation in
703 *Prdm9*$^{-/-}$ mice were discarded. Only hotspots with read coverage > 0 for all
704 histone marks were used for regression analyses because we performed
705 regression on the log10 transformed coverage values. The leaps package in R
706 was used to perform an all-subsets regression using the seven histone marks
707 enriched at DSB hotspots.
708

709 **Principal Component Analysis**

710 All H3K4me3 peaks from LE, ZY, EP and LP were used for analysis. Each
711 interval was resized to ± 250 bp around the center. Sequencing reads for each
712 histone mark were counted at each feature. Input DNA reads were also counted
713 and subtracted from the count for each mark, following NCIS[48] correction. DSB
714 hotspots and GENCODE TSSs were expanded to ± 1,500 bp to determine
715 H3K4me3 peaks that overlapped hotspots and TSSs, respectively. Peaks that
716 overlapped both a hotspot and TSS were discarded as the potential compound
717 signal would confound these analyses. Only autosomal peaks were used.
718 H3K36me3 and H3K4me3 were excluded. The R prcomp command was used for
719 Principal Component Analysis. Variables were scaled to have unit variance and
720 zero centered.

721

722 TSSs were split to into Active (TSS(On)) and Inactive (TSS(Off)); active TSS
723 were those upstream of genes expressed in early MPI (expression > 0 TPM at
724 pre-leptotene, leptotene, zygotene or pachytene; gene expression data from
725 RNA-Seq[33]). DSB hotspots were split into two sets by strength; the top 25%; Hot
726 HS and the rest; HS.

727

728 ROC curves were built by ranking intervals by the dependent variable. Intervals
729 were ranked either from high to low or from low to high and the ROC with the
730 higher area under the curve was used.

731

732

733

734 ## References

735 1. Handel, M. A., Eppig, J. J. & Schimenti, J. C. Applying 'Gold Standards' to In-Vitro-
736 Derived Germ Cells. *Cell* **157,** 1257–1261 (2014).
737 2. Bellve, A. R., Cavicchia, J. C. & Millette, C. F. Spermatogenic cells of the prepuberal
738 mouse: isolation and morphological characterization. *J. Cell Biol.* (1977).
739 3. Goetz, P., Chandley, A. C. & Speed, R. M. Morphological and temporal sequence of
740 meiotic prophase development at puberty in the male mouse. *J. Cell Sci.* **65,** 249–
741 263 (1984).
742 4. Kluin, P. M., Kramer, M. F. & de Rooij, D. G. Spermatogenesis in the immature
743 mouse proceeds faster than in the adult. *Int. J. Androl.* **5,** 282–294 (1982).
744 5. Zelazowski, M. J. *et al.* Age-Dependent Alterations in Meiotic Recombination Cause
745 Chromosome Segregation Errors in Spermatocytes. *Cell* **171,** 601–614.e13 (2017).
746 6. Hogarth, C. A. *et al.* Turning a spermatogenic wave into a tsunami: synchronizing
747 murine spermatogenesis using WIN 18,446. *Biol. Reprod.* **88,** 40–41 (2013).
748 7. Bastos, H. *et al.* Flow cytometric characterization of viable meiotic and postmeiotic
749 cells by Hoechst 33342 in mouse spermatogenesis. *Cytometry A* **65,** 40–49 (2005).
750 8. Barchi, M., Geremia, R., Magliozzi, R. & Bianchi, E. Isolation and analyses of
751 enriched populations of male mouse germ cells by sedimentation velocity: the
752 centrifugal elutriation. *Methods Mol. Biol.* **558,** 299–321 (2009).
753 9. Gaysinskaya, V., Soh, I. Y., Heijden, G. W. & Bortvin, A. Optimized flow cytometry
754 isolation of murine spermatocytes. *Cytometry A* **85,** 556–565 (2014).
755 10. Bonn, S. *et al.* Cell type–specific chromatin immunoprecipitation from multicellular
756 complex samples using BiTS-ChIP. *Nat. Protoc.* **7,** 978 (2012).
757 11. Meuwissen, R. L. *et al.* A coiled-coil related protein specific for synapsed regions of
758 meiotic prophase chromosomes. *EMBO J.* **11,** 5091–5100 (1992).

759 12. Dobson, M. J., Pearlman, R. E., Karaiskakis, A., Spyropoulos, B. & Moens, P. B.
760 Synaptonemal complex proteins: occurrence, epitope mapping and chromosome
761 disjunction. *J. Cell Sci.* **107 (Pt 10),** 2749–2760 (1994).
762 13. Inselman, A., Eaker, S. & Handel, M. A. Temporal expression of cell cycle-related
763 proteins during spermatogenesis: establishing a timeline for onset of the meiotic
764 divisions. *Cytogenet. Genome Res.* **103,** 277–284 (2003).
765 14. Mark, M. *et al.* STRA8-deficient spermatocytes initiate, but fail to complete, meiosis
766 and undergo premature chromosome condensation. *J. Cell Sci.* **121,** 3233–3242
767 (2008).
768 15. Powers, N. R. *et al.* The Meiotic Recombination Activator PRDM9 Trimethylates
769 Both H3K36 and H3K4 at Recombination Hotspots In Vivo. *PLoS Genet.* **12,**
770 e1006146 (2016).
771 16. Grey, C. *et al.* In vivo binding of PRDM9 reveals interactions with noncanonical
772 genomic sites. *Genome Res.* **27,** 580–590 (2017).
773 17. Margolin, G., Khil, P. P., Kim, J., Bellani, M. A. & Camerini-Otero, R. D. Integrated
774 transcriptome analysis of mouse spermatogenesis. *BMC Genomics* **15,** 39 (2014).
775 18. Sun, F. *et al.* Nuclear localization of PRDM9 and its role in meiotic chromatin
776 modifications and homologous synapsis. *Chromosoma* **124,** 397–415 (2015).
777 19. Moens, P. B. *et al.* The time course and chromosomal localization of recombination-
778 related proteins at meiosis in the mouse are compatible with models that can
779 resolve the early DNA-DNA interactions without reciprocal recombination. *J. Cell*
780 *Sci.* **115,** 1611–1622 (2002).
781 20. Cole, F. *et al.* Homeostatic control of recombination is implemented progressively in
782 mouse meiosis. *Nat. Cell Biol.* **14,** 424–430 (2012).
783 21. Buard, J., Barthès, P., Grey, C. & de Massy, B. Distinct histone modifications define

784 initiation and repair of meiotic recombination in the mouse. *EMBO J.* **28,** 2616–2624
785 (2009).
786 22. Moens, P. B. Histones H1 and H4 of surface-spread meiotic chromosomes.
787 *Chromosoma* **104,** 169–174 (1995).
788 23. Adli, M. & Bernstein, B. E. Whole-genome chromatin profiling from limited numbers
789 of cells using nano-ChIP-seq. *Nat. Protoc.* **6,** 1656–1668 (2011).
790 24. Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone
791 modifications in human and mouse. *Cell* **120,** 169–181 (2005).
792 25. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the
793 human genome. *Nature* **489,** 57–74 (2012).
794 26. Smagulova, F. *et al.* Genome-wide analysis reveals novel molecular features of
795 mouse recombination hotspots. *Nature* **472,** 375–378 (2011).
796 27. Brick, K., Smagulova, F., Khil, P., Camerini-Otero, R. D. & Petukhova, G. V. Genetic
797 recombination is directed away from functional genomic elements in mice. *Nature*
798 **485,** 642–645 (2012).
799 28. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor
800 binding patterns. *Nat. Genet.* **43,** 264–268 (2011).
801 29. Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and
802 modENCODE consortia. *Genome Res.* **22,** 1813–1831 (2012).
803 30. Mimitou, E. P., Yamada, S. & Keeney, S. A global view of meiotic double-strand
804 break end resection. *Science* **355,** 40–45 (2017).
805 31. Wojtasz, L. *et al.* Mouse HORMAD1 and HORMAD2, two conserved meiotic
806 chromosomal proteins, are depleted from synapsed chromosome axes with the help
807 of TRIP13 AAA-ATPase. *PLoS Genet.* **5,** e1000702 (2009).
808 32. Keeney, S., Lange, J. & Mohibullah, N. Self-Organization of Meiotic Recombination

809 Initiation: General Principles and Molecular Pathways. *Annu. Rev. Genet.* **48,** 187–
810 214 (2014).
811 33. Gaysinskaya, V. *et al.* Transient reduction of DNA methylation at the onset of
812 meiosis in male mice. *Epigenetics Chromatin* **11,** 15 (2018).
813 34. Tibshirani, R., Walther, G. & Hastie, T. Estimating the number of clusters in a data
814 set via the gap statistic. *J. R. Stat. Soc. Series B Stat. Methodol.* **63,** 411–423
815 (2001).
816 35. Vastenhouw, N. L. *et al.* Chromatin signature of embryonic pluripotency is
817 established during genome activation. *Nature* **464,** 922–926 (2010).
818 36. Diagouraga, B. *et al.* PRDM9 Methyltransferase Activity Is Essential for Meiotic DNA
819 Double-Strand Break Formation at Its Binding Sites. *Mol. Cell* **69,** 853–865.e6
820 (2018).
821 37. Walker, M. *et al.* Affinity-seq detects genome-wide PRDM9 binding sites and
822 reveals the impact of prior chromatin modifications on mammalian recombination
823 hotspot usage. *Epigenetics Chromatin* **8,** 31 (2015).
824 38. Getun, I. V. *et al.* Functional Roles of Acetylated Histone Marks at Mouse Meiotic
825 Recombination Hot Spots. *Mol. Cell. Biol.* **37,** (2017).
826 39. Merker, J. D. *et al.* The histone methylase Set2p and the histone deacetylase
827 Rpd3p repress meiotic recombination at the HIS4 meiotic recombination hotspot in
828 Saccharomyces cerevisiae. *DNA Repair* **7,** 1298–1308 (2008).
829 40. Yamada, S., Ohta, K. & Yamada, T. Acetylated Histone H3K9 is associated with
830 meiotic recombination hotspots, and plays a role in recombination redundantly with
831 other factors including the H3K4 methylase Set1 in fission yeast. *Nucleic Acids Res.*
832 **41,** 3504–3517 (2013).
833 41. Lange, J. *et al.* The Landscape of Mouse Meiotic Double-Strand Break Formation,

834 Processing, and Repair. *Cell* **167,** 695–708.e16 (2016).
835 42. Yamada, S. *et al.* Genomic and chromatin features shaping meiotic double-strand
836 break formation and repair in mice. *Cell Cycle* **16,** 1870–1884 (2017).
837 43. Thacker, D., Mohibullah, N., Zhu, X. & Keeney, S. Homologue engagement controls
838 meiotic DNA break number and distribution. *Nature* **510,** 241–246 (2014).
839 44. Kauppi, L. *et al.* Numerical constraints and feedback control of double-strand breaks
840 in mouse meiosis. *Genes Dev.* **27,** 873–886 (2013).
841 45. Baker, C. L., Walker, M., Kajita, S., Petkov, P. M. & Paigen, K. PRDM9 binding
842 organizes hotspot nucleosomes and limits Holliday junction migration. *Genome Res.*
843 **24,** 724–732 (2014).
844 46. Gates, L. A., Foulds, C. E. & O'Malley, B. W. Histone Marks in the 'Driver's Seat':
845 Functional Roles in Steering the Transcription Cycle. *Trends Biochem. Sci.* **42,** 977–
846 989 (2017).
847 47. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-
848 MEM. *arXiv [q-bio.GN]* (2013).
849 48. Liang, K. & Keleş, S. Normalization of ChIP-seq data with control. *BMC*
850 *Bioinformatics* **13,** 199 (2012).
851 49. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137
852 (2008).
853 50. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing
854 genomic features. *Bioinformatics* **26,** 841–842 (2010).
855 51. Smagulova, F., Brick, K., Pu, Y., Camerini-Otero, R. D. & Petukhova, G. V. The
856 evolutionary turnover of recombination hot spots contributes to speciation in mice.
857 *Genes Dev.* **30,** 266–280 (2016).
858 52. Mudge, J. M. & Harrow, J. Creating reference gene annotation for the mouse

859 C57BL6/J genome assembly. *Mamm. Genome* **26,** 366–378 (2015).

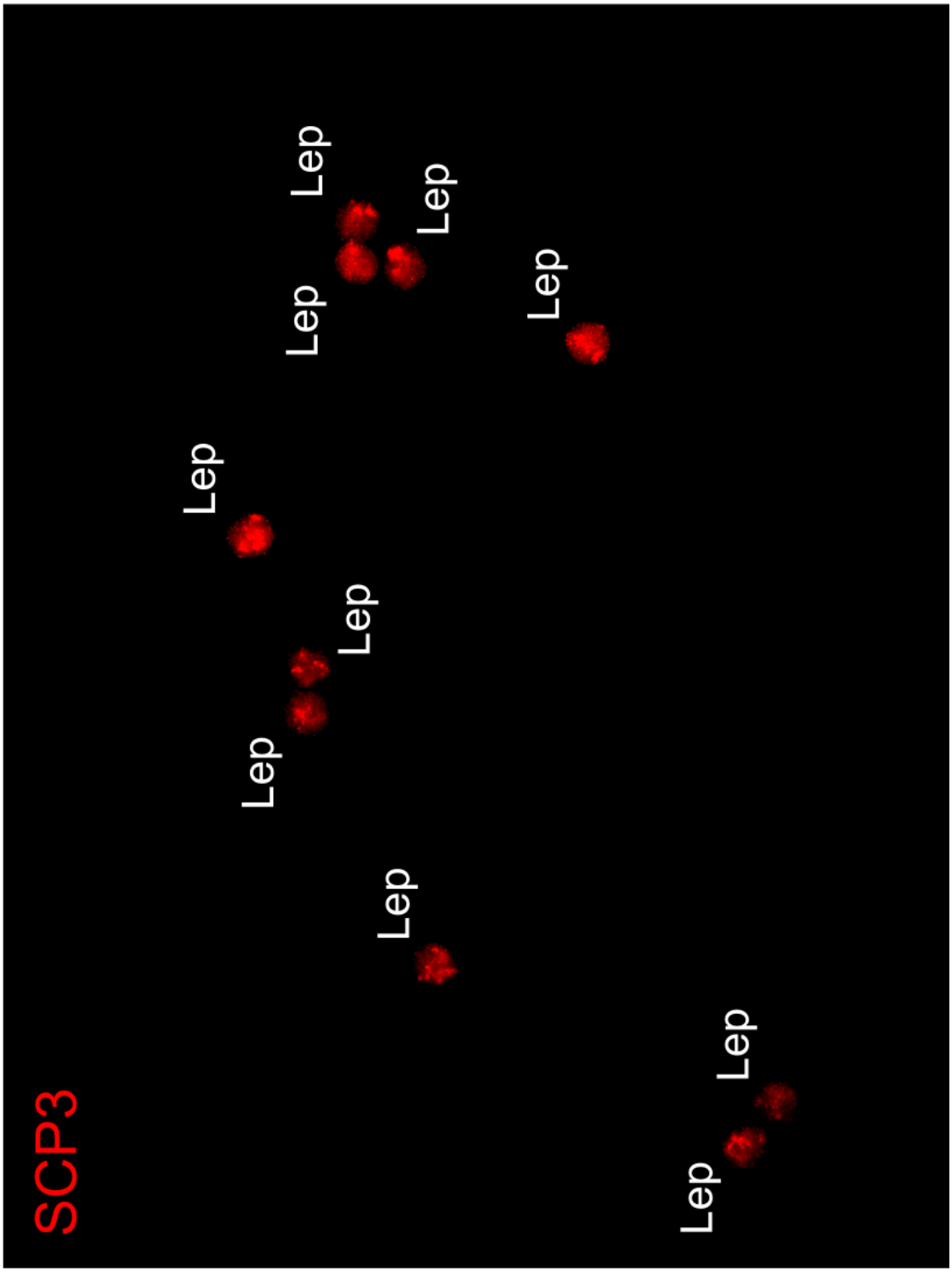
860 53. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible

861 platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42,** W187–91

862 (2014).

863

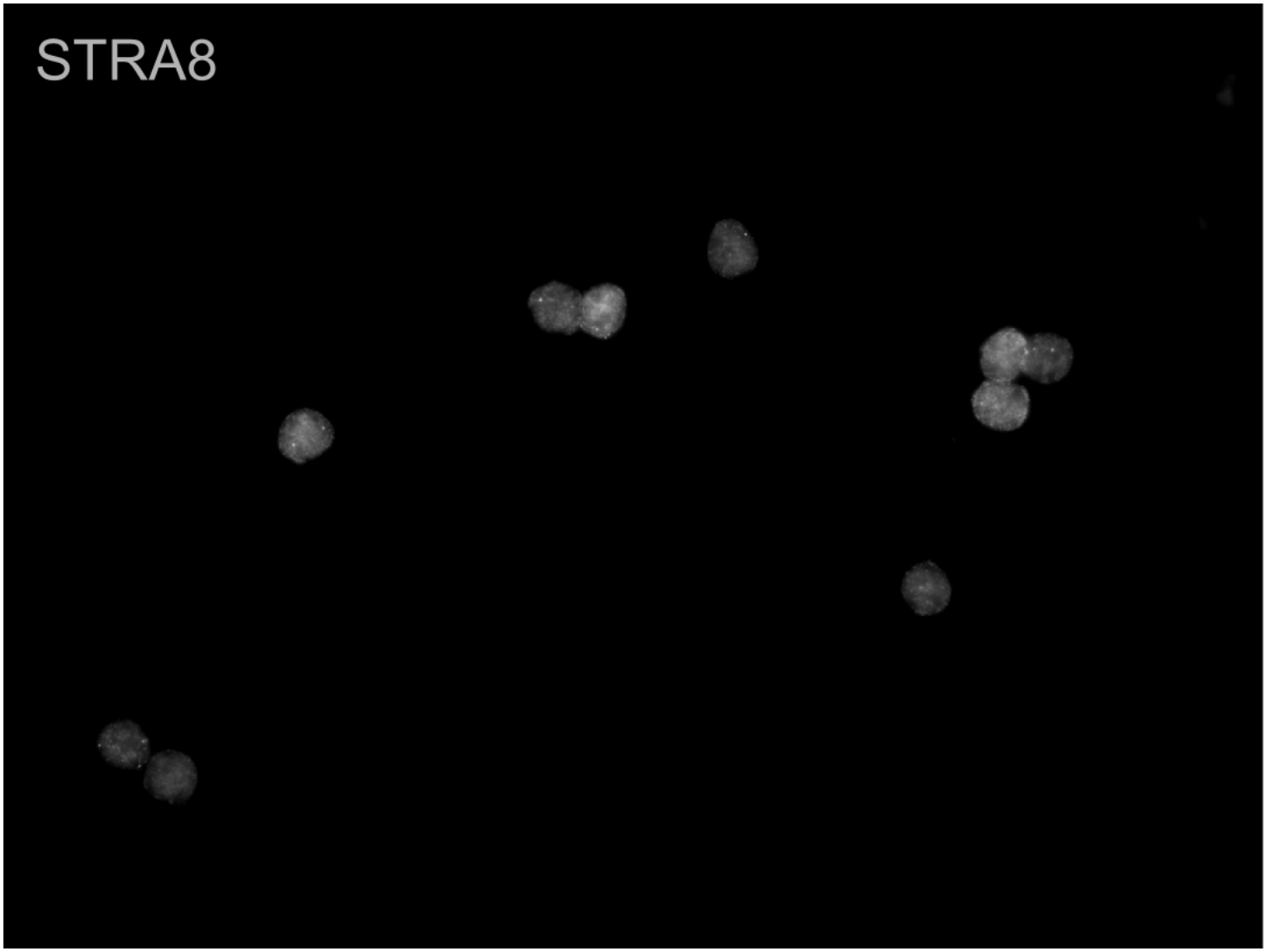
Supplementary Figures

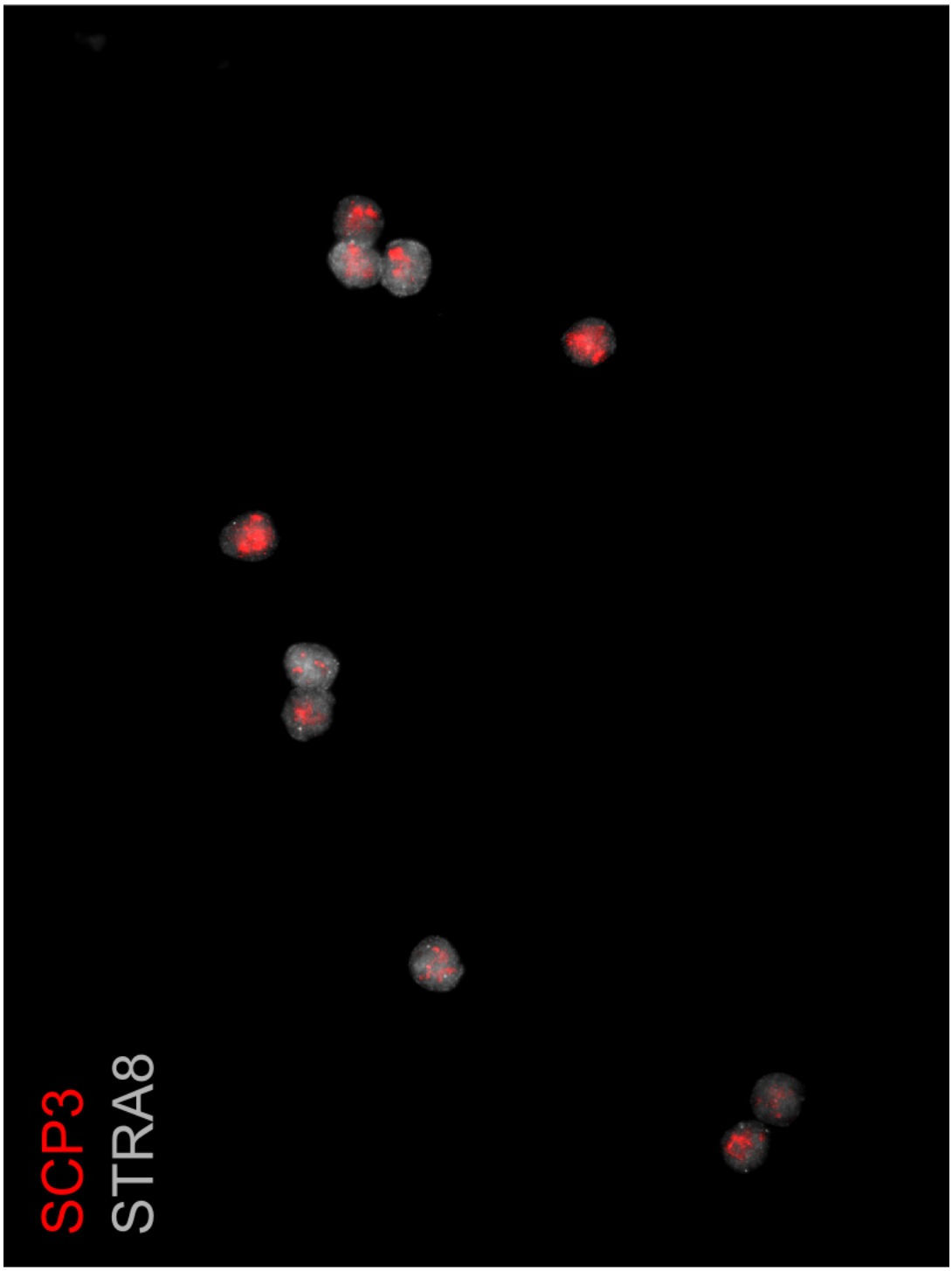


A (i)

866

A (ii)

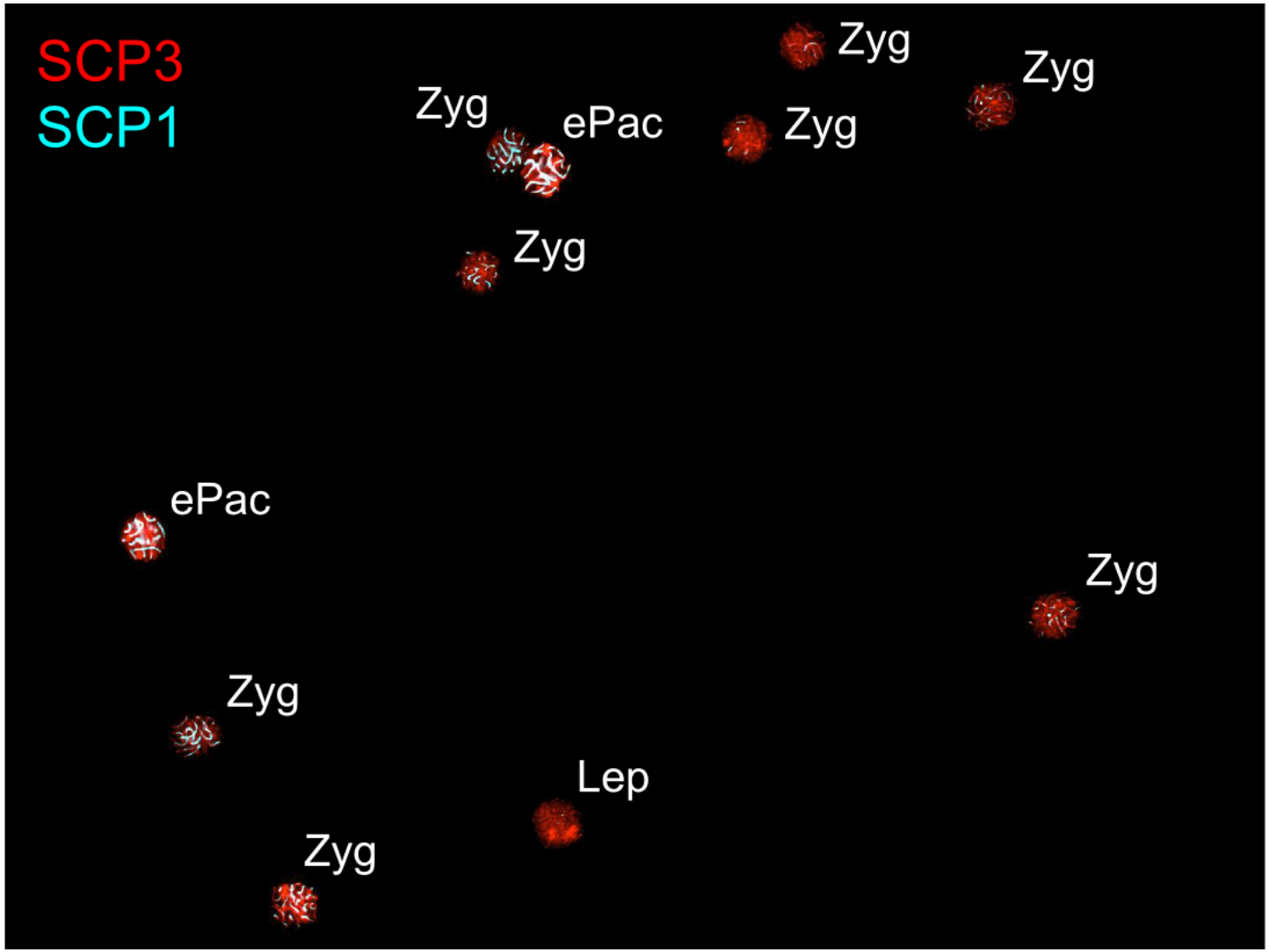


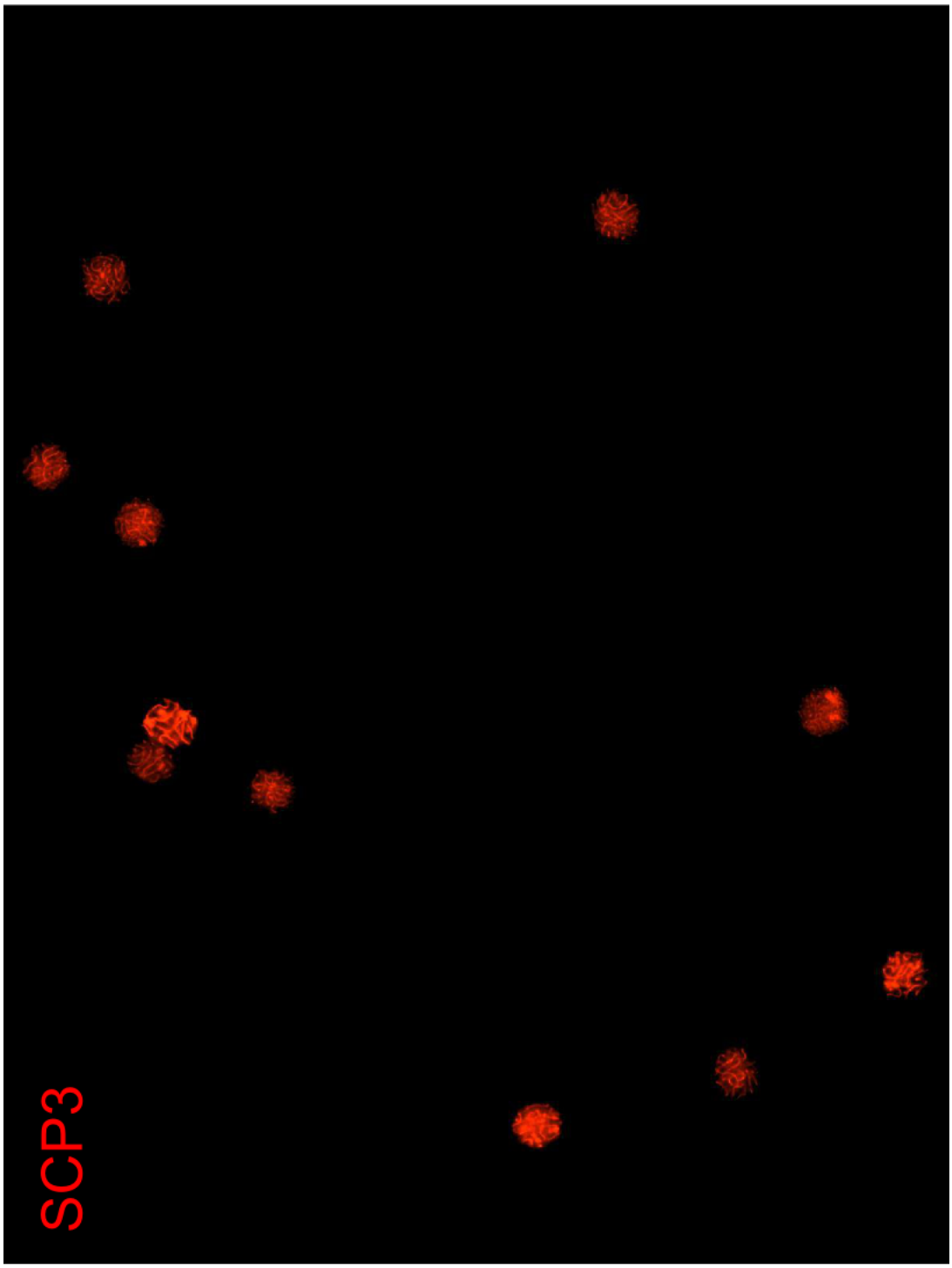


A (iii)

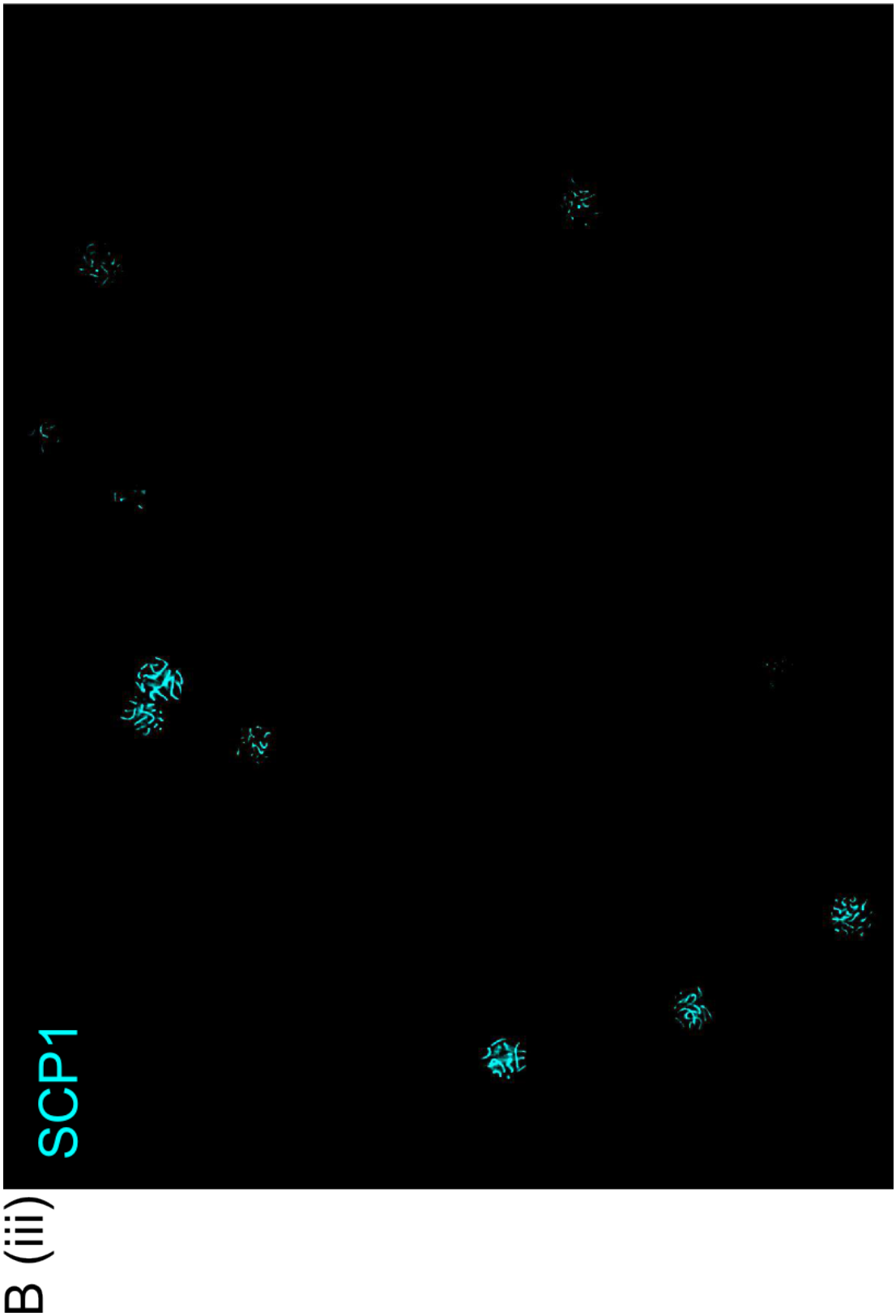
868

B (i)



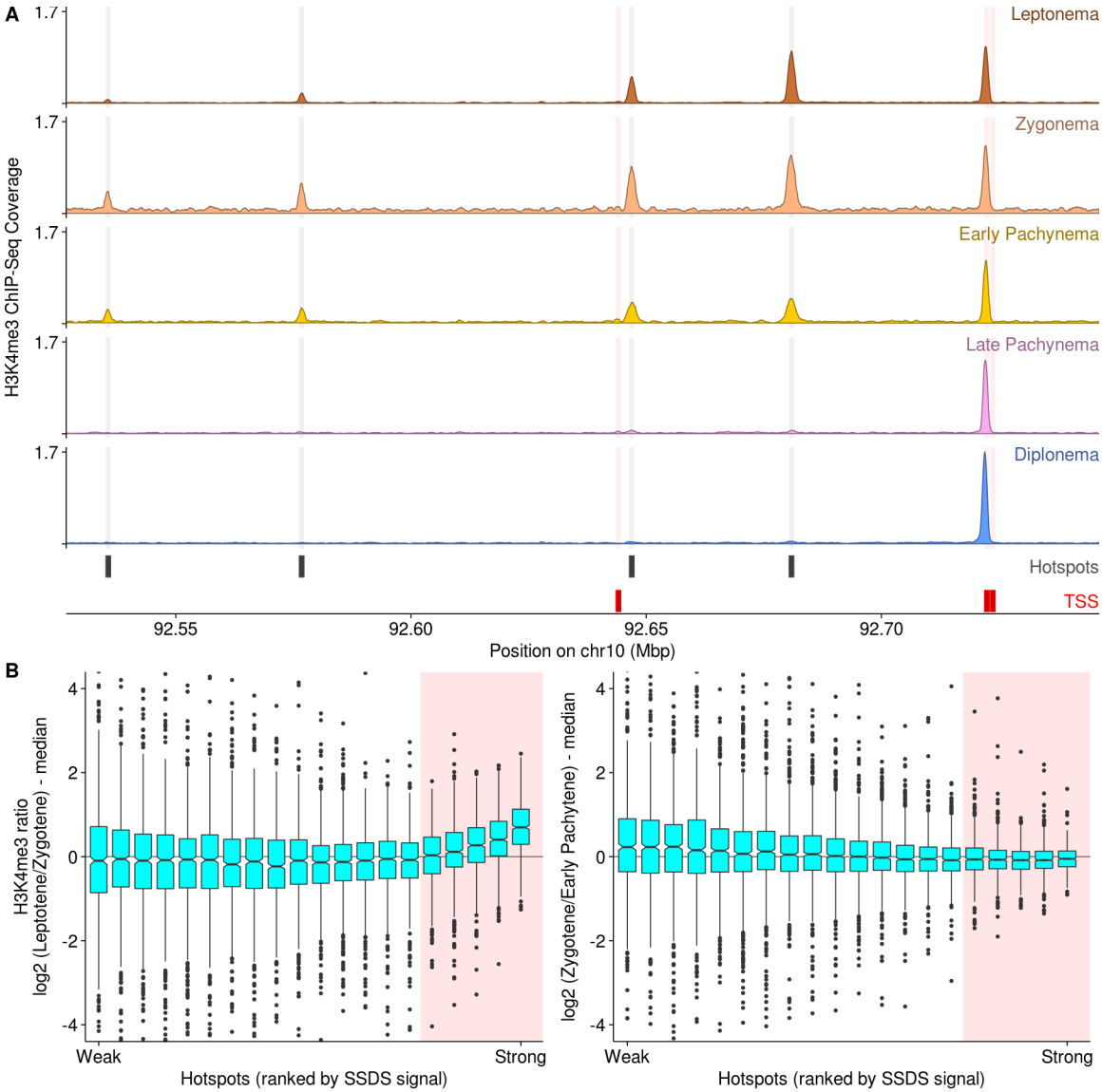


B (ii)

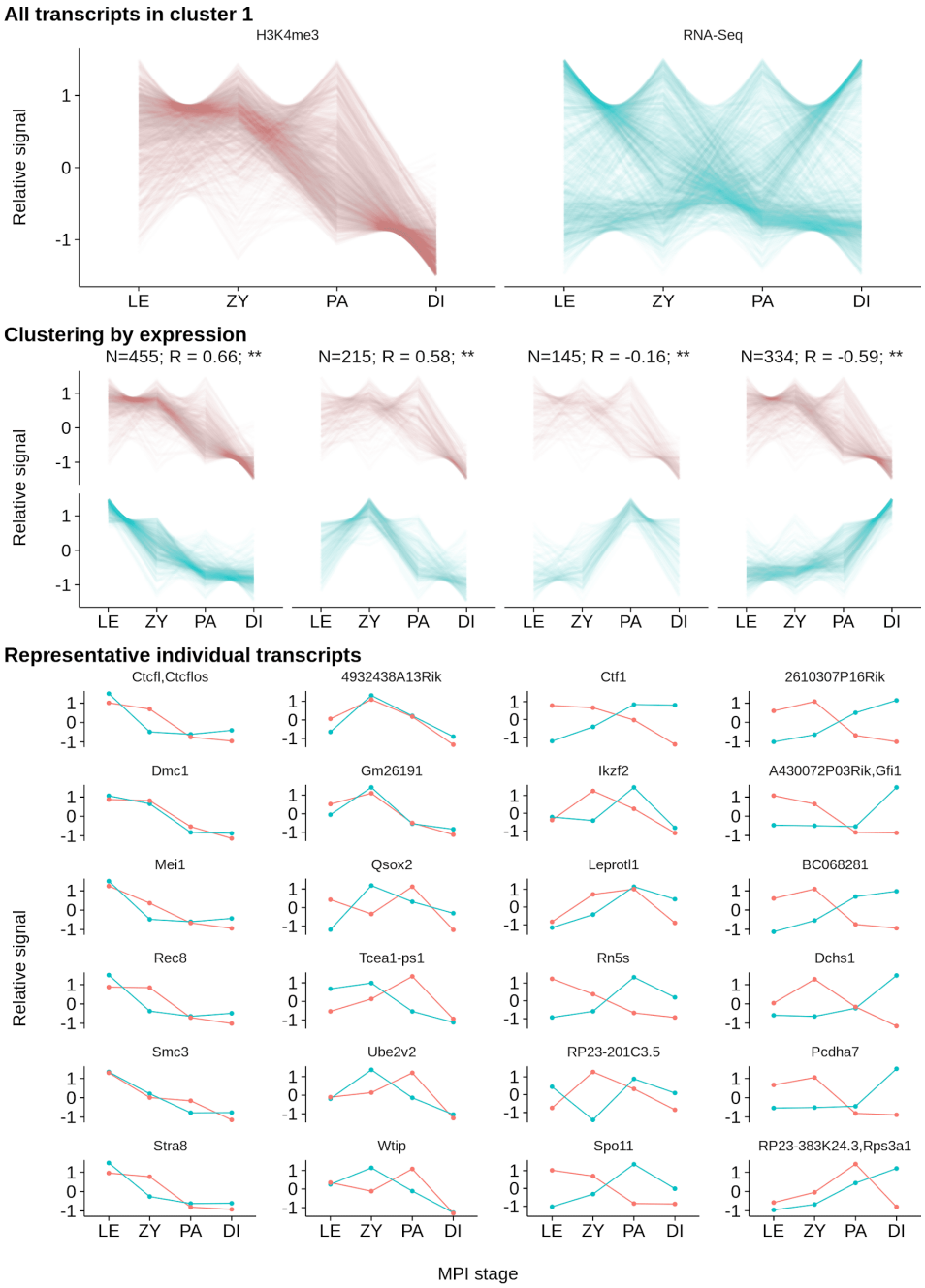


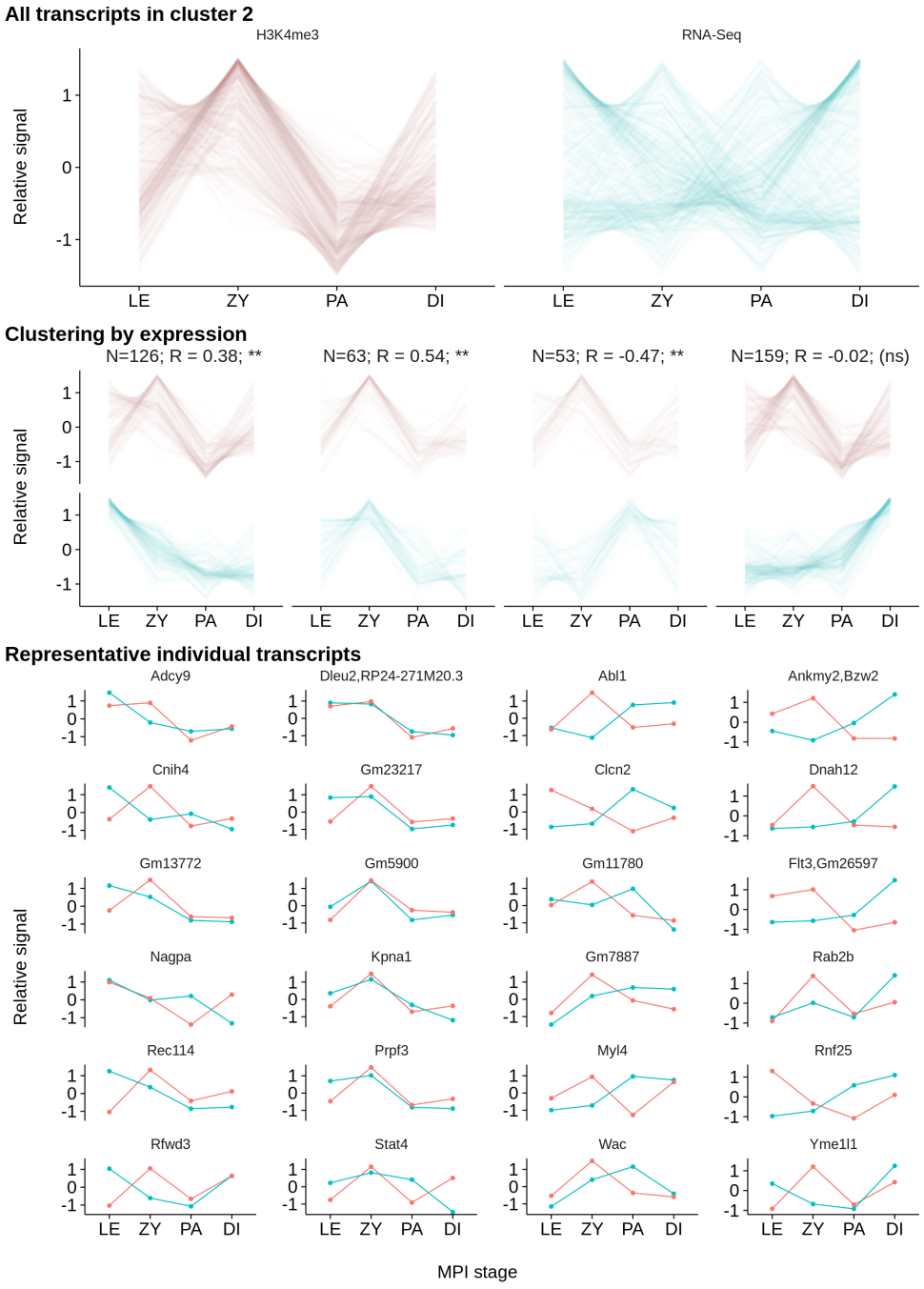
Supplementary Fig 1: Microscopic examination of sorted nuclei after FACS.

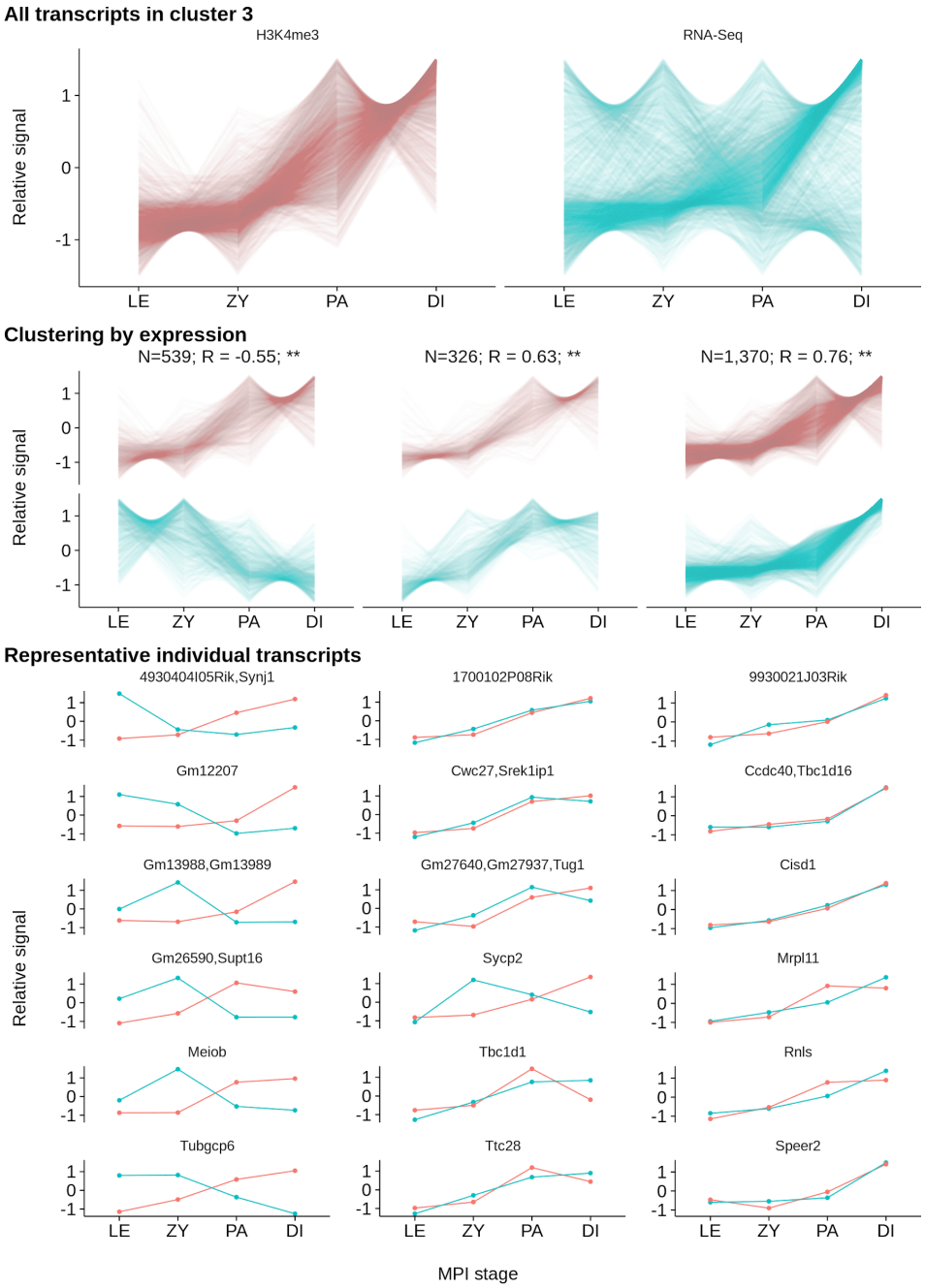
873 Whole-field immunofluorescence images of sorted nuclei in the populations of (A)
874 leptonema and (B) zygonema. Lep=leptonema; Zyg=zygonema; ePac=early
875 pachynema.
876
877



Supplementary Fig 2: 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 weakest hotspot (left) the H3K4me3 signal is maximal at Zygonema. At the strongest hotspot (right) the H3K4me3 signal is strongest at Leptonema. (B) Systematic trend towards stronger Leptonema H3K4me3 at stronger DSB hotspots. On the left panel, the H3K4me3 ratio between Leptonema and Zygonema is shown. Red shaded area shows the trend region. This trend is not seen when comparing Zygonema to Early Pachynema H3K4me3.


All transcripts in cluster 1
H3K4me3
RNA-Seq
Relative signal
LE
ZY
PA
DI
Clustering by expression
N=455; R = 0.66; **
N=215; R = 0.58; **
N=145; R = -0.16; **
N=334; R = -0.59; **
Representative individual transcripts
Ctcfl,Ctcflos
4932438A13Rik
Ctf1
2610307P16Rik
Dmc1
Gm26191
Ikzf2
A430072P03Rik,Gfi1
Mei1
Qsox2
Leprotl1
BC068281
Rec8
Tcea1-ps1
Rn5s
Dchs1
Smc3
Ube2v2
RP23-201C3.5
Pcdha7
Stra8
Wtip
Spo11
RP23-383K24.3,Rps3a1
MPI stage

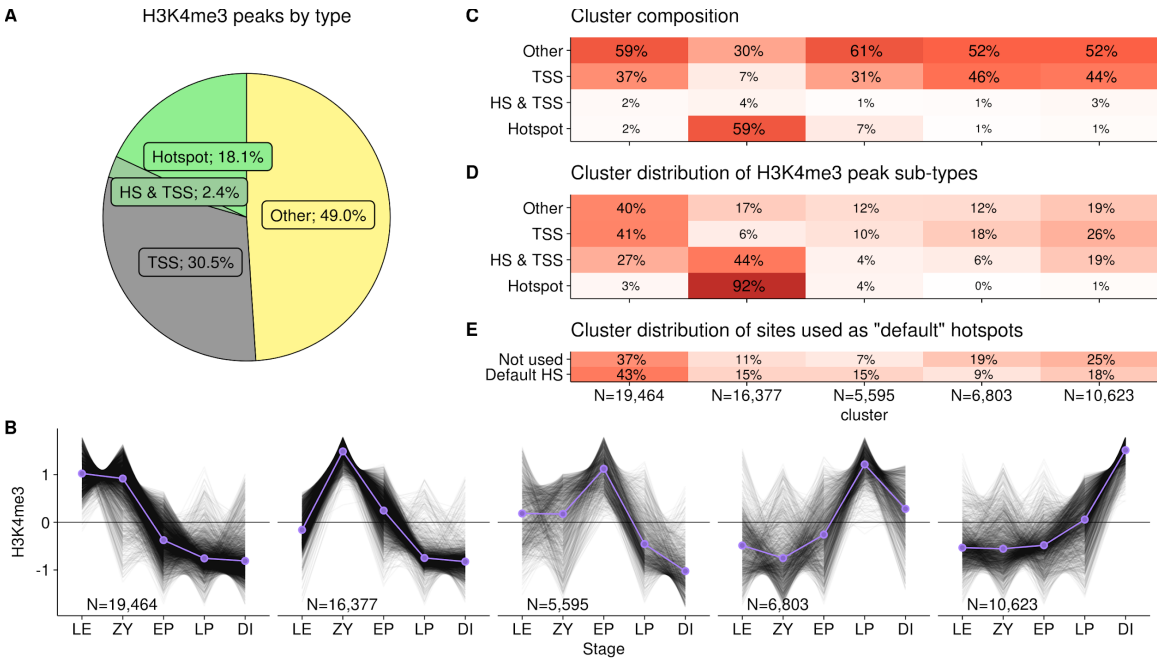

All transcripts in cluster 2
H3K4me3
RNA-Seq
Relative signal
LE
ZY
PA
DI
Clustering by expression
N=126; R = 0.38; **
N=63; R = 0.54; **
N=53; R = -0.47; **
N=159; R = -0.02; (ns)
Representative individual transcripts
Adcy9
Dleu2,RP24-271M20.3
Abl1
Ankmy2,Bzw2
Cnih4
Gm23217
Clcn2
Dnah12
Gm13772
Gm5900
Gm11780
Flt3,Gm26597
Nagpa
Kpna1
Gm7887
Rab2b
Rec114
Prpf3
Myl4
Rnf25
Rfwd3
Stat4
Wac
Yme1l1
MPI stage



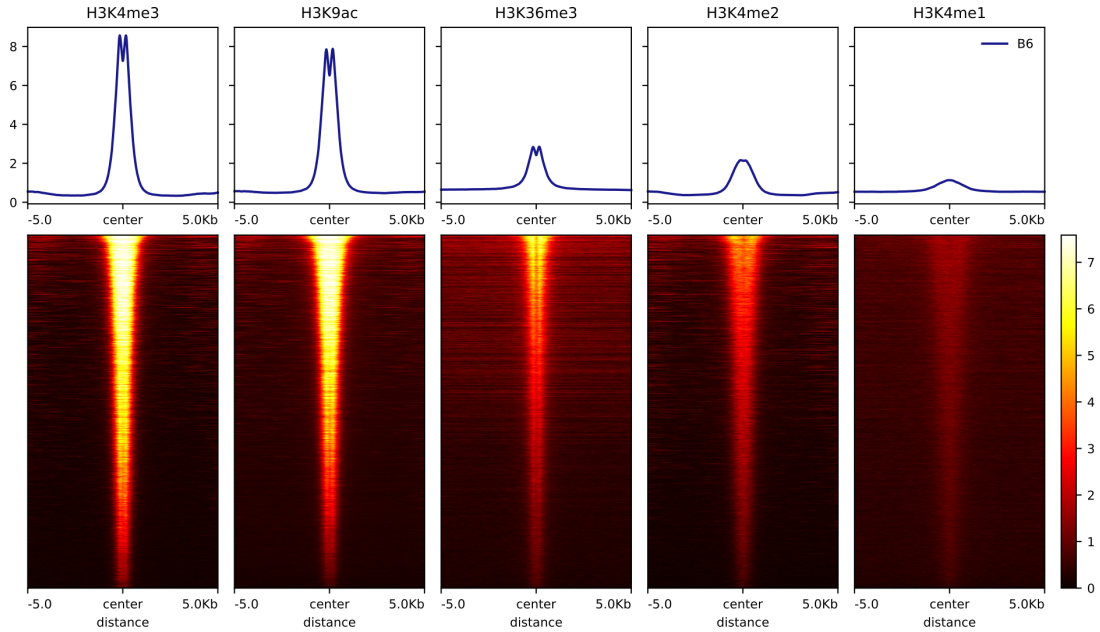
Supplementary Fig 3: H3K4me3 dynamics correlate with gene expression profiles from male juvenile mice.

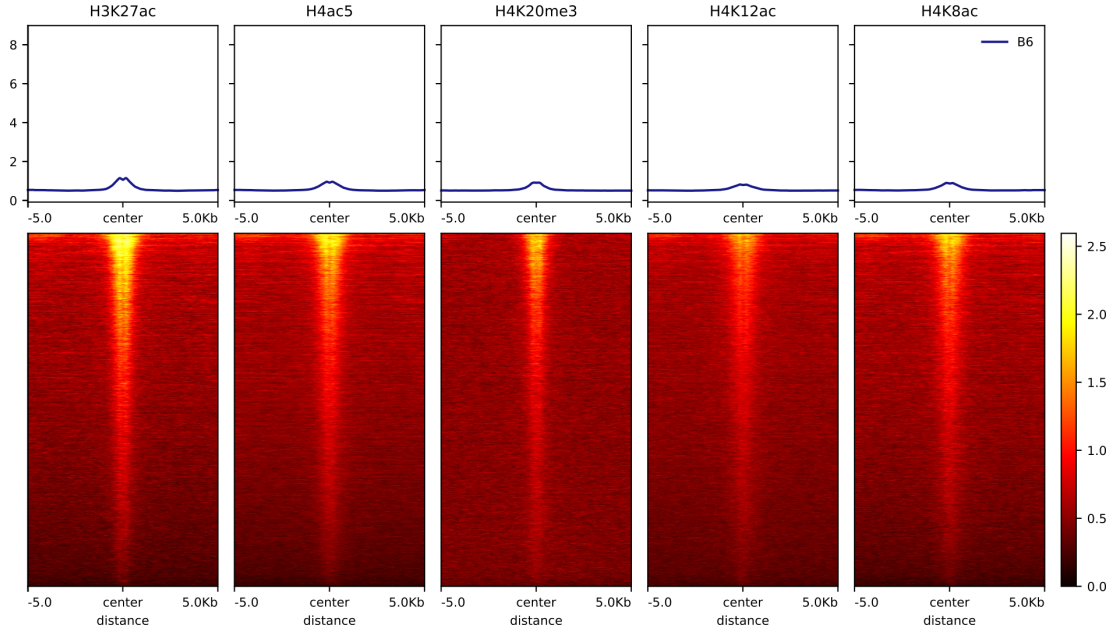
Each of the four plots represents a single cluster of transcripts from Fig. 2. Throughout, red depicts H3K4me3 ChIP-Seq signal and blue depicts gene expression from RNA-Seq. Each top panels shows all transcripts in the cluster; lines represent individual transcripts. Transcripts were next sub-clustered by gene expression profile (middle panels; k-means clustering; optimal k chosen

903 using gap statistic). Clusters in which profiles correlate more or less than
904 expected are highlighted (* = P < 0.01, ** = P < 0.001; 1,000 bootstraps of
905 shuffling stages for each gene). Six representative transcripts from each cluster
906 are shown in the lower panel below each cluster. Transcripts of genes known to
907 play a role in MPI were preferentially selected; if < 6 such genes were present in
908 a cluster, the remainder were randomly selected.

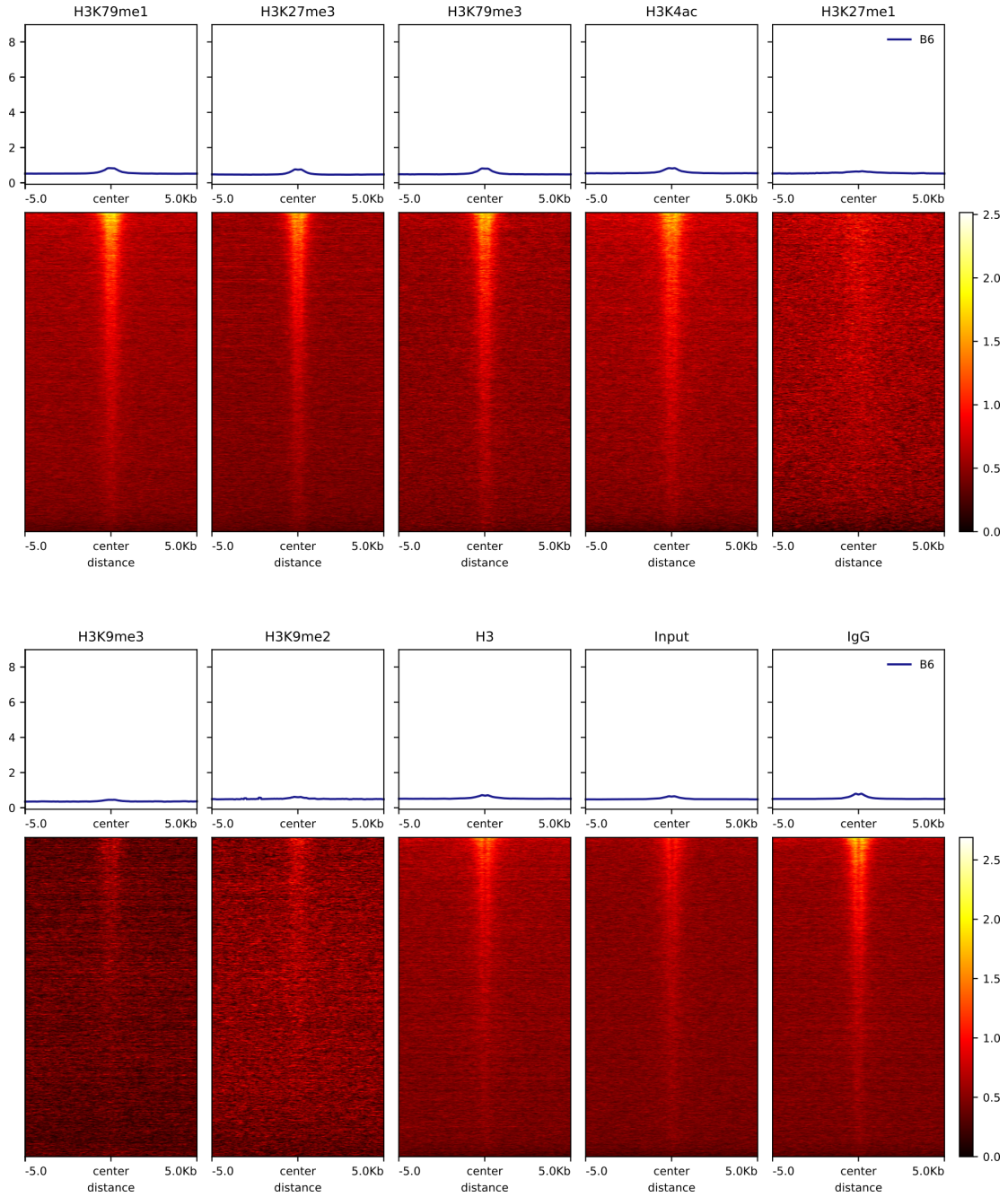


Supplementary Fig 4: Extensive dynamics revealed by unbiased clustering of H3K4me3 peaks. (A) Almost half (49%) of all H3K4me3 peaks occur at sites not defined at DSB hotspots or transcription start sites (TSS). These are classified as “Other”. 2.4% of H3K4me3 peaks coincide with both a DSB hotspot and a GENCODE TSS (HS & TSS). (B) Unbiased k-means clustering of MPI H3K4me3 profiles. 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). (C) Cluster composition shown as the percentage of each type of H3K4me3 peak within each cluster. Cluster 2 is the only cluster composed primarily of peaks at DSB hotspots. (D) The percentage of peaks of each type across clusters is shown. Notably, 92% of DSB hotspots occur in cluster 2. 44% 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[27]. Default hotspots form more frequently than expected at H3K4me3 peaks that are present in early MPI (cluster 1-3). Peaks with a late dynamic (cluster 4, 5) 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.


H3K4me3
H3K9ac
H3K36me3
H3K4me2
H3K4me1
B6
-5.0
center
5.0Kb
distance


H3K27ac
H4ac5
H4K20me3
H4K12ac
H4K8ac
B6
-5.0
center
5.0Kb
distance

932

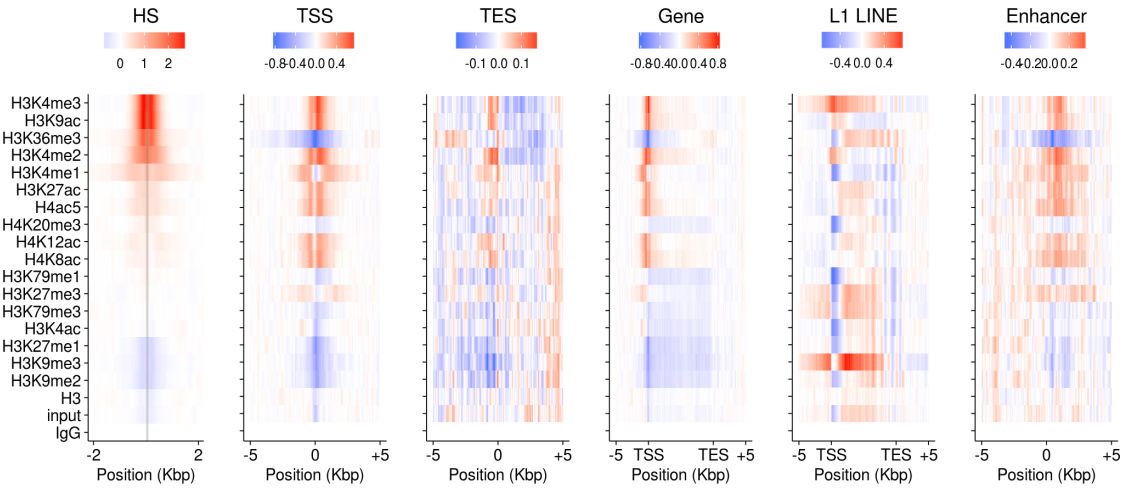


933

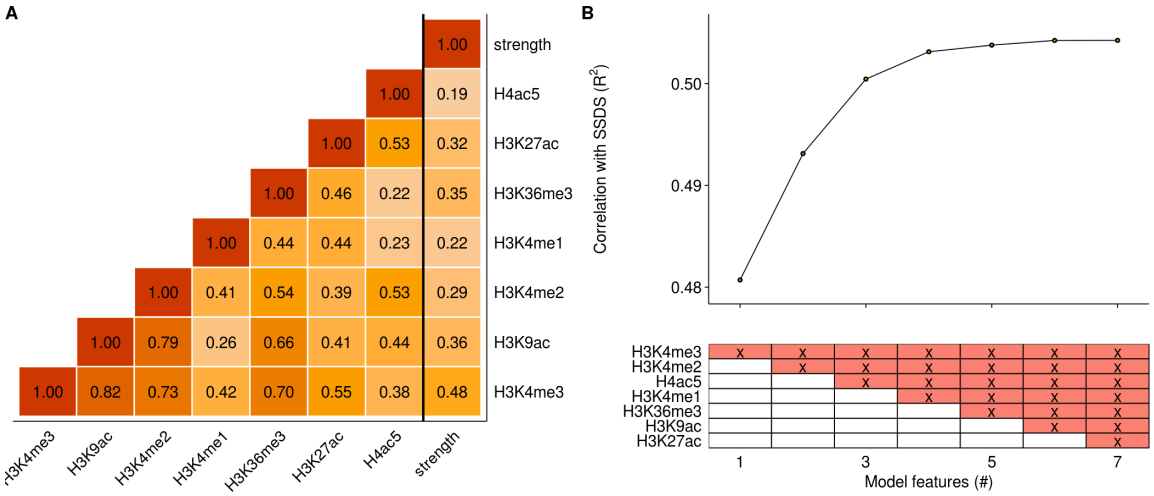
934

935 **Supplementary Fig 5: Heatmaps of histone modification ChIP-Seq at DSB**
936 **hotspots.** Coverage data were averaged in 150bp windows and converted to
937 reads per Kbp per million (RPKM) to facilitate cross-comparison. Heatmaps were
938 generated using deeptools[53]. It is important to note that some enrichment at DSB
939 hotspots is seen when using a non-specific antibody (IgG; last panel).

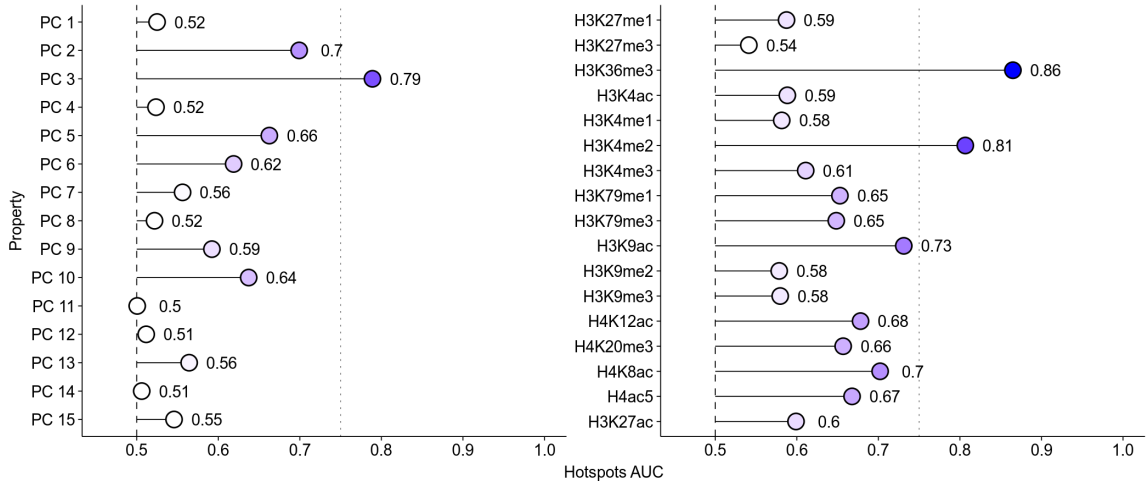
940



Supplementary Fig 6: 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 7: 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 8: Hotspots are distinguished from other H3K4me3 sites.

The area under the ROC curve (AUC) for discriminating hotspots from other H3K4me3 sites in the genome. AUCs for each principal component and for each histone modification are shown. Darker blue indicates higher AUC.