



HAL
open science

High-salt Recovered Sequences are associated with the active chromosomal compartment and with large ribonucleoprotein complexes including nuclear bodies

Marie Baudement, Axel Cournac, Franck Court, Marie Seveno, Hugues Parrinello, Christelle Reynes, Robert Sabatier, Tristan Bouchet, Zhou Yi, Sephora Sallis, et al.

► To cite this version:

Marie Baudement, Axel Cournac, Franck Court, Marie Seveno, Hugues Parrinello, et al.. High-salt Recovered Sequences are associated with the active chromosomal compartment and with large ribonucleoprotein complexes including nuclear bodies. *Genome Research*, 2018, 28 (11), pp.1733 - 1746. 10.1101/gr.237073.118 . hal-01912447

HAL Id: hal-01912447

<https://hal.science/hal-01912447>

Submitted on 5 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **High-salt Recovered Sequences are associated with the active chromosomal**
2 **compartment and with large ribonucleoprotein complexes including**
3 **nuclear bodies**

4
5 Baudement¹ M-O., Cournac² A., Court¹ F., Seveno¹ M., Parrinello³ H., Reynes⁴ C., Sabatier⁴
6 R., Bouschet⁴ T., Yi⁵ Z., Sallis¹ S., Tancelin¹ M., Rebouissou¹ C., Cathala¹ G., Lesne^{1,5} A.,
7 Mozziconacci^{5*} J., Journot^{3,4*} L. and Forné^{1*} T.

8
9 ¹ IGMM, Univ. Montpellier, CNRS, Montpellier, France.

10 ² Institut Pasteur, Paris, France.

11 ³ MGX, Univ. Montpellier, CNRS, INSERM, Montpellier, France.

12 ⁴ IGF, Univ. Montpellier, CNRS, INSERM, Montpellier, France.

13 ⁵ Sorbonne Université, CNRS, Laboratoire de Physique Théorique de la Matière Condensée, LPTMC, F-75252,
14 Paris, France.

15
16 * Correspondence should be sent to forne@igmm.cnrs.fr; Phone: +33 434 35 96 82,
17 or to laurent.journot@igf.cnrs.fr, mozziconacci@lptmc.jussieu.fr
18

19
20 **[Supplemental material is available for this article]**

21
22
23 **Running title:** The active compartment associates with large RNPs

24
25 **Keywords:** Active chromosomal compartment, super-enhancers, nuclear bodies, High-salt
26 Recovered Sequences
27

1 **Abstract**

2 The mammalian cell nucleus contains numerous discrete suborganelles named nuclear bodies. While
3 recruitment of specific genomic regions into these large ribonucleoprotein (RNP) complexes critically
4 contributes to higher-order functional chromatin organization, such regions remain ill-defined. We
5 have developed the HRS-seq method (High-salt Recovered Sequences-sequencing), a straightforward
6 genome-wide approach whereby we isolated and sequenced genomic regions associated with large
7 high-salt insoluble RNP complexes. Using mouse embryonic stem cells (ESC), we showed that these
8 regions essentially correspond to the most highly expressed genes, and to *cis*-regulatory sequences like
9 super-enhancers, that belong to the active A chromosomal compartment. They include both cell type-
10 specific genes, such as pluripotency genes in ESC, and housekeeping genes associated with nuclear
11 bodies, such as histone and snRNA genes that are central components of Histone Locus Bodies and
12 Cajal bodies. We conclude that High-salt Recovered Sequences are associated with the active
13 chromosomal compartment and with large ribonucleoprotein complexes including nuclear bodies.
14 Association of such chromosomal regions with nuclear bodies is in agreement with the recently
15 proposed phase separation model for transcription control and might thus play a central role in
16 organizing the active chromosomal compartment in mammals.

17

1 **Introduction**

2 The interphasic nucleus of mammalian cells is a highly compartmentalized organelle. Chromosome
3 Conformation Capture (3C)-derived technologies (Lieberman-Aiden et al. 2009) as well as molecular
4 imaging methods (Wang et al. 2016b) have revealed several layers of chromosome organization. At
5 the megabase (Mb) scale, chromosomes are segregated into active (A) and inactive (B) compartments
6 (median size ~3 Mb), while at the sub-megabase scale, they are partitioned into discrete
7 “Topologically Associating Domains” (TADs, median size ~880 kb) (Dixon et al. 2012; Nora et al.
8 2012). However, the molecular determinants and organization principles that control these two layers
9 of organization remain enigmatic. In contrast to TADs, chromosomal compartments are cell type-
10 specific, even if only a subset of genes is affected by A/B compartment changes during cell
11 differentiation (Dixon et al. 2012; Bonev et al. 2017). While cohesin and the CCCTC-binding factor
12 (CTCF) are required for TAD organization, the A/B chromosomal compartments remain intact upon
13 depletion of these factors, indicating that compartmentalization of mammalian chromosomes emerges
14 independently of proper insulation of TADs (Nora et al. 2017; Schwarzer et al. 2017). It has been
15 proposed that genome partitioning into chromosomal compartments may arise from contacts with
16 specific nuclear bodies or other important architectural components of the nucleus, such as the
17 nucleolus and the nuclear lamina for the B-compartment, or transcription factories for the A
18 compartment (Gibcus and Dekker 2013; Ea et al. 2015a).

19 Nuclear bodies are composed of large ribonucleoprotein (RNP) complexes self-assembled onto
20 specific chromatin regions, and recruitment of some genomic loci into nuclear bodies is known to be
21 crucial for proper gene expression (Mao et al. 2011). One emblematic example is the *U7* snRNA gene
22 that is recruited, together with histone genes (for which it is maturing the pre-mRNAs) into “Histone
23 Locus Bodies” (Frey and Matera 1995; Nizami et al. 2010). Impairment of nuclear body assembly has
24 been evidenced in several pathologies, including Spinal Muscular Atrophy (Sleeman and Trinkle-
25 Mulcahy 2014). Despite their importance for nuclear functions, the genomic sequences associated with
26 nuclear bodies remain largely unknown. Indeed, genomic profiling of such sequences is challenging
27 because purification of nuclear bodies is laborious and complex.

1

2 **Results**

3 **The HRS-seq method**

4 We previously showed that high-salt treatment of nuclei preparations allows the mapping of active
5 regulatory elements at mammalian imprinted genes (Weber et al. 2003; Braem et al. 2008; Court et al.
6 2011). More recently, extensive proteomic analyses have shown that high-salt treatments enable the
7 recovery of known protein components of nuclear bodies, such as the nucleolus, the Cajal bodies, or
8 the nuclear lamina (Engelke et al. 2014). We adopted this approach to develop a high-throughput
9 method aiming at profiling nuclear bodies-associated genomic sequences. The method, which avoids
10 formaldehyde crosslinking used in many currently available techniques (Dobson et al. 2017), involves
11 three experimental steps:

12 (i) The HRS assay make large RNP complexes, including nuclear bodies, insoluble through high-
13 salt treatments in order to trap, purify and sequence the genomic DNA associated with them
14 (HRS=High-salt Recovered Sequences) (Fig. 1A). A detailed protocol is given in the *Supplemental*
15 *Methods*. Briefly, a suspension containing 10^5 purified nuclei is placed onto an ultrafiltration unit and
16 is treated with a 2M NaCl buffer. Each nucleus forms a so-called “nuclear halo” composed of a dense
17 core containing insoluble complexes to which parts of the genomic DNA remain tightly associated,
18 surrounded by a pale margin of DNA loops corresponding to the rest of the genome (Fig. 1A). We
19 digested nuclear halos with the StyI restriction enzyme (for enzyme choice, see *Supplemental Methods*
20 and Supplemental Fig. S3C) and washed through the DNA loops (Loop fraction), leaving on the filter
21 the insoluble complex-associated fraction containing the High-salt Recovered Sequences (HRS-
22 containing fraction). Genomic DNA from each fraction is purified by proteinase K digestion,
23 phenol/chloroform extraction and ethanol precipitation.

24 (ii) Quality controls are performed in order to check the correct efficiency of each HRS assay. We
25 used quantitative (q)PCR reactions targeting two positive controls corresponding to DNA sequences
26 known to be constitutively enriched within the HRS-containing fraction in a wide range of
27 experimental conditions (Weber et al. 2003; Court et al. 2011). The enrichment level of these controls

1 (ratio of HRS to Loop fractions) was calculated for each HRS assay and normalized to the enrichment
2 level of a negative control (Weber et al. 2003).

3 (iii) The construction of DNA libraries for high-throughput sequencing is detailed in the *Methods*
4 section. Briefly, a first biotinylated StyI DNA adaptor, containing a binding site for the MmeI type IIS
5 restriction enzyme, is ligated to both sides of the StyI fragments (Fig. 1B). Ligated products are
6 captured onto streptavidin beads and digested with MmeI to homogenize the size of the StyI restriction
7 fragments (~18 to 20 nucleotides). A second sequencing adaptor is ligated to MmeI restricted sites and
8 DNA fragments are amplified on streptavidin beads using GEX PCR primers. The PCR reaction is
9 purified on an acrylamide gel and used for high-throughput sequencing (50-nucleotide single reads).

10 We applied our approach to the well-characterized e14Tg2a male mouse ESC (Gaspard et al.
11 2008). We made three HRS-seq experiments, each performed on a distinct ESC nuclei preparation
12 (biological replicates, see Fig. 1C). In order to obtain enough material, for each of the three nuclei
13 preparation, we selected 12 HRS assays displaying high enrichment levels of the positive controls (see
14 Supplemental Fig. S1 and *Methods* section). The HRS-containing fractions, on one side, or the Loop
15 fractions, on the other side, of these 12 assays were then pooled and 150ng of genomic DNA from
16 each pooled fractions were used for constructing sequencing libraries (see *Methods* section).
17 Therefore, for each of the three HRS-seq experiments, two sequencing libraries were prepared (HRS
18 and Loop fractions). Each pair of libraries thus represents a biological replicate since it is made from
19 only one of the three nuclei preparations (Fig. 1C).

20 The reads obtained from each fraction in each replicate were mapped to the reference genome of
21 e14Tg2a mouse ESC (129P2 built from the mm9 assembly) (see Table 1 in *Methods* section) and the
22 number of reads mapping to each StyI fragment was counted. Among a total of 3,053,742 StyI
23 fragments known in this reference mouse genome, 2,544,227 (83%) fragments were represented in the
24 experiments performed on ESC (509,515 StyI fragments were not be sequenced and/or their
25 corresponding reads did not map to a unique position on the mouse genome). Read counts of StyI
26 fragments in both the Loop and the HRS-containing fractions were highly reproducible between
27 biological replicates ($R > 0.90$) (Supplemental Fig. S2A/B and Supplemental Table S1) as well as in

1 control libraries (gDNA control) constructed from StyI digested genomic DNA ($R>0.90$)
2 (Supplemental Fig. S2D). In contrast, a poor correlation ($R=0.50$) was found between read counts
3 obtained from the HRS-containing and Loop fractions of each replicate, indicating that many StyI
4 fragments were efficiently segregated into one of the two fractions (Supplemental Fig. S2C/D and
5 Supplemental Table S1). Using the *edgeR* and *DESeq R* packages (Anders and Huber 2010; Robinson
6 et al. 2010), we determined, for each informative StyI fragment, the significance of the
7 overrepresentation of read counts in the HRS-containing fraction compared to the Loop fraction (see
8 *Supplemental Methods*). The same approach was used to determine the overrepresentation of reads
9 counts in the HRS-containing fraction compared to the gDNA control. As a result, 61,080 genomic
10 regions overrepresented in the HRS-containing fraction relative to the gDNA control and/or to the
11 Loop fraction have been identified in ESC (Benjamini-Hochberg corrected p -value <0.05)
12 (Supplemental Table S2). They were termed High-salt Recovered Sequences (HRS). This ESC HRS
13 set was used for subsequent bioinformatic analyses.

14

15 **HRS display chromosomal clustering**

16 We first looked at the size distribution (Supplemental Fig. S3A) and nucleotide composition
17 (Supplemental Fig. S3B) of the 61,080 ESC HRS. We found that they are barely different from those
18 obtained from 100 sets of 61,080 StyI fragments randomly selected in the mouse genome. We
19 conclude that HRS have sizes similar to those of regular StyI fragments and that their nucleotide
20 composition is not globally biased toward A/T or G/C-rich sequences, even if one can note that a small
21 subset of HRS is overrepresented in the range of 59% to 73% of G/C (p -value <0.01) (Supplemental
22 Fig. S3B). Globally, the G/C content of HRS is distributed around 43%, a value similar to the mean
23 G/C content of the mouse genome that corresponds to the value expected for sequences located around
24 regular StyI sites (Supplemental Fig. S3C). To demonstrate that the distribution of StyI sites in the
25 mouse genome does not introduce biases for HRS identification, we performed a correlation study
26 between StyI site density *versus* HRS density (*i.e.* the density of StyI sites associated with HRS) in
27 100 kb bins. This analysis showed that HRS density does not correlate with the density of StyI sites in

1 these bins (Spearman's correlation coefficient $R=0.129$) (Supplemental Fig. S3D). Consistently, StyI
2 density of HRS-containing bins is distributed around the mean StyI density in the mouse genome
3 (117.33 StyI/100 kb, vertical red line in Supplemental Fig. S3D). Overall, this demonstrates that HRS
4 are not specially found in bins with either high or low StyI density.

5 We then looked at the distribution of the 61,080 ESC HRS along mouse chromosomes and
6 found that they are spread over all chromosomes (Fig. 2A) with a mean genome-wide density of 23.47
7 HRS per megabase (Mb). However, the mean density of HRS was higher on chromosome 7, 11, 17,
8 and 19 and lower on chromosome 12, 14 and 18 (Fig. 2A). Furthermore, HRS seem to be not
9 uniformly distributed along the chromosomes, but they appeared to cluster at specific loci. To
10 demonstrate HRS clustering, we calculated the median distance between two consecutive HRS (6 kb)
11 and showed that it is much lower than the median distance obtained from 61,080 StyI fragments
12 randomly selected in the mouse genome (32 kb) (Fig. 2B). This analysis demonstrated that HRS are
13 highly clustered in the genome of mouse ESC.

14

15 **HRS are associated with the active A chromosomal compartment**

16 We then assessed whether HRS also cluster in the tridimensional (3D) space of the nucleus. Using
17 published Hi-C data obtained from mouse ESC (Dixon et al. 2012), we took the 100 kb bins most
18 highly enriched in HRS (hereafter called HRS bins) and calculated the mean score of
19 interchromosomal contact frequency for all possible pairs of HRS bins in these cells (see *Supplemental*
20 *Methods*). The score obtained was found to be significantly higher ($p < 10^{-2}$) than the scores obtained
21 from 100 sets of an equal number of 100 kb bins taken at random in the mouse genome (Fig. 2C, box-
22 plot on the left), thus demonstrating that HRS located on distinct chromosomes are closer together in
23 the 3D space of the nucleus. Among the 1125 HRS bins, 1102 (98%) were located in the active A
24 chromosomal compartment. HRS bins located in this compartment also have a contact score higher
25 than randomizations (Fig. 2C, box-plot on the right), indicating that HRS bins found within the A
26 compartment are also spatially clustered.

1 A global survey of ESC HRS in a genome browser then suggested that HRS are associated with
2 gene-rich regions and with the active A chromosomal compartment (Lieberman-Aiden et al. 2009;
3 Dixon et al. 2012) (Fig. 3A). To assess this point, we determined the overlap score between this
4 compartment and the HRS (*i.e.* the number of base pairs located in HRS and corresponding to the
5 active A chromosomal compartment, divided by the total number of base pair of the active
6 compartment in the entire mouse genome) and found that, for each chromosome, the overlap score is
7 systematically higher than the score obtained for a random set of StyI fragments. This demonstrates
8 that HRS are strongly associated with the active A compartment (Fig. 3B). In sharp contrast, HRS are
9 underrepresented in the inactive B compartment (Fig. 3C). We conclude that sequences identified by
10 HRS-seq correspond to regions essentially associated with the mouse active A chromosomal
11 compartment.

12

13 **HRS are associated with highly expressed genes and super-enhancers**

14 The preferential overlap of HRS with the active A chromosomal compartment (Fig. 3B) and their
15 weak overrepresentation in some G/C-rich sequences (Supplemental Fig. S3B), suggest that they
16 might be associated with CpG islands and gene-rich regions. Indeed, 4817 HRS (7.9%) are
17 overlapping with CpG islands, which is significantly different from the mean count (623 ± 26) (1.1%)
18 obtained from 1000 sets of 61,080 StyI fragments randomly selected in the mouse genome (Fig. 4A).
19 Moreover, we found 3625 genes for which at least one TSS is located inside a HRS in ESC,
20 henceforth termed HRS-associated genes (listed in Supplemental Table S3). A randomization analysis
21 showed that this number is much higher than expected by chance (632 ± 20) (Fig. 4B). Similar results
22 were obtained separately for each individual chromosome (Supplemental Fig. S4). In contrast, overlap
23 score analyses indicated that HRS are underrepresented into Lamina Associated Domains (LADs)
24 (Supplemental Fig. S5) which are associated with the inactive B compartment (Peric-Hupkes et al.
25 2010).

26 To assess whether HRS-associated genes belong to active or inactive genes, we used available
27 RNA-seq data from ESC (Wamstad et al. 2012) to design 3 sets of genes: the first set corresponds to

1 the 3000 most highly expressed genes, the second set to 3000 genes that display moderate expression
2 and the third set to the 3000 genes that display the weakest expression levels. This analysis showed
3 that HRS-associated genes are largely overrepresented in the first set of highly-expressed genes. In
4 contrast, the number of HRS-associated genes is comparable to those obtained from random sets in the
5 moderately-expressed gene set while they are strongly underrepresented in the weakly-expressed gene
6 set (Fig. 4C). A similar result was obtained when GRO-seq data from ESC (Min et al. 2011) were used
7 instead of RNA-seq data (Fig. 4D). Furthermore, on each chromosome, HRS are overrepresented in
8 exon sequences (Supplemental Fig. S6A) and underrepresented in introns (Supplemental Fig. S6B).

9 Using the list of all super-enhancers known in the mouse genome (Khan and Zhang 2016), we
10 found that super-enhancers are globally underrepresented in HRS (5225 are overlapping with HRS
11 while randomizations show that 5812 ± 78 should be expected) (Fig. 4E, left panel). However, among
12 the 231 super-enhancers that possess active epigenetic marks in ESC (Khan and Zhang 2016), 153
13 (66%) are found overlapping with HRS, and this number is much higher than expected by chance
14 (67 ± 6) (Fig. 4E, middle panel), while super-enhancers active in other cell types, like the cortex (Fig.
15 4E right panel), are not over-represented in ESC HRS. Therefore, super-enhancers active in the ESC
16 are strongly associated with ESC HRS.

17 Finally, using data available in the literature (ENCODE project), we showed that, in ESC, HRS
18 are not correlated with tri-methylation of lysine 9 on histone 3 (H3K9me3), that marks constitutive
19 heterochromatin. In contrast, they overlap with tri-methylation of lysine 36 on histone 3 (H3K36me3)
20 (Fig. 5A/B), which marks transcriptionally active exon regions (Hon et al. 2009). This latter result was
21 confirmed on each chromosome using appropriate randomizations (Supplemental Fig. S6C). We
22 conclude that HRS are associated with TSS of highly expressed genes and active super-enhancers.

23

24 **HRS-associated genes are housekeeping as well as cell type-specific genes**

25 Using DAVID functional annotation tool (Huang da et al. 2009), we carried out Gene Ontology (GO)
26 analyses on genes with HRS-associated TSS (Supplemental Table S4). Most of the ontology terms
27 correspond to housekeeping genes often linked to known nuclear bodies, with terms such as “covalent

1 chromatin modification” ($p=3.2 \cdot 10^{-7}$), “intracellular RNP complexes” ($p=1.5 \cdot 10^{-20}$), “spliceosome”
 2 ($p=3.9 \cdot 10^{-7}$), “nucleolus” ($p=1.6 \cdot 10^{-25}$), “cell cycle” ($p=9.3 \cdot 10^{-17}$) and also “nuclear speckles” ($p=3.3$
 3 10^{-3}) and “promyelocytic leukemia (PML) body” ($p=3.9 \cdot 10^{-2}$) (Fig. 6). We also noted the term “stem
 4 cell population maintenance” ($p=2.6 \cdot 10^{-4}$) (black arrow in Fig. 6), which reflects the presence in ESC
 5 HRS of the TSS of many pluripotency genes, *e.g.* *Nanog*, *Tet1* or *Sox2*, which are very highly
 6 expressed in ESC (Wamstad *et al.* 2012). HRS-seq data indicated the presence of HRS at the *Sox2*,
 7 *Klf4*, *Pou5f1* and *Nanog* loci (Fig. 5A and Supplemental Fig. S7). In contrast, DAVID ontology
 8 analysis of highly-expressed genes that are not associated with HRS (1656 genes among the 3000
 9 genes used in Set1 of Fig. 4C) indicates that they essentially correspond to housekeeping genes
 10 involved in cell metabolism or cytoskeleton and membrane associated processes, and no indication of
 11 cell-type specific or nuclear body-associated genes could be evidenced (Supplemental Fig. S8A and
 12 Supplemental Table S5). Using the *i-cisTarget* tool (Herrmann *et al.* 2012), we then showed that the
 13 promoters of highly expressed genes associated with the HRS preferentially bind cell-cycle regulators
 14 of the E2F family, while the promoters of highly-expressed genes not associated with HRS bind a
 15 whole series of factors belonging to the ETS family (ELF, ELK, GABPA...) (Supplemental Fig. S8B
 16 and Supplemental Table S6). This suggests that many HRS-associated genes are tightly regulated
 17 during cell cycle progression.

18 Pluripotency genes are well-known to be largely repressed when ESC are differentiated into
 19 cortical neurons (see for example (Bonev *et al.* 2017)). To further assess the functional significance of
 20 the association of pluripotency genes with HRS, we differentiated ESC into cortical neurons (Gaspard
 21 *et al.* 2009). In both cell types, we performed detailed analyses at the *Sox2*, *Pou5f1*, *Nanog* and *Klf4*
 22 loci using quantitative PCR to determine the relative enrichment levels in the HRS-containing fraction
 23 of StyI fragments spread along these loci (HRS-qPCR experiments). At the *Sox2* locus, we found that
 24 the enrichment levels within the gene body are 11 to 23 times higher than the mean local background
 25 in ESC (Fig. 7A) (see *Supplemental Methods* for background definition), but they fall to 5 times the
 26 mean local background in cortical neurons (Fig. 7B). This region, which also maps with super-
 27 enhancers (Khan and Zhang 2016), is known to contain two *Sox2 Regulatory Regions* (SRR) (Zhou *et*

1 al. 2014). As suggested by HRS-seq data (Fig. 5A), a second HRS region was found 107 kb
2 downstream of the gene. It corresponds to a known super-enhancer (Khan and Zhang 2016), the *Sox2*
3 *Regulatory Region* 107 (SRR107), which is required to maintain a high expression level of this gene in
4 ESC (Zhou et al. 2014). Its enrichment level was 3 times lower in neurons (Fig. 7B) than in ESC (Fig.
5 7A) confirming the previously described cell specificity of this region (Zhou et al. 2014). Similar
6 results were obtained at the *Pou5f1* and *Nanog* loci, where enrichment levels of the gene body and
7 associated super-enhancers were high in ESC but drastically reduced in neurons (Supplemental Fig.
8 S9A/D). At the *Klf4* locus, the gene body also displays much lower enrichment levels in neurons than
9 in ESC. However, the promoter as well as some part of the associated super-enhancer remained highly
10 enriched in neurons (Supplemental Fig. S9E/F), suggesting that this locus may remain associated to a
11 large ribonucleoprotein complex in cortical neurons even if transcription levels are largely reduced in
12 this cell type.

13 Overall, these results indicate that, while the vast majority of genes-associated HRS correspond to
14 housekeeping activities, some of them correspond to cell type-specific genes (such as pluripotency
15 genes in ESC) and to their *cis*-regulatory sequences (such as super-enhancers).

16

17 **HRS include nuclear body-associated sequences**

18 Former proteomic analyses indicated that high-salt treatments of nuclei preparations enable the
19 recovery of protein components of nuclear bodies (Engelke et al. 2014). To check whether nuclear
20 bodies are also recovered under our experimental conditions, we performed immunofluorescence
21 microscopy on nuclear halos prepared as described above for the HRS assays (also see *Method*
22 section). These experiments showed that Coilin, SMN (Survival of Motor Neuron) and PML foci are
23 present within the insoluble material obtained after high-salt treatment of our nuclei preparations (Fig.
24 8A), indicating that PML bodies as well as nuclear bodies of the Cajal body family (Histone Locus
25 Bodies, gems or Cajal bodies) are indeed retained under these experimental conditions.

26 Since nuclear bodies remain within the high-salt insoluble material of the HRS-containing
27 fraction, we assessed whether known nuclear body-associated sequences are present among HRS.

1 From the list of the 3625 TSS that map to HRS in ESC (Supplemental Table S3), Gene Ontology
2 analyses showed that 14 are associated with the term “nucleosome assembly” ($p=2.4 \cdot 10^{-3}$) (Fig. 6).
3 This term is linked to the presence of histone genes known to be an essential component of the Histone
4 Locus Bodies, a class of nuclear bodies that share strong structural similarities with Cajal bodies
5 (Nizami et al. 2010). Indeed, 35 U7-dependent histone genes (whose mRNA maturation involves the
6 U7 snRNA) among 73, but only 2 TSS of histone variant genes among 15 (including *H2afx*), are
7 found into HRS. In the mouse, histone genes are clustered into three major histone loci: the *Hist1*
8 locus on chromosome 13qA3.1, the *Hist2* locus on chromosome 3qF1-3qF2.1 and the *Hist3* locus on
9 chromosome 11qB1.3. TSS of genes belonging to all three clusters are found associated with HRS.
10 Therefore, in ESC, histone genes are strongly associated with HRS.

11 Given the strong association of the *Hist1* locus genes with HRS (Supplemental Fig. S10A), we
12 performed a detailed analysis of this locus by HRS-qPCR. We found that the enrichment levels along
13 the whole *Hist1* locus, from *Hist1h4h* to *Hist1hle* genes, are 1.5 to 4 times higher than the mean local
14 background (Fig. 8B). While devoid from any coding genes, the upstream part of the *Hist1* locus
15 nevertheless displayed the highest enrichment levels, with some values rising to almost 100 times
16 above the local background. These sequences, having high enrichment levels in the HRS-containing
17 fraction, systematically contain tRNA genes. We conclude that, at the *Hist1* locus, both histone genes
18 and tRNA genes have high enrichment levels within the HRS-containing fraction.

19 Since, at the *Hist1* locus, the tRNA genes appear highly enriched within the HRS-containing
20 fraction, we assessed whether this is particular to this locus or whether a wider association of HRS
21 with tRNA genes exists in the mouse genome. Because tRNA genes are classified among repeat
22 sequences, we analyzed the repeat content of the HRS. The enrichment levels of 1554 mouse repeat
23 families (UCSC classification) in the HRS were calculated and compared to the mean enrichment level
24 obtained from 1000 random sets with the same number of StyI fragments. Repeats families that are
25 significantly enriched ($p\text{-value}<10^{-3}$) in HRS belong to 9 classes including SINE repeats, that are
26 known to be overrepresented in gene-rich regions and in the A compartment (Cournac et al. 2016), as
27 well as *snRNA* genes and tRNA genes (Fig. 8C). We conclude that HRS are associated with snRNA

1 and tRNA genes as well as SINE repeats, at the genome-wide level. This observation suggests that
2 some repeat regions are associated with large RNP complexes in the mouse nucleus.

3 HRS-seq data indicated that one of the largest HRS clusters in ESC maps to the *Malat1*
4 (Metastasis associated lung adenocarcinoma transcript 1) / *Neat1* (Nuclear paraspeckle assembly
5 transcript 1) locus (Supplemental Fig. S10B). Indeed, both genes are found in the HRS-associated
6 genes (Supplemental Table S3). Given the importance of these genes and transcripts for the assembly
7 of nuclear speckles and paraspeckles (Hutchinson et al. 2007; Clemson et al. 2009), we performed
8 HRS-qPCR experiments and showed that both genes are strongly enriched within the HRS-containing
9 fraction throughout their gene bodies and that enrichment levels for their TSS are particularly high
10 (Fig. 8D). Therefore, nuclear speckles and paraspeckles might also be retained in HRS assays.

11 Overall, these results confirm that nuclear halos contain insoluble nuclear bodies and show that
12 many HRS correspond to genomic regions known to be associated with nuclear bodies, including the
13 *Hist 1* locus that is part of the Histone Locus Bodies (Nizami et al. 2010), as well as *snRNA* genes that
14 are integral components of Cajal bodies, *tRNA* genes that are known to contact the perinucleolar
15 compartment (Nemeth et al. 2010; Padeken and Heun 2014) and the *Malat1/Neat1* genes that are
16 required for the assembly of nuclear speckles and paraspeckles.

17 We conclude that the sequences identified by HRS-seq correspond to genomic regions associated
18 with large high-salt insoluble RNP complexes, including nuclear bodies, that display preferential
19 physical proximity and association with the mouse active A chromosomal compartment.

20

21 **Discussion**

22 The HRS-seq method provides an original genome-wide approach to identify genomic sequences
23 physically associated *in vivo* with large RNP complexes, including several nuclear bodies. Our method
24 is probing higher-order chromatin architecture at the supranucleosomal scale. It is therefore clearly
25 different from previous methods such as FAIRE-seq (Giresi et al. 2007), ATAC-seq (Buenrostro et al.
26 2013) or MNase-seq (Henikoff et al. 2011; Valouev et al. 2011; Gaffney et al. 2012) that all aim at
27 investigating accessibility of the chromatin nucleofilament (nucleosomal scale). Contrary to most

1 genome-wide approaches developed so far to investigate higher-order chromatin architecture (*e.g.*
2 DamID mapping or Hi-C) (Vogel et al. 2007; Lieberman-Aiden et al. 2009), the HRS-seq method is
3 based on a simple and straightforward principle: high-salt treatments of nuclei preparations. Previous
4 works in *Drosophila* (Henikoff et al. 2009), as well as proteomic analyses (Engelke et al. 2014) and
5 immunofluorescence microscopy (Dobson et al. 2017) (Fig. 8A) indicate that such treatments make
6 large RNP complexes insoluble. Combined with restriction digestion (StyI) and ultrafiltration (Fig. 1),
7 they allow an easy separation of DNA sequences that are trapped within such complexes (HRS-
8 containing fraction) from sequences that are not interacting with them (Loop fraction). Our method
9 avoids long purification procedures of nuclear bodies that may bias retention of genomic sequences.
10 Such approaches proved to be efficient for the identification of the Nucleolus Associated Domains
11 (NADs) (Nemeth et al. 2010), but they were so far mostly unsuccessful for other nuclear bodies. The
12 HRS-seq method is also avoiding delicate chemical crosslinking procedures (Dobson et al. 2017) or
13 the use of specific antibodies that may restrict retention of some genomic sequences. It should thus be
14 very helpful to explore in further detail the impact of this level of chromatin organization on gene
15 regulation and cell fate determination in a variety of physiological and pathological situations.

16 We have here performed the first global genomic profiling of High-salt Recovered Sequences
17 (HRS) and found that, in mouse ESC, the major components of HRS are (i) the histone genes and
18 snRNA genes, that are known to contact the Histone Locus Bodies and the Cajal bodies respectively,
19 (ii) tRNA genes, that are known to spatially cluster into the perinucleolar compartment (Thompson et
20 al. 2003; Nemeth et al. 2010), and (iii) many transcriptionally active genes contacting large RNP
21 complexes that may correspond to RNA polymerase II foci. This latter result is in agreement with
22 pioneering works in human HeLa cells (Linnemann et al. 2009) and *Drosophila* S2 cells (Henikoff et
23 al. 2009) showing that salt-insoluble chromatin is enriched in actively transcribed regions. However,
24 our method replaces nuclease treatments (Henikoff et al. 2009) and microarray-based profiling
25 (Henikoff et al. 2009; Linnemann et al. 2009) used in these earlier works by restriction digestions and
26 a high-throughput sequencing approach, which allow a more powerful and high resolution genome-
27 wide profiling. Recent genome-wide identification of the so-called Matrix Attachment Regions

1 (MAR-seq) in human mammary epithelial cells showed that MARs are A/T-rich sequences that are
2 overrepresented in the active A chromosomal compartment, even though no correlation was found
3 between such MARs and active or inactive epigenetic marks (Dobson et al. 2017). While MAR-seq
4 technique also involves high-salt treatments, it makes use of extensive crosslinking, and therefore, like
5 Chromosome Conformation Capture (3C) assays, it may also capture sequences involved in long-
6 range chromatin contacts that are not necessarily associated with large RNP complexes. HRS are not
7 only associated with actively transcribed regions (Fig. 4C), but also with cell-type specific super-
8 enhancers (Fig. 4E) and, opposite to MARs (Dobson et al. 2017), with genomic regions that display
9 active epigenetic marks like H3K36me3 (Fig. 5B). Finally, HRS are not biased toward A/T-rich
10 sequences (Supplemental Fig. S3B) and therefore they appear to be clearly distinct from A/T-rich
11 sequences like MARs or LADs. Indeed, although lamins are known to remain associated with the
12 insoluble material upon high-salt extractions of nuclei preparations (Engelke et al. 2014), HRS are
13 underrepresented in LADs (Supplemental Fig. S5). Since it has been demonstrated that lamins are not
14 required for LAD organization in mouse ESC (Amendola and van Steensel 2015), it may be possible
15 that LAD organization in these cells depends on a factor that is soluble upon high-salt treatments.

16 Our work shows that genomic sequences that are associated with large RNP complexes (HRS) are
17 in close proximity in the 3D space of the nucleus, and that such sequences are overrepresented in the
18 active A chromosomal compartment. It has been suggested that this latter level of chromatin
19 organization may be coordinated through contacts with some nuclear bodies (Gibcus and Dekker
20 2013; Ea et al. 2015a). It is indeed known that chromosomal compartments are established during the
21 early G1 phase of the cell cycle, at the so-called “Timing Decision Point” (TDP) (Dileep et al. 2015),
22 when replication-timing programs are fixed and several major nuclear bodies, like Cajal bodies, are
23 reassembled (Carmo-Fonseca et al. 1993). Recruitment into nuclear bodies may confine specific
24 chromatin regions, thus limiting their diffusion into the nuclear space and favouring functional
25 interactions required for genomic regulations during the interphase. This is particularly true for long-
26 range inter-TAD chromatin interactions, since chromatin dynamics at this level displays extremely
27 low contact frequencies while being essential for many genomic functions (Dixon et al. 2012; Nora et

1 al. 2012; Rao et al. 2014; Ea et al. 2015b). Using 4C-seq, it was for example shown that Cajal bodies-
2 associated regions are enriched in highly expressed histone genes and snRNA loci, thus forming intra-
3 and inter-chromosomal clusters (Wang et al. 2016a). In the interphasic cell, chromosomal partitioning
4 into the active or inactive compartments is cell type-specific (Lieberman-Aiden et al. 2009). Hi-C
5 experiments using an ESC differentiation model (Wamstad et al. 2012) have recently suggested that
6 such hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in
7 cellular differentiation (Fraser et al. 2015; Bonev et al. 2017). Pluripotency genes (including *Sox2*,
8 *Pou5f1*, *Nanog* and *Klf4*) that are highly transcribed in ESC (Wamstad et al. 2012) are found among
9 HRS in this cell type (Fig. 5A and Supplemental Fig. S7), but their association is drastically reduced in
10 cortical neurons (Fig. 7 and Supplemental Fig. S9). This observation indicates that HRS-associated
11 genes do not only correspond to highly expressed housekeeping genes, but also to cell type-specific
12 genes that require high transcription levels. HRS also include some *cis*-regulatory elements required
13 for maintaining high expression levels, like cell-type specific super-enhancers (Fig. 4E) and the
14 SRR107 region found at the *Sox2* locus (Fig. 7). This region, which is also described as a super-
15 enhancer (Wei et al. 2016), is located within a major distal cluster of enhancers, named the *Sox2*
16 *Control Region* (SCR) (Zhou et al. 2014). It has been shown by 3C and Hi-C experiments that the
17 SCR is required for maintaining high *Sox2* expression levels in ESC through long-range chromatin
18 interactions with this gene (Zhou et al. 2014; Stadhouders et al. 2018). Furthermore, only the most
19 highly expressed genes are found overrepresented in HRS, while genes that are expressed at moderate
20 or weak expression levels are not (Fig. 4C). These findings are in agreement with the recently
21 proposed phase separation model for transcription control (Hnisz et al. 2017) suggesting the existence
22 of a cell-type specific transcriptional compartment where a subset of genes and their regulatory
23 elements, including super-enhancers, are associated with large RNP complexes allowing high
24 expression levels. Such complexes are likely to correspond to RNA polymerase II foci visualized by
25 immunofluorescence microscopy (transcription factories/active chromatin hubs). However, given that
26 sequences at the *Malat1/Neat1* gene locus are found highly enriched within the HRS-containing
27 fraction (Fig. 8D), one possibility could be that they correspond to RNA polymerase II complexes

1 contacting nuclear speckles/paraspeckles, serving as “hubs” to link active transcription sites
2 (Sutherland and Bickmore 2009; Cook 2010; Mao et al. 2011; Quinodoz et al. 2018).

3 Overall, our results provide a strong experimental support in favour of a model whereby nuclear
4 bodies, and/or large RNP complexes associated with RNA polymerase II, play an important role in
5 organizing the active chromosomal compartment through recruitment of highly expressed genes,
6 including housekeeping and cell type-specific genes with their *cis*-regulatory regions.

7

8 **Methods**

9 **Cell culture**

10 Cultures and *in vitro* corticogenesis of mouse ESC (e14Tg2a strain, 129P2 genomic background) were
11 performed as previously described (Gaspard et al. 2009) (see *Supplemental Methods* for details). Cells
12 were tested for the absence of mycoplasma contamination and their identity was confirmed by
13 immunofluorescence microscopy (Supplemental Fig. S11 and *Supplemental Methods*). All
14 experimental designs and procedures are in agreement with the guidelines of the animal ethics
15 committee of the French “Ministère de l’Agriculture” (European directive 2010/63/EU).

16

17 **HRS assay**

18 Nuclei preparations used for HRS assays were made from undifferentiated mouse ESC or neurons as
19 previously described for C2C12 myoblasts (Milligan et al. 2000). Such nuclei preparations are snap
20 frozen into liquid nitrogen and can be stored at -80°C for several months. They were formerly used for
21 nuclear run-on experiments to investigate transcriptional activity of mammalian genes (Milligan et al.
22 2000; Milligan et al. 2002). The HRS assays were adapted from our previous publications (Weber et
23 al. 2003; Braem et al. 2008) (see *Supplemental Methods* for details).

24

25 **Real-time quantitative PCR and quality check**

26 The quality of each HRS assay was checked by real-time quantitative PCR targeting StyI fragments
27 that are known, from previously published works (Court et al. 2011), to be either highly enriched

1 (positive control) or not enriched (negative control) in the HRS-containing fraction relative to the
 2 Loop fraction in diverse experimental conditions. Primer sequences used for HRS-qPCR analyses at
 3 *Sox2*, *Pou5f1*, *Nanog*, *Klf4*, *Histone 1* and *Malat1/Neat1* loci (Fig. 7, Supplemental Fig. S9 and Fig.
 4 8B/D) are given in Supplemental Tables S7, S8, S9, S10, S11 and S12 respectively.

5 The enrichment levels were calculated as the ratio of the amount of DNA target in the HRS-containing
 6 fraction *versus* the Loop fraction. They were normalized to the local background level according to an
 7 algorithm adapted from a previous work (Braem et al. 2008) (see *Supplemental Methods* for details).

8

9 **HRS-seq library construction**

10 HRS assays were performed from three distinct ESC nuclei preparations. For each nuclei preparation,
 11 DNA extracted from Loop fractions on one side and from HRS-containing fractions on the other side
 12 of 12 high-quality HRS assays were pooled (see Fig. 1C). These samples were used to make HRS-seq
 13 libraries (Supplemental Fig. S1C/B). Each biological replicate was thus composed of two HRS-seq
 14 libraries: one built from the DNA pooled from the Loop fractions and one built from that pooled from
 15 the HRS-containing fractions. Construction of HRS-seq libraries is done as follows: DNA samples are
 16 first re-digested with StyI (Eco130I at 10u/μl, Fermentas ref ER0411) in order to ensure complete
 17 digestion. 30 pmol of biotinylated adaptor 1 with complementarity for StyI restriction sites (5'P-
 18 **CWWGTCGGACTGTAGAACTCTGAACCTGTCCAAGGTGTGA-Biotin-3'** and 3'-
 19 **AGCCTGACATCTTGAGACTTGGACA-5'**) are ligated during 15 min. at Room Temperature (RT)
 20 to 150 ng of StyI digested genomic DNA (Quick Ligation™ Kit, NEB ref. M2200S). 100 μg of
 21 streptavidine beads (Dynabeads© MyOne™ Streptavidin C1 from Invitrogen, ref. 650.01) are
 22 resuspended into 50μl of BW 2X buffer (10 mM Tris-HCl pH 7.5; 1mM EDTA; 2M NaCl) and the
 23 adaptor 1 ligation reaction is added and incubated with the beads during 15 min. at RT on a rotation
 24 wheel. Beads are washed 3 times by one volume of BW 1X buffer (10mM Tris-HCl pH 7.5; 1mM
 25 EDTA; 2M NaCl) and 2 times by one volume of TE buffer (10 mM Tris-HCl pH7.5; 1mM EDTA).
 26 They are resuspended into 10μl of NEBuffer 4 and 10μl of 10X SAM (made of 5 μl of 32 mM S-
 27 AdenosylMethionin diluted in water to a final volume of 325 μl) are added, as well as 76 μl of water

1 and 4 μ l of MmeI restriction enzyme (NEB, ref. R0637S) (final volume of 100 μ l). This reaction is
2 incubated 90 min. at 37°C under agitation. The supernatant is then removed and beads are washed 3
3 times with 50 μ l of 1X BW buffer and 2 times with one volume of TE buffer. The following is then
4 added to the beads: 5 μ l of 10X T4 ligase buffer, 2 μ l of 15 μ M (30pmol) of GEX adaptor 2 (5'
5 CAAGCAGAAGACGGCATAACGANN 3' and 3' GTTCGTCTTCTGCCGTATGCT-P 5'), 1 μ l of T4
6 DNA ligase (NEB M0202S) and 42 μ l of water (final volume 50 μ l). This reaction is incubated 2 h at
7 20°C and agitated 15 sec. each 2 minutes. Beads are then washed 3 times with 1X BW buffer and 2
8 times with one volume of TE buffer before being resuspended into 10 μ l of distilled water. 2 μ l of this
9 reaction (DNA on beads) are then mixed with 10 μ l of 5X HF Phusion Buffer and 0.5 μ l of Phusion
10 DNA Polymerase (Finnzymes, ref. F-530), 0.5 μ l of dNTP mix (25 mM each), 0.5 μ l of 25 μ M GEX
11 PCR primer 1 (5' CAAGCAGAAGACGGCATAACGA 3'), 0.5 μ l of 25 μ M GEX PCR primer 2 (5'
12 AATGATACGGCGACCACCGACAGGTTTCAGAGTTCTACAGTCCGA 3') and 36 μ l of water
13 (final volume of 50 μ l). This reaction is amplified in a thermocycler as follows: 30 sec. at 98°C,
14 followed by 15 cycles of [10 sec. at 98°C / 30 sec. at 60°C / 15 sec. at 72°C] and 10 min. at 72°C. The
15 PCR reaction is then run on a 6% 1X TBE acrylamide gel (NOVEX, Invitrogen) (Supplemental Fig.
16 S1C) and the main DNA band (expected size 95-97 bp) is cut and purified (Spin-X-filter column from
17 Sigma, and ethanol precipitation) before being resuspended into 10 μ l of water. DNA concentration of
18 the HRS-seq library is checked with an Agilent Bioanalyzer apparatus before being used for high-
19 throughput sequencing (50 nucleotide single reads) on a HiSeq 2000 apparatus (Illumina). The
20 following primer was used to generate the clusters: 5'
21 CCACCGACAGGTTTCAGAGTTCTACAGTCCGAC 3'. Five control sequencing libraries were also
22 constructed exactly as described above by using 150 ng of mouse genomic DNA cut by StyI enzyme
23 (gDNA libraries) and 3 of them were used for high-throughput sequencing (Supplemental Fig. S1B).

24

25 **Raw data filtering**

26 Sequencing tags were trimmed and aligned on the mouse reference genome of e14Tg2a mouse ESC
27 (129P2 built from mm9 assembly) and read positions were determined. Reads mapping to multiple

1 positions and reads with more than 2 mismatches were removed. Two bioinformatic filters were then
2 applied to exclude potentially aberrant reads. Indeed, according to our protocol, all relevant reads
3 should have a size of 18 to 20 nt due to MmeI digestion (filter 1) and they should have one of their
4 extremities next to a StyI site (filter 2). Table 1 is summarizing the number of reads obtained at each
5 step of data filtering. Note that tag alignments to the mm10 assembly would not be expected to
6 improve raw data processing since 97% of StyI sites are identical between the two assemblies and that
7 the missing 3% corresponds to additional sites that are essentially lying in telomeric regions of
8 chromosomes and not in gene rich regions where most HRS are located.

9 For each fraction independently, we then calculated the total number of reads obtained for
10 each StyI restriction fragments in the mouse genome (mm9 assembly) by including reads sequenced
11 from both the 5' and 3' extremities. These processed data were then checked for technical
12 reproducibility (Supplemental Fig. S2 and Supplemental Table S1).

13

14 **Statistical analyses**

15 The processed data are discrete, consisting for each StyI fragment in read counts for three different
16 biological replicates. The aim is to compare, for each experiment, the number of reads between the
17 HRS-containing fraction and the Loop fraction or between the HRS-containing fraction and the gDNA
18 control. Statistical significance of the overrepresentation of read counts for StyI fragments in the HRS-
19 containing fraction compared to the Loop fraction (or gDNA control) has been assessed using the *R*
20 packages *DESeq* (Anders and Huber 2010) and *edgeR* (Robinson et al. 2010) (see *Supplemental*
21 *Methods*). Only fragments being identified as differential between compared conditions (*i.e.* having a
22 Benjamini-Hochberg corrected *p*-value lower than 5%) for both tests have been kept for further
23 bioinformatic analyses (*i.e.* 61,080 HRS for ESC) (Supplemental Table S2).

24

25 **Bioinformatic analyses**

26 Mean inter-chromosomal contact scores were calculated from the Hi-C data obtained on mouse ESC
27 (SRR400251 to SRR400255, replicate 2 from (Dixon et al. 2012)). The analysis of 3D proximity of

1 HRS (Fig. 2C) was based on the comparison between mean contact scores of HRS and randomizations
2 (see *Supplemental Methods*).

3 The overlap score used in Figures 3/4 and Supplemental Figure S6 is the fraction of the genomic
4 feature of interest that is covered by HRS (or random sets with the same number of elements) *i.e.* the
5 base-pair number of HRS regions corresponding to the genomic feature of interest divided by the total
6 base-pair number of the genomic feature in the mouse genome. The null model used to generate the
7 null hypothesis distribution was based on a random swapping procedure (see *Supplemental Methods*).
8 For HRS distribution with respect to A/B compartments (Fig. 3B/C), all HRS were uniformly
9 randomized on the whole genome. To test the significance of the overlap between HRS and
10 H3K36me3 (ENCF001KDY), exons or introns (UCSC mm9 assembly), the HRS present in gene
11 bodies were randomized only to gene body sequences. The distributions corresponding to 1000
12 random realizations were represented by their mean and the 95% confidence interval around this
13 mean. A/B compartments were computed as described in Lieberman-Aiden *et al.* (Lieberman-Aiden *et al.*
14 2009) using Hi-C datasets from mouse ESC (Dixon *et al.* 2012) (see *Supplemental Methods*). All
15 tracks were plotted with the WashU epigenome browser (mm9 assembly). Lamina Associated
16 Domains (LAD) analyses were performed on data available from DamID maps of lamin B1 in mouse
17 ESC (NimbleGen microarray probes) (Peric-Hupkes *et al.* 2010).

18 Gene content analyses (Fig. 4B, Supplemental Fig. S4) were performed using UCSC annotation
19 data (reFlat.txt file, mm9 built). RNA-seq data from ESC (used in Fig. 4C) were downloaded from the
20 Gene Expression Omnibus repository [GSE47948] (Wamstad *et al.* 2012) and GRO-seq data from
21 ESC (used in Fig. 4D) from [GSE27037] (Min *et al.* 2011). ChIP-seq data used in Figure 5B and
22 Supplemental Fig. S6C were downloaded from the ENCODE project [ENCF001KEV ;
23 ENCF001KFB ; ENCF001KFH ; ENCF001KFN ; ENCF001KFT ; ENCF001ZHE ;
24 ENCF001ZID, <http://genome.cse.ucsc.edu/encode>]. Super-enhancer data (Fig. 4E) were downloaded
25 from the dbSUPER database (<http://asntech.org/dbsuper/>) (Khan and Zhang 2016). Gene Ontology
26 analyses (Fig. 6 and Supplemental Fig. S8A) were performed using the Functional Annotation Tool on
27 the DAVID 6.8 ontology server (settings: GOTERM_BP_DIRECT, KEGG_PATHWAY, Fold

1 Enrichment and Benjamini-Hochberg-corrected Fisher's exact test; all other settings were defaults)
2 (<https://david.ncifcrf.gov/home.jsp>) (Huang da et al. 2009). Comparative analysis of predicted
3 transcription factor binding (Supplemental Fig. S8B) were performed using the *i-cisTarget* tool
4 (<https://gbiomed.kuleuven.be/apps/lcb/i-cisTarget/>) (Herrmann et al. 2012). All settings were defaults
5 with a Normalized Enrichment Score (NES) threshold of 0.3 corresponding to a *p*-value < 0.01. For
6 repeat analyses (Fig. 8C), we first identified repeat sequences mapping within ESC HRS and the
7 number of repeats was then determined for each of the 1554 repeat families (UCSC, mm9 built).
8 Several bioinformatic analyses performed on ESC HRS were assessed statistically by randomization
9 tests, with n=100 or n=1000 uniformly random resampling (see *Supplemental Methods*).

10

11 **Immunofluorescence microscopy**

12 Immunofluorescence microscopy on nuclear halos was performed on silanized cover slip, using
13 antibodies targeting p80 coilin (polyclonal rabbit antibody, 1:400 dilution, gift from R. Bordonné)
14 (Boulisfane et al. 2011), SMN (monoclonal antibody from BD Transduction Laboratories #610646,
15 1:1000 dilution), PML (mouse monoclonal antibody, clone 36.1-104, Millipore #MAB3738, 1:500
16 dilution, gift from P. Lomonte) (see *Supplemental Methods* for further details).

17

18 **Data access**

19 Raw data and processed data from this study have been submitted to the Gene Expression Omnibus
20 repository under accession number GSE106751 (<https://www.ncbi.nlm.nih.gov/geo>). Fully processed
21 data supporting the findings of this study are available within Supplemental Material files.

22

23 **Acknowledgments**

24 We thank Isabelle Degors, Sébastien This, Emeric Dubois, Françoise Carbonell, Marie-Noëlle Lelay-
25 Taha and Marjorie Drac (DNA combing platform) for technical assistance. We thank Florence Rage,
26 Rémy Bordonné and Patrick Lomonte for providing antibodies, Cyril Esnault, Jean-Christophe Andrau
27 for help with bioinformatics, Eric Soler and Mounia Lagha for critical reading of the manuscript. This
28 work was supported by grants from the *Institut National du Cancer* [PLBIO 2012-129, INCa_5960 to

1 T.F.], the *AFM-Téléthon* [N°21024 to T.F.], the *Ligue contre le cancer*, the *Agence Nationale de la*
 2 *Recherche* [CHRODYT, ANR-16-CE15-0018-04] and the *C.N.R.S.* M.O.B. was supported by the
 3 University of Montpellier and Z.Y. by a fellowship from the IDEX Super 3DRNA.

5 **Author contributions**

6 MOB performed HRS assays, sequencing libraries, raw data filtering, bioinformatic analyses and
 7 edited the manuscript. AC performed bioinformatic analyses. FC designed the project, performed raw
 8 data filtering and bioinformatic analyses. MS conceived and optimized sequencing library protocol.
 9 HP performed NGS sequencings. ChR and RS analysed genomic data and performed statistical
 10 analyses. TB provided the biological material and performed immunofluorescence microscopy (IF).
 11 ZY performed bioinformatic analyses. MT and SS performed HRS-qPCR. CoR performed HRS-qPCR
 12 and IF. GC designed the project and edited the manuscript. AL designed/assessed statistical analyses
 13 and edited the manuscript. JM performed bioinformatic analyses and edited the manuscript. LJ
 14 designed the project, managed NGS sequencing and edited the manuscript. TF conceived and designed
 15 the project, analysed genomic data, performed bioinformatic analyses and wrote the manuscript.

17 **Disclosure declaration**

18 The authors declare no conflicts of interest.

20 **References**

- 21 Amendola M, van Steensel B. 2015. Nuclear lamins are not required for lamina-associated domain
 22 organization in mouse embryonic stem cells. *EMBO Rep* **16**: 610-617.
- 23 Anders S, Huber W. 2010. Differential analysis for sequence count data. *Genome Biol* **11**: R106.
- 24 Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot JP,
 25 Tanay A et al. 2017. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*
 26 **171**: 557-572.e524. doi: 510.1016/j.cell.2017.1009.1043.
- 27 Boulisfane N, Choleza M, Rage F, Neel H, Soret J, Bordonne R. 2011. Impaired minor tri-snRNP
 28 assembly generates differential splicing defects of U12-type introns in lymphoblasts derived
 29 from a type I SMA patient. *Hum Mol Genet* **20**: 641-648.
- 30 Braem C, Recolin B, Rancourt RC, Angiolini C, Barthes P, Branchu P, Court F, Cathala G, Ferguson-
 31 Smith AC, Forne T. 2008. Genomic matrix attachment region and chromosome conformation
 32 capture quantitative real time PCR assays identify novel putative regulatory elements at the
 33 imprinted *Dlk1/Gtl2* locus. *J Biol Chem* **283**: 18612-18620.
- 34 Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin
 35 for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and
 36 nucleosome position. *Nat Methods* **10**: 1213-1218.
- 37 Carmo-Fonseca M, Ferreira J, Lamond AI. 1993. Assembly of snRNP-containing coiled bodies is
 38 regulated in interphase and mitosis--evidence that the coiled body is a kinetic nuclear
 39 structure. *J Cell Biol* **120**: 841-852.
- 40 Clemson CM, Hutchinson JN, Sara SA, Ensminger AW, Fox AH, Chess A, Lawrence JB. 2009. An
 41 architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of
 42 paraspeckles. *Mol Cell* **33**: 717-726.

- 1 Cook PR. 2010. A model for all genomes: the role of transcription factories. *J Mol Biol* **395**: 1-10. .
- 2 Cournac A, Koszul R, Mozziconacci J. 2016. The 3D folding of metazoan genomes correlates with the
3 association of similar repetitive elements. *Nucleic Acids Res* **44**: 245-255.
- 4 Court F, Baniol M, Hagege H, Petit JS, Lelay-Taha MN, Carbonell F, Weber M, Cathala G, Forné T.
5 2011. Long-range chromatin interactions at the mouse *Igf2/H19* locus reveal a novel
6 paternally expressed long non-coding RNA. *Nucleic Acids Res* **39**: 5893-5906.
- 7 Dileep V, Ay F, Sima J, Vera DL, Noble WS, Gilbert DM. 2015. Topologically associating domains and
8 their long-range contacts are established during early G1 coincident with the establishment
9 of the replication-timing program. *Genome Res* **25**: 1104-1113.
- 10 Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in
11 mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376-380.
- 12 Dobson JR, Hong D, Barutcu AR, Wu H, Imbalzano AN, Lian JB, Stein JL, van Wijnen AJ, Nickerson JA,
13 Stein GS. 2017. Identifying Nuclear Matrix-Attached DNA Across the Genome. *J Cell Physiol*
14 **232**: 1295-1305.
- 15 Ea V, Baudement MO, Lesne A, Forné T. 2015a. Contribution of Topological Domains and Loop
16 Formation to 3D Chromatin Organization. *Genes (Basel)* **6**: 734-750. .
- 17 Ea V, Sexton T, Gostan T, Herviou L, Baudement MO, Zhang Y, Berlivet S, Le Lay-Taha MN, Cathala G,
18 Lesne A et al. 2015b. Distinct polymer physics principles govern chromatin dynamics in
19 mouse and Drosophila topological domains. *BMC Genomics* **16**:607.
- 20 Engelke R, Riede J, Hegermann J, Wuerch A, Eimer S, Dengjel J, Mittler G. 2014. The quantitative
21 nuclear matrix proteome as a biochemical snapshot of nuclear organization. *J Proteome Res*
22 **13**: 3940-3956.
- 23 Fraser J, Ferrai C, Chiariello AM, Schueler M, Rito T, Laudanno G, Barbieri M, Moore BL, Kraemer DC,
24 Aitken S et al. 2015. Hierarchical folding and reorganization of chromosomes are linked to
25 transcriptional changes in cellular differentiation. *Mol Syst Biol* **11**: 852.
- 26 Frey MR, Matera AG. 1995. Coiled bodies contain U7 small nuclear RNA and associate with specific
27 DNA sequences in interphase human cells. *Proc Natl Acad Sci U S A* **92**: 5915-5919.
- 28 Gaffney DJ, McVicker G, Pai AA, Fondufe-Mittendorf YN, Lewellen N, Michelini K, Widom J, Gilad Y et
29 al. 2012. Controls of nucleosome positioning in the human genome. *PLoS Genet* **8**: e1003036.
- 30 Gaspard N, Bouschet T, Herpoel A, Naeije G, van den Aemele J, Vanderhaeghen P. 2009. Generation
31 of cortical neurons from mouse embryonic stem cells. *Nat Protoc* **4**: 1454-1463.
- 32 Gaspard N, Bouschet T, Hourez R, Dimidschstein J, Naeije G, van den Aemele J, Espuny-Camacho I,
33 Herpoel A, Passante L, Schiffmann SN et al. 2008. An intrinsic mechanism of corticogenesis
34 from embryonic stem cells. *Nature* **455**: 351-357.
- 35 Gibcus JH, Dekker J. 2013. The hierarchy of the 3D genome. *Mol Cell* **49**: 773-782.
- 36 Giresi PG, Kim J, McDaniell RM, Iyer VR, Lieb JD. 2007. FAIRE isolates active regulatory elements from
37 human chromatin. *Genome Res* **17**: 877-885.
- 38 Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, Henikoff S. 2011. Epigenome characterization at
39 single base-pair resolution. *Proc Natl Acad Sci U S A* **108**: 18318-18323.
- 40 Henikoff S, Henikoff JG, Sakai A, Loeb GB, Ahmad K. 2009. Genome-wide profiling of salt fractions
41 maps physical properties of chromatin. *Genome Res* **19**: 460-469.
- 42 Herrmann C, Van de Sande B, Potier D, Aerts S. 2012. i-cisTarget: an integrative genomics method for
43 the prediction of regulatory features and cis-regulatory modules. *Nucleic Acids Res* **40**: e114.
- 44 Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA. 2017. A Phase Separation Model for
45 Transcriptional Control. *Cell* **169**: 13-23.
- 46 Hon G, Wang W, Ren B. 2009. Discovery and annotation of functional chromatin signatures in the
47 human genome. *PLoS Comput Biol* **5**: e1000566.
- 48 Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists
49 using DAVID bioinformatics resources. *Nat Protoc* **4**: 44-57.

- 1 Hutchinson JN, Ensminger AW, Clemson CM, Lynch CR, Lawrence JB, Chess A. 2007. A screen for
2 nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing
3 domains. *BMC Genomics* **8**: 39.
- 4 Khan A, Zhang X. 2016. dbSUPER: a database of super-enhancers in mouse and human genome.
5 *Nucleic Acids Res* **44**: D164-171.
- 6 Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragozcy T, Telling A, Amit I, Lajoie BR,
7 Sabo PJ, Dorschner MO et al. 2009. Comprehensive mapping of long-range interactions
8 reveals folding principles of the human genome. *Science* **326**: 289-293.
- 9 Linnemann AK, Platts AE, Krawetz SA. 2009. Differential nuclear scaffold/matrix attachment marks
10 expressed genes. *Hum Mol Genet* **18**: 645-654.
- 11 Mao YS, Zhang B, Spector DL. 2011. Biogenesis and function of nuclear bodies. *Trends Genet* **27**: 295-306.
- 12 Milligan L, Antoine E, Bisbal C, Weber M, Brunel C, Forné T, Cathala G. 2000. H19 gene expression is
13 up-regulated exclusively by stabilization of the RNA during muscle cell differentiation.
14 *Oncogene* **19**: 5810-5816.
- 15 Milligan L, Forné T, Antoine E, Weber M, Hemonnot B, Dandolo L, Brunel C, Cathala G. 2002.
16 Turnover of primary transcripts is a major step in the regulation of mouse H19 gene
17 expression. *EMBO Rep* **3**: 774-779.
- 18 Min IM, Waterfall JJ, Core LJ, Munroe RJ, Schimenti J, Lis JT. 2011. Regulating RNA polymerase
19 pausing and transcription elongation in embryonic stem cells. *Genes Dev* **25**: 742-754.
- 20 Nemeth A, Conesa A, Santoyo-Lopez J, Medina I, Montaner D, Peterfia B, Solovei I, Cremer T, Dopazo
21 J, Langst G. 2010. Initial genomics of the human nucleolus. *PLoS Genet* **6**: e1000889.
- 22 Nizami Z, Deryusheva S, Gall JG. 2010. The Cajal body and histone locus body. *Cold Spring Harb*
23 *Perspect Biol* **2**: a000653.
- 24 Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA,
25 Bruneau BG. 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome
26 Domains from Genomic Compartmentalization. *Cell* **169**: 930-944.
- 27 Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J,
28 Sedat J et al. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation
29 centre. *Nature* **485**: 381-385.
- 30 Padeken J, Heun P. 2014. Nucleolus and nuclear periphery: velcro for heterochromatin. *Curr Opin Cell*
31 *Biol* **28**: 54-60.
- 32 Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SW, Solovei I, Brugman W, Graf S, Flicek P,
33 Kerkhoven RM, van Lohuizen M et al. 2010. Molecular maps of the reorganization of
34 genome-nuclear lamina interactions during differentiation. *Mol Cell* **38**: 603-613.
- 35 Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, Lai MM, Shishkin AA, Bhat P,
36 Takei Y et al. 2018. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in
37 the Nucleus. *Cell* **4**: 30636-30636.
- 38 Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I,
39 Omer AD, Lander ES et al. 2014. A 3D Map of the Human Genome at Kilobase Resolution
40 Reveals Principles of Chromatin Looping. *Cell* **159**: 1665-1680.
- 41 Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential
42 expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.
- 43 Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber
44 W, C HH, Mirny L et al. 2017. Two independent modes of chromatin organization revealed by
45 cohesin removal. *Nature* **551**: 51-56.
- 46 Sleeman JE, Trinkle-Mulcahy L. 2014. Nuclear bodies: new insights into assembly/dynamics and
47 disease relevance. *Curr Opin Cell Biol* **28**: 76-83.
- 48 Stadhouders R, Vidal E, Serra F, Di Stefano B, Le Dily F, Quilez J, Gomez A, Collombet S, Berenguer C,
49 Cuartero Y et al. 2018. Transcription factors orchestrate dynamic interplay between genome
50 topology and gene regulation during cell reprogramming. *Nat Genet* **15**: 017-0030.

- 1 Sutherland H, Bickmore WA. 2009. Transcription factories: gene expression in unions? *Nat Rev Genet*
 2 **10**: 457-466.
- 3 Thompson M, Haeusler RA, Good PD, Engelke DR. 2003. Nucleolar clustering of dispersed tRNA
 4 genes. *Science* **302**: 1399-1401.
- 5 Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. 2011. Determinants of nucleosome
 6 organization in primary human cells. *Nature* **474**: 516-520.
- 7 Vogel MJ, Peric-Hupkes D, van Steensel B. 2007. Detection of in vivo protein-DNA interactions using
 8 DamID in mammalian cells. *Nat Protoc* **2**: 1467-1478.
- 9 Wamstad JA, Alexander JM, Truty RM, Shrikumar A, Li F, Eilertson KE, Ding H, Wylie JN, Pico AR,
 10 Capra JA et al. 2012. Dynamic and coordinated epigenetic regulation of developmental
 11 transitions in the cardiac lineage. *Cell* **151**: 206-220.
- 12 Wang Q, Sawyer IA, Sung MH, Sturgill D, Shevtsov SP, Pegoraro G, Hakim O, Baek S, Hager GL, Dundr
 13 M. 2016a. Cajal bodies are linked to genome conformation. *Nat Commun* **7**:10966.
- 14 Wang S, Su JH, Beliveau BJ, Bintu B, Moffitt JR, Wu CT, Zhuang X. 2016b. Spatial organization of
 15 chromatin domains and compartments in single chromosomes. *Science* **353**: 598-602.
- 16 Weber M, Hagège H, Murrell A, Brunel C, Reik W, Cathala G, Forné T. 2003. Genomic imprinting
 17 controls matrix attachment regions in the Igf2 gene. *Mol Cell Biol* **23**: 8953-8959.
- 18 Wei Y, Zhang S, Shang S, Zhang B, Li S, Wang X, Wang F, Su J, Wu Q, Liu H et al. 2016. SEA: a super-
 19 enhancer archive. *Nucleic Acids Res* **44**: D172-179.
- 20 Zhou HY, Katsman Y, Dhaliwal NK, Davidson S, Macpherson NN, Sakthidevi M, Collura F, Mitchell JA.
 21 2014. A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential.
 22 *Genes Dev* **28**: 2699-2711.

23

24 **Figure legends**

25 **Figure 1: Flowchart of the HRS-seq method.** The HRS-seq method consists in high-throughput
 26 sequencing of genomic DNA issued from HRS assays. **(A)** HRS assay principle. Each HRS assay
 27 involves 10^5 nuclei that are treated with a 2M NaCl buffer to obtain the so-called nuclear halos.
 28 Nuclear halos are digested with a restriction enzyme (here StyI) and the insolubilized fraction (HRS-
 29 containing fraction) is separated from the soluble Loop fraction by ultrafiltration. Genomic DNA is
 30 purified from each fraction and controls are performed to ensure the quality of each assay. **(B)**
 31 Construction of HRS-seq libraries for deep-sequencing. For each HRS-seq experiment, two
 32 sequencing libraries are prepared: one from the HRS-containing fraction and one from the Loop
 33 fraction. A StyI adaptor containing a MmeI binding site is ligated to the StyI restriction fragments.
 34 Ligated fragments are captured on streptavidin beads and digested with MmeI to obtain StyI fragments
 35 having homogenous sizes (18 to 20 nucleotides). The beads are washed several times, a MmeI adaptor
 36 is ligated and these StyI/MmeI fragments are eluted from the beads. The StyI and MmeI adaptors are
 37 used for deep-sequencing. **(C)** Preparation of biological replicates. Each biological replicate (here
 38 Rep.1) is prepared from a different nuclei preparation (here Prep.1). A first sequencing library (here
 39 HRS Rep.1) is prepared from HRS fractions pooled from 12 HRS assays (technical replicates) and
 40 another one (Loop Rep.1) is prepared from Loop fractions pooled from the same 12 HRS assays. This
 41 procedure was applied on three distinct nuclei preparations to obtain 6 sequencing libraries
 42 representing three biological replicates.

43

1 **Figure 2: Chromosomal mapping of HRS identified in mouse ESC.** (A) HRS identified by HRS-
 2 seq performed in mouse ESC have been mapped (brown bars) on mouse chromosomes. The mean
 3 densities of HRS on each chromosome (HRS/Mb) are indicated on the figure. (B) The distance
 4 between consecutive HRS (d) was determined. The graph shows the genome-wide distribution (1kb
 5 bins) of non-null values for d corresponding to HRS (blue) and random (brown) StyI fragments. The
 6 median values of d for each distribution are indicated on the figure. The difference between the two
 7 distributions is highly significant, featuring a p -value lower than 10^{-100} (Wilcoxon rank sum test). (C)
 8 The mean inter-chromosomal contact scores of 100 kb bins enriched in HRS (red dots) were calculated
 9 from Hi-C data available for the same cell type (ESC) (Dixon et al. 2012) and compared to the mean
 10 contact scores obtained from 100 random sets of the same number of 100 kb bins (box-plots). The
 11 box-plot on the right represents the mean contact score and randomizations obtained when HRS and
 12 random StyI fragments are taken only in the A compartment while the box-plot on the left represents
 13 the mean contact score and randomizations obtained from the whole genome. Bars represent the
 14 minimum and maximum values obtained in the 100 randomizations. The number of 100 kb bins (n)
 15 used for each randomizations is indicated on the figure. The p -value indicates the significance of the
 16 difference between the mean contact scores obtained for HRS *vs* randomizations.

17

18 **Figure 3: HRS are associated with the active chromosomal compartment.** (A) Comparison
 19 between A and B compartments, StyI fragment, HRS and gene densities along mouse chromosome 1.
 20 (B) For each chromosome, the overlap score between ESC HRS and the active A compartment has
 21 been calculated (red dot) and compared to the overlap scores obtained for 1000 randomizations (box
 22 plots). The overlap score represents the fraction of the genomic feature of interest (here A
 23 compartment) that is covered by HRS. The p -value (valid independently for each chromosome)
 24 assesses the difference between the overlap scores obtained for HRS *vs* 1000 randomizations. (C)
 25 Analyses of overlap scores for the inactive B compartment were performed as described above.

26

27 **Figure 4: HRS are associated with actives genes and exonic regions.** (A) The number of HRS that
 28 overlap with CpG islands (UCSC, mm9 built) was counted (4817) (left pie-chart) and compared to the
 29 mean counts (623 ± 26 , SD) obtained from “random permutation tests” with $n=1000$ random
 30 resampling (1000 sets of equivalent number of random StyI fragments) (right pie-chart). The p -value
 31 indicates the significance of the difference between the counts obtained for HRS *vs* 1000
 32 randomizations. (B) The number of TSS that map into the ESC HRS set was counted (brown dot).
 33 This number was compared to the counts obtained from “random permutation tests” with $n=1000$
 34 random resampling (1000 sets of 61,080 random StyI fragments) (box-plots, median value obtained
 35 from randomizations is indicated in purple). The p -value indicates the significance of the difference
 36 between the counts obtained for HRS *vs* 1000 randomizations. (C) Based on RNA-seq data available

1 from ESC (Wamstad et al. 2012), mouse genes were classified into 3 sets. The first set corresponds to
 2 the 3000 genes having the highest expressed levels, the second to 3000 moderately expressed genes
 3 and the last to the 3000 genes with the weakest expression levels (mean of two replicates). For each
 4 set, the number of HRS-associated genes were counted and compared to the counts obtained for
 5 equivalent numbers of genes taken at random. The *p*-value indicates the significance of the difference
 6 between the counts obtained for HRS vs 100 randomizations (box-plots). It is valid independently for
 7 the differences observed in the highly and weakly expressed gene sets. **(D)** Identical analysis as
 8 described above in (C) was performed using ESC GRO-seq data (Min et al. 2011). **(E)** The numbers of
 9 super-enhancers (Khan and Zhang 2016) that overlap with the ESC HRS (brown dots) were counted
 10 for all super-enhancers known in the mouse genome (left panel), for those that are active in ESC
 11 (middle panel) or in the cortex (right panel). These numbers were compared to “random permutation
 12 tests” (1000 random sets of 61,080 StyI fragments) (box-plots, median value indicated in purple). The
 13 *p*-value indicates the significance of the difference between the counts obtained for HRS vs 1000
 14 randomizations.

15

16 **Figure 5: HRS are associated with active epigenetic marks.** **(A)** Browser snapshot showing the
 17 HRS density at the *Sox2* gene locus on mouse chromosome 3 as determined by HRS-seq experiments
 18 performed in ESC. Tracks displaying DNase I sensitive sites, RNA PolIII peaks as well as ChIP-seq
 19 data for the indicated epigenetic marks (ENCODE E14 ESC data) were plotted using the WashU
 20 epigenome browser. **(B)** Heat map depicting the Pearson correlation coefficients obtained between
 21 ESC HRS and sequences (10 kb bins) enriched in distinct epigenetic marks as indicated on the figure
 22 (black/red: high positive correlation coefficient; white/blue: low null/negative correlation coefficient).

23

24 **Figure 6: HRS-associated genes are housekeeping as well as cell type-specific genes.** DAVID
 25 ontology analyses were performed on genes for which at least one TSS was mapping in the ESC HRS.
 26 KEGG pathways (green), and GO terms related to Biological Processes (red), Molecular Functions
 27 (blue) and Cellular Components (yellow) are depicted by circles as a function of fold enrichments. For
 28 each indicated term, circle areas is proportional to gene counts. Only the most significant terms (*p*-
 29 value < 0.05 and Fold Enrichment > 1.80) are shown. *P*-values < $5 \cdot 10^{-6}$ are depicted by red squares
 30 (for exact values see Supplemental Table S4).

31

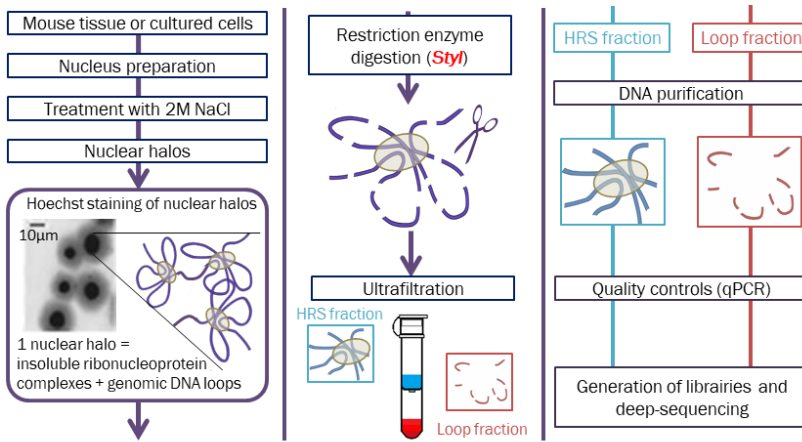
32 **Figure 7: HRS are associated with highly expressed cell-specific genes.** The enrichment levels
 33 (HRS vs Loop fractions) of StyI fragments at the *Sox2* locus (chr3:34400000-34670000, mm9
 34 assembly) were determined by qPCR (HRS-qPCR) on ESC (brown bars) **(A)** or in neurons **(B)** (blue
 35 bars). The red horizontal line corresponds to the mean local background level (value 1) and dashed
 36 lines depict the mean noise band as defined in the *Supplemental Methods*. The positions of StyI

1 fragments identified by HRS-seq, as well as RefSeq genes, StyI sites, Super-Enhancers (SE) (Khan
2 and Zhang 2016) and *Sox2 Regulatory Regions* (SRR) (Zhou et al. 2014) are indicated below the
3 histogram. The HRS-qPCR track indicates StyI fragments investigated in the experiment. Green bars
4 represent HRS (StyI fragments having enrichment levels above the noise band), red bars indicate StyI
5 fragments that are not HRS. n=3 (technical replicates) for each experiment, error bars represent s.e.m.

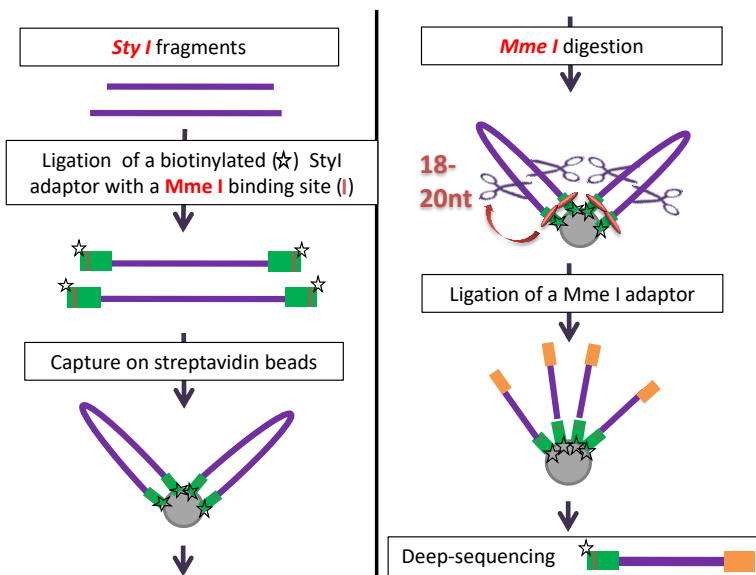
6

7 **Figure 8: HRS include nuclear body-associated sequences.** (A) Immunofluorescence (IF)
8 microscopy experiments were performed on nuclear halos using the following antibodies: α SMN
9 (Cajal Bodies and Gems) (upper panel), α Coilin (Histone Locus Bodies and Cajal bodies) (middle
10 panel), α PML (PML bodies) (bottom panel). DAPI staining is shown on the left, IF in the middle and
11 the merged picture on the right. (B) The enrichment levels (HRS vs Loop fractions) of StyI fragments
12 at the *Histone 1* locus were determined by qPCR (HRS-qPCR) on ESC (brown bars). Tracks below the
13 histogram are as described in Fig. 7. *tRNA* genes are also indicated below the histogram. n=3
14 (technical replicates), error bars represent s.e.m. (C) The enrichment (given in % of total family
15 members) in ESC HRS of 1554 repeat families were calculated and compared to enrichments obtained
16 from “random permutation tests” with n=1000 random resampling (1000 random sets of equivalent
17 numbers of StyI fragments). For each class of repeats, the percentage of repeat families that was found
18 significantly overrepresented in HRS (p -value<0.001) compared to randomizations was determined.
19 (D) HRS-qPCR experiments were performed at the *Malat1/Neat1* gene locus as indicated above (Fig.
20 8B). n=3 (technical replicates), error bars represent s.e.m. Tracks below the histogram are as described
21 above (Fig. 7 and Fig. 8B) (no *tRNA* gene is mapping to this locus).

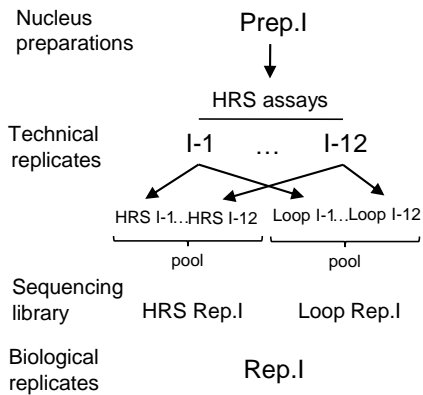
A



B



C



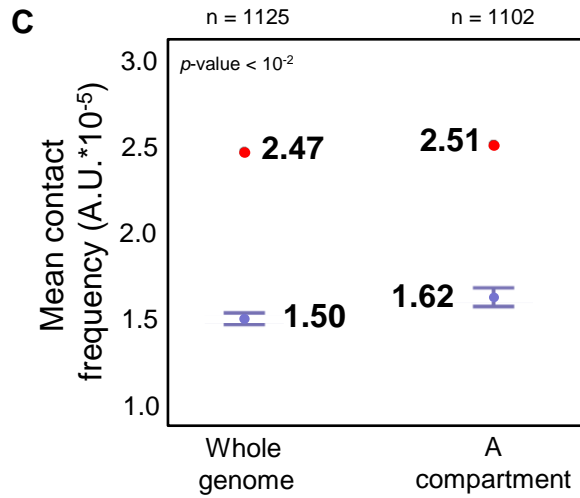
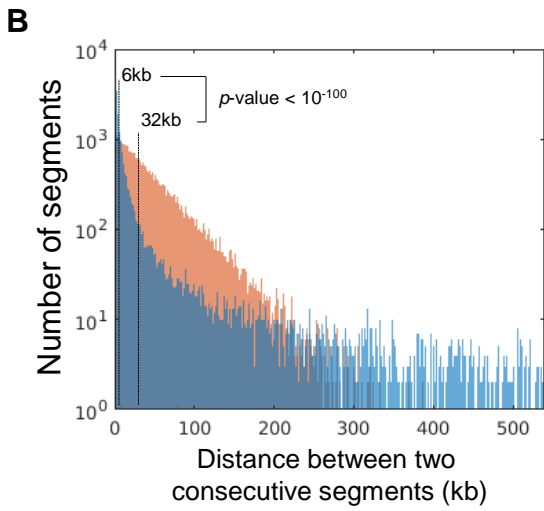
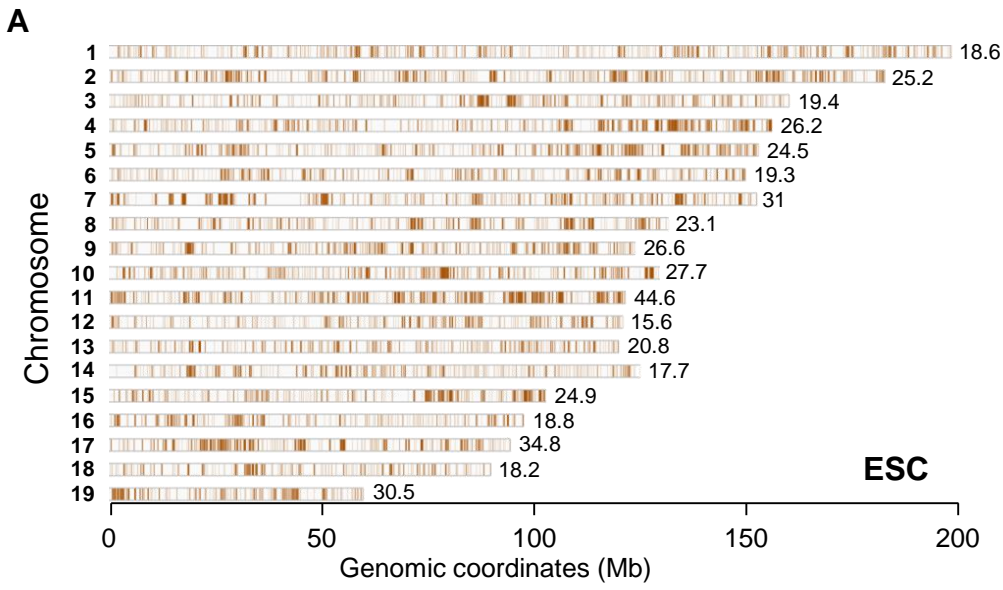
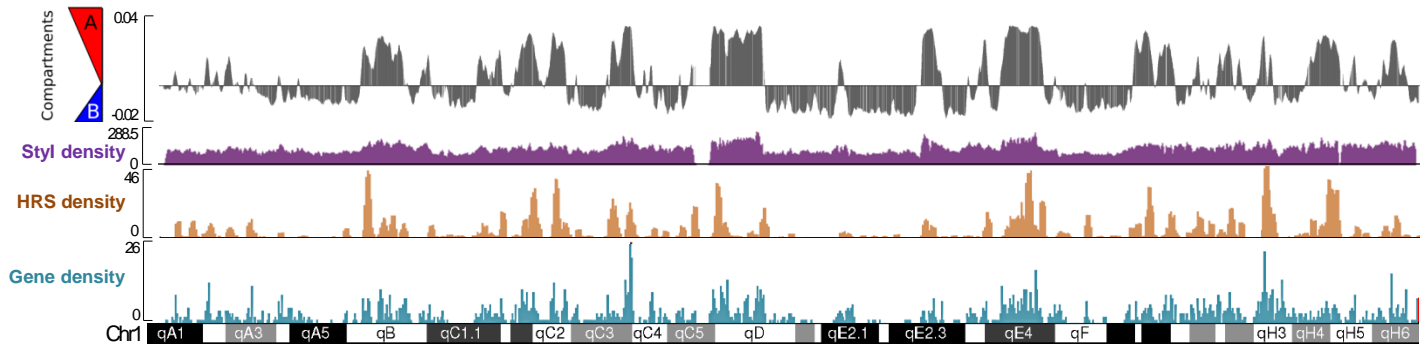
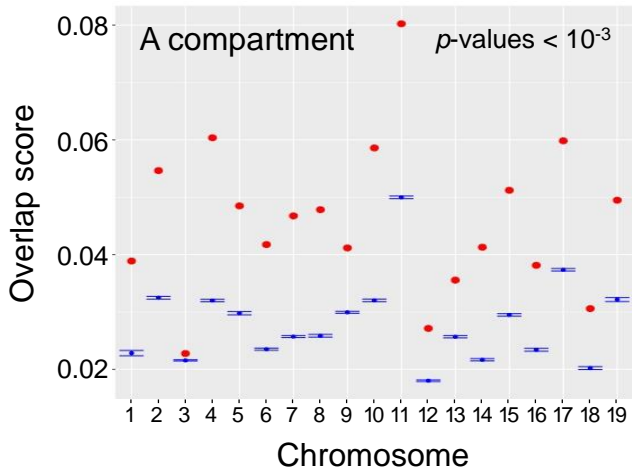


Figure 3

A



B



C

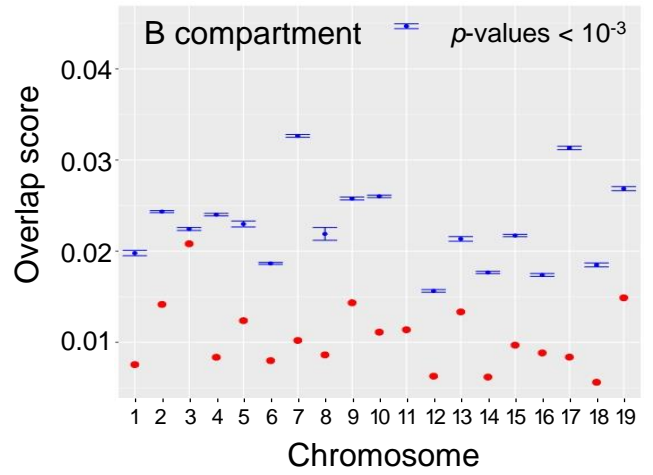


Figure 4

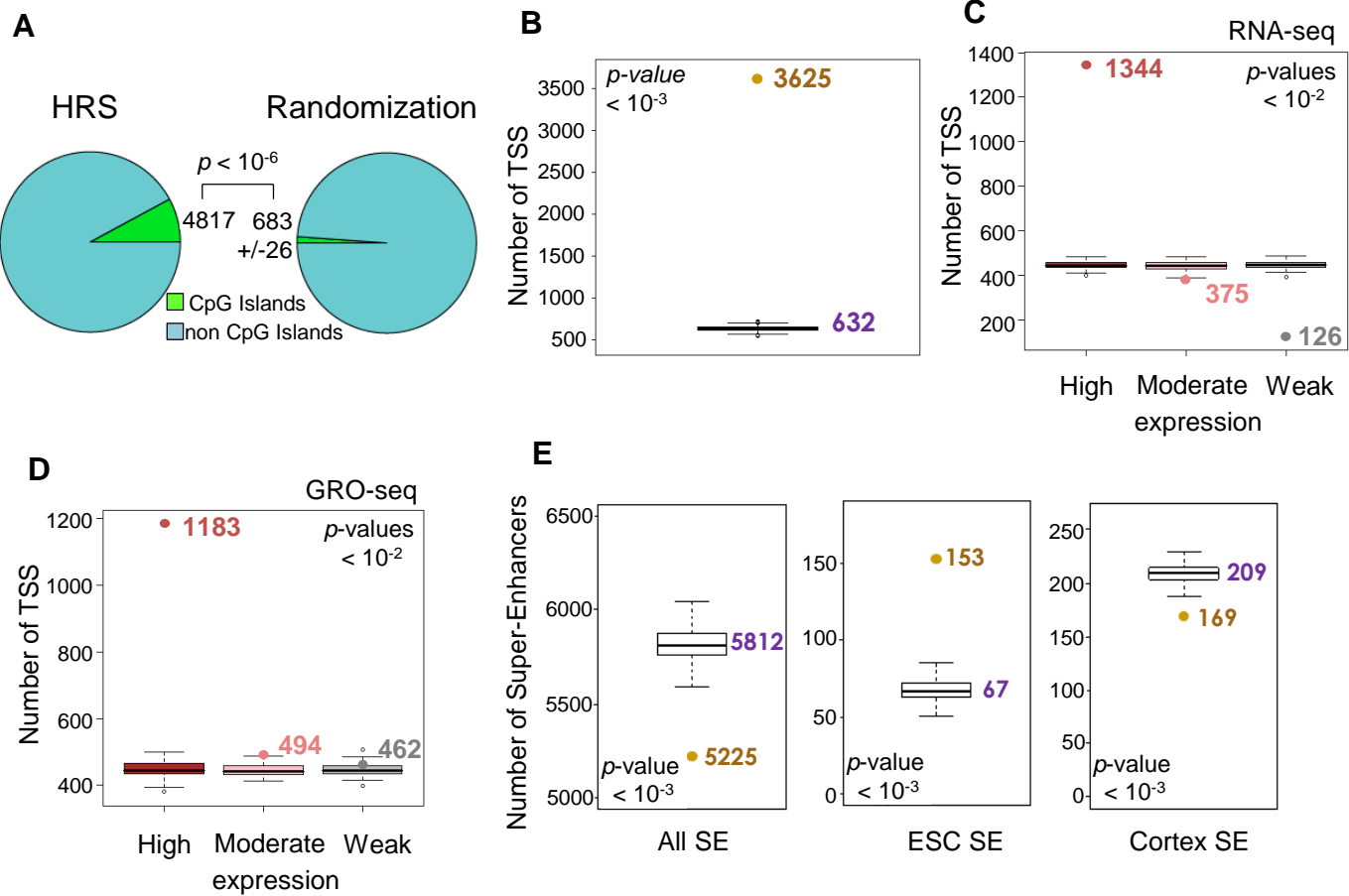


Figure 5

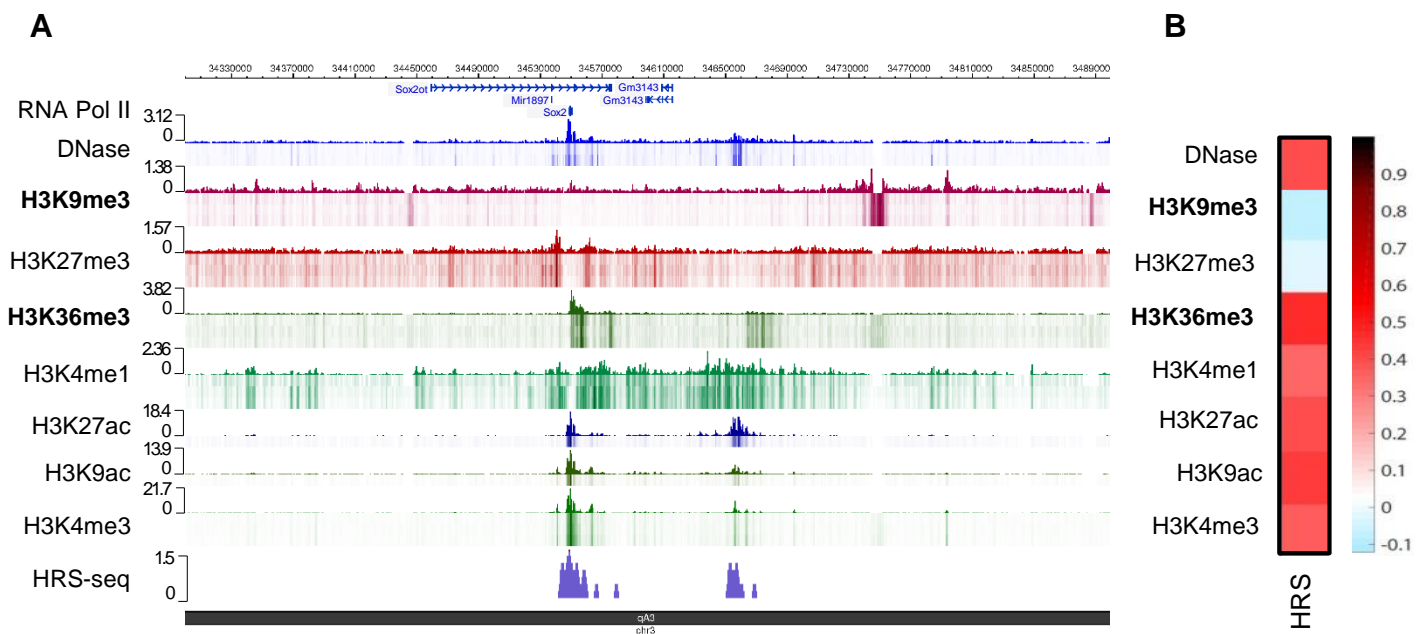


Figure 6

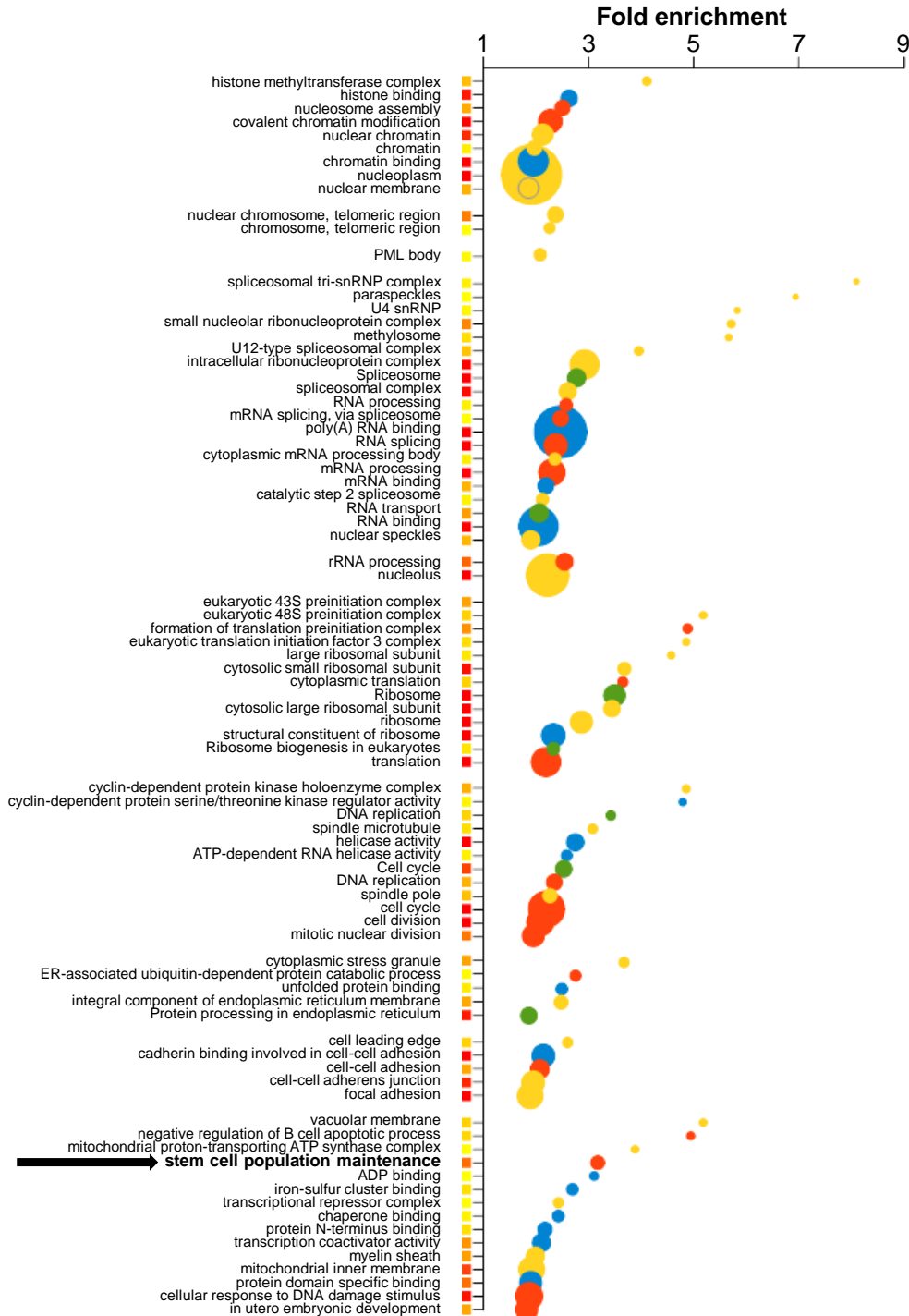
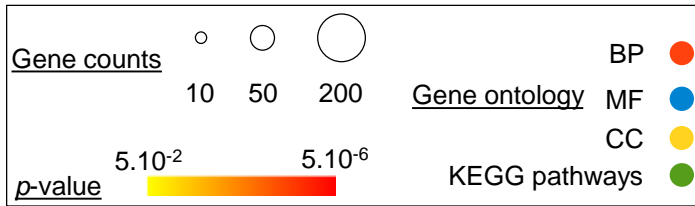
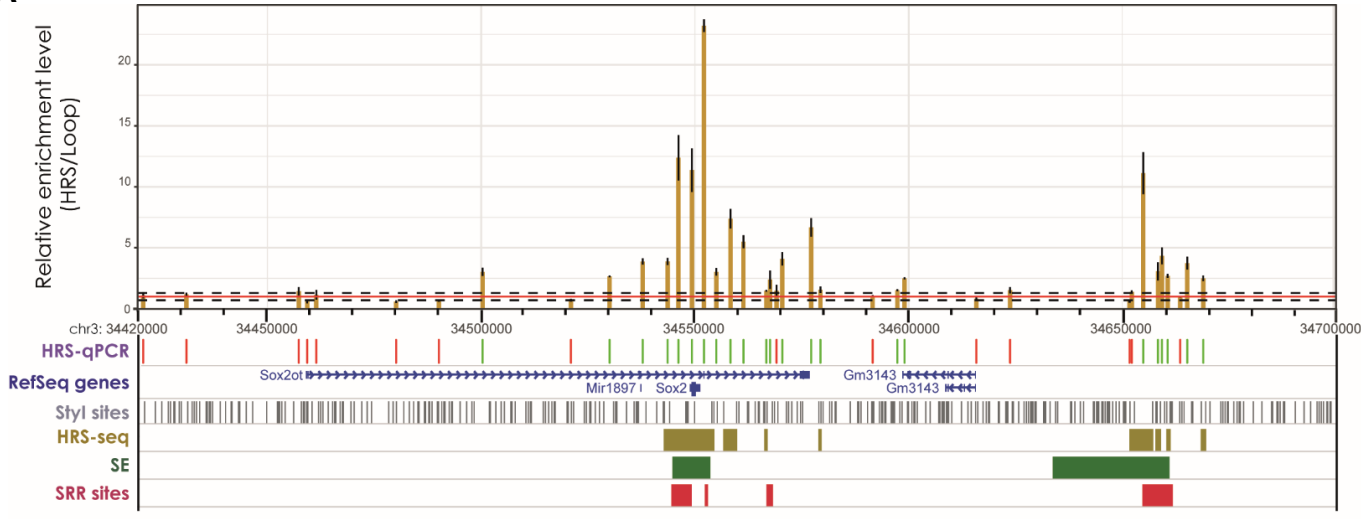


Figure 7

A



B

