Tissular chromatin states cartography based on double-barcoded DNA arrays capturing unloaded PA-Tn5 Transposase

Maria Grazia Mendoza-Ferri [1], Gwendoline Lozachmeur [1,2], Maximilien Duvina [1], Laetitia Perret [2], Didier Merciris [2], Anne Gigout [2] and Marco Antonio Mendoza-Parra [1,*]

[1] UMR 8030 Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, University of Evry-val-d'Essonne, University Paris-Saclay, 91057 Évry, France.

[2] In vivo pharmacology, Novalix, Avenue Gaston Roussel 102, Romainville, France.

[*] Corresponding author: Marco Antonio Mendoza-Parra

E-mail: mmendoza@genoscope.cns.fr

Phone: +33 (0) 1 60 76 42 92

Keywords: epigenomics, Cut&tag, chromatin states, spatial omics

Abstract

(150 words)

Recent developments in spatial omics are revoluzionating our understanding of tissue structures organization and their deregulation in disease. Here, we present a strategy for capturing chromatin histone modification signatures across tissue sections by taking advantage of a double-barcoded DNA arrays design compatible with in situ protein A-Transposase Tn5 tagmentation. This approach has been validated in presence of fresh-frozen mouse brain tissues but also in decalcified formalin-fixed paraffin-embedded (FFPE) mouse paws samples, where either the histone modification H3K4 tri-methylation or H3K27-acethylation has been used as proxy for interrogating active promoter signatures. Furthermore, since combinatorial enrichment of multiple histone modifications were shown to code for various states of gene transcriptional status (active, bivalent, repressed), we have integrated several histone modifications issued from consecutive mouse embryos to reveal changes in chromatin states across the tissue. Overall, this spatial epigenomics technology combined with the use of a spatial chromatin states analytical strategy paves the way for future epigenetics studies for addressing tissue architecture complexity.

Main text (5000 words)

Introduction

Developments in spatially-resolved transcriptomics (SrT), and more recently on spatially-resolved epigenomics (Deng et al. 2022) are providing means to interrogate organ/tissue architecture from the angle of the gene programs defining their molecular complexity. Despite the proved performance of the available spatially-resolved omics technologies in a variety of tissue samples; there are still major challenges to address. Indeed, beyond the over-discussed aspects related to the resolution of these technologies, their capacity to access to other type of molecular readouts (e.g. chromatin accessibility, epigenetics), their technical limitations issued from the way in which the biological specimens were preserved (FFPE vs fresh frozen tissues), but also their current elevated costs, make of these technologies applicable to few tissue sections, thus avoiding to generate multi-omic integrative views of their molecular complexity.

Previously we have presented a double-barcoded DNA array strategy for spatially-resolved transcriptomics (SrT), allowing to interrogate large number of tissue sections at a cost-efficient manner, as illustrated by the setup of a three-dimensional molecular cartography of human cerebral organoids (Lozachmeur et al. 2023). Herein we present the use of double-barcoded DNA arrays to perform spatial epigenomics landscaping. Specifically, this implies (i) a modified version of double-barcoded DNA arrays for capturing cleaved chromatin with the CUT&Tag chemistry (Kaya-Okur et al. 2019); (ii) a modified bioinformatics pipeline allowing to demultiplex spatial read-counts retrieved around gene promoter regions and their visualization within our tool MULTILAYER (Moehlin et al. 2021b); (iii) the implementation of combinatorial epigenetic enrichment analysis issued from consecutive sections, allowing to reveal chromatin states associated to gene promoters.

Our double-barcoded spatial epigenomics strategy has been validated in fresh-frozen tissues, but also in decalcified FFPE conserved material (bone-related tissues), known to be

incompatible with spatial transcriptomics assays because of RNA degradation. Finally, this technology has been used for performing spatial epigenomics profiling in mouse embryo samples, targeting several histone modifications assessed in consecutive sections. This strategy allows to reveal relevant chromatin state signatures associated to either active, repressive or bivalent chromatin status across the tissue, thus providing a detailed molecular view of the chromatin regulation on the grounds of its physical location.

Overall, this study covers both, the development of a cost-efficient strategy for generating spatial epigenomic maps on tissue sections, and a bioinformatics strategy for a pioneer integrative effort in spatial chromatin states profiling, essential for the field of “spatial omics”, anticipating the needs for this type of efforts with the democratization of these technologies.

Results

A double-barcoded spatial epigenomics strategy allowing to capture histone modification signatures across large tissue sections.

With the aim of capturing histone modifications across tissue sections, we have engineered a DNA array manufacturing strategy based on the use of two complementary oligonucleotides printed at a high accuracy with the help of the pico-litter spotter Sciflexarrayer S3 (Scienion). Like in our previous design for double-barcoded spatial transcriptomics (Lozachmeur et al. 2023), we have first deposited as rows, oligonucleotides harboring an amino C6 linker modification at the 5’ end for UV crosslinking, a T7promoter sequence, a unique molecular barcode per deposited row and a 30-nt-length adapter sequence, herein labelled as “Gibson” (**Figure 1A**). Then, we have printed per columns and on top of the previous deposited probes, the second set of oligonucleotides harboring a complementary Gibson sequence at

the 3'-end, a unique molecular barcode per deposited column and a specific 19-nt transposase recognition sequence (Mosaic End, herein labelled as "MOS").

Printed DNA arrays are first UV irradiated to crossking the DNA probes, then they are covered with a solution containing T4 DNA polymerase for elongating the hybridized probes, leading to a double-stranded DNA molecule harboring a unique combination of two molecular barcodes, as well as a MOS sequence at the exposed extremity of the DNA probe (**Figure 1B**). This strategy allows to manufacture DNA arrays composed by several thousands of DNA probes at a cost-effective manner since the molecular barcodes complexity is obtained by a combinatorial manner.

Tissue sections deposited on top of the double-barcoded DNA arrays are permeabilized and incubated with an antibody against the histone modification of interest, followed by a secondary antibody binding allowing to enhance the tethering of the Protein A-Transposase Tn5 (PA-Tn5) (**Figure 1C**). In contrast to the previously described spatial CUT&Tag strategy (Deng et al. 2022) , herein we use a PA-Tn5 moeity devoid of loaded DNA sequences, such that the loading takes place with the MOS sequences retrieved in the exposed ends of the printed DNA probes (**Figure 1D**). In that manner, the cleaved chromatin in presence of Magnesium chloride remains captured by the proximal DNA probes (**Figure 1E**). After tissue degradation by Proteinase K treatment, the DNA array is exposed to alkaline conditions (NaOH), revealing single-DNA ends for library preparation *in situ*. This is performed by extending DNA extremities by poly (C) tailing with Terminal Transferase, hybridization of a poly (G) primer presenting a known adapter sequence at its 5'-extremity (Adpt1), followed by Klenow Polymerase extension (**Figure1E**). Finally, DNA arrays are exposed again to alkaline conditions for releasing the elongated complementary DNA sequence, and collected in a tube for performing two rounds of PCR amplification. The first PCR amplification step relies in a primer complementary to the T7-promoter sequence and presenting a known adapter fragment (Adpt2) as well as in a primer targeting the Adpt1. The second PCR amplification

uses the Adpt1 and Adpt2 sequences for introducing the P5 and P7 illumina adapters, including a molecular barcode sequence, specific to the prepared library (**Figure 1E**).

This strategy has been validated in an adult mouse brain fresh-frozen tissue section deposited on top of a DNA array covering a surface of 16×8 mm, presenting a total of 4,096 unique combinations of dual barcodes (**Figure 2A**). The manufactured DNA array presented a pitch distance of ~177 microns, with a DNA probe resolution of 100 microns (diameter spot) (**Figure 2B**). Tissue section has been immunolabelled with an antibody targeting the histone modification H3K4me3, known to be enriched in actively transcribing or bivalent promoter regions. For enhancing the tethering of the PA-Tn5, a secondary antibody has been added (1/100 dilution), complemented by the same secondary antibody conjugated to a fluorophore (AF-488) but loaded at a higher dilution level (1/1,000). In that manner, the performance of the histone modification labelling can be visualized in the same sample that will be used for the spatial epigenomics profiling (**Figure 2C**).

Once the immunolabelled mouse brain section was scanned for revealing the histone modification localization (FITC channel; **Figure 2C**) as well as for capturing its physical position within the DNA array thanks to Cy3-labelled fiducial oligonucleotides deposited in the borders of the double-barcoded deposited probes (TRITC channel; **Figure 2A**), the aforementioned procedure has been applied, leading to a sequencing library presenting DNA fragments ranging between 300 and 1,000 bp-length (**Figure 2D**). More than 250 million reads were Illumina sequenced and a custom-made bioinformatics pipeline has been applied, targeting the detection of the Gibson sequence, the presence of both unique molecular barcodes and the mapping to the mouse genome of the captured genomic sequence (**Supplemental Figure S1A** & **Figure 2E**). Despite the major loss of sequenced reads during this primary data processing, more than 40 million reads mapped to the mouse genome, and were processed as bulk with the peak caller MACS2 (Zhang et al. 2008), followed by the identification of promoter regions presenting confident peaks. Demultiplexing of these confident regions gave rise to ~2,000 read counts per SpExel (spatial Epigenomics

element, in analogy to pixel for picture element), and ~1,500 promoters per SpExel in average (**Figure 2F**).

Spatial epigenomic maps issued from confident peaks in promoters were first quantile normalized for correcting for potential technical variations across printed probes, then adjusted for cell density variation across the tissue by integrating immunostaining pixel information with the assesed read count levels (**Supplemental Figure S1B**). The obtained mouse brain spatial H3K4me3 epigenomic map revealed adjusted read counts fluctuating between 11 to 18 (log_2 scale), in which the *corpus callosum* structure is visible due to its high adjusted read counts intensity (**Figure 2G**). Promoters enriched for the H3K4me3 histone modification across the spatial map were compared with those described in two public bulk epigenomic maps assessed on mouse brain samples, revealing 8,097 common genes, corresponding to > 60% of all retreived peaks in either conditions (bulk or spatial setting) (**Figure 2H**). Importantly, a peak-centered heatmap display of these common H3K4me3 peaks revealed preferentially the presence of two-peak shape signature around the TSS, which has been further stratified in 9 peak classes based on their intensity, but also in their surrounding signatures (**Figure 2I**). Indeed, Beyond than having a double-peak TSS-centered signature, H3K4me3 enrichment around the TSS can further expand in the direction of the active transcription, correlated with an increased transcription elongation (Chen et al. 2015). This has been further illustrated in the case of the gene *Adcy5*; known to encode for a member of the membrane-bound adenylyl cyclase enzymes and being expressed in the brain. As illustrated in **Figure 2K**, the H3K4me3 enrichment around the promoter region of *Adcy5* analysed in the aforementioned ChIP-seq datasets, reveals a broader signature in one of the datasets (GSM1656749) and notably skewed towards the gene transcription orientation. Similarly the pseudo-bulk profiles obtained either from a public H3K4me3 spatial data (GSM5622964) or the one generated in this study, present H3K4me3 enrichment at the level of the TSS of *Adcy5*, but also further enrichment signatures within the gene transcription orientation (**Figure 2K**).

By comparing the pseudo-bulk profile obtained from a public H3K4me3 spatial data (GSM5622964) with the one obtained in this study revealed between 5,000 to 8,000 common peaks (depending on the enrichment confidence threshold in use), corresponding to >40 % of the peaks retrieved in the public data, despite the few number of read counts used for the analysis (~40 Million reads after mapping) (**Figure 2J**).

Finally, the spatial H3K4me3 enrichment signature associated to the gene *Adcy5* has been compared with either the one generated by Deng et al. (GSM5622964 (Deng et al. 2022)); as well as with data provided by the Allen mouse brain atlas. Indeed, *Adcy5* has been previously shown to be over-expressed in the mouse brain basal ganglia (**Figure 2O**), concordant with the spatial signature revealed in the obtained H3K4me3 map (**Figure 2L&2M)**, as well as in the H3K4me3 spatial map published by Deng et al; despite the fact that this last published data covers a rather small part of the left lobe of the mouse brain section (**Figure 2N**).

Similarly H3K4me3 enrichment in the promoter region associated to the gene *Thap2* (THAP Domain Containing), *G6pc3* (Glucose 6 Phosphatase Catalytic Subunit 3), or the transcription factore *Sox10* presented spatial signatures mainly concentrated around the *corpus callosum* structure in aggreement with the data provided in the H3K4me3 spatial map published by Zhang et al (Zhang et al. 2023) (**Supplemental Figures S2A&S2B**). Furthermore, pseudo-bulk profiles obtained either from the public H3K4me3 spatial data (GSM5622964) or the one generated in this study, revealed similar enrichment profiles around the promoter regions of the aforementioned genes; in coherence with bulk H3K4me3 profiles obtained from two public mouse brain datasets (**Supplemental Figure S2C**). In the particular case of *Sox10*, our H3K4me3 spatial enrichment profile appeared extremely sparse, potentially due to the low sequencing coverage in use, but it could also be explained by a low gene expression behavior, as revealed by *in situ* hybridization data issued from the Allen mouse brain atlas (**Supplementary Figure S2E**).

Overall, we have demonstrated herein that despite a low coverage used for the spatial profile reconstruction, our spatial epigenomics profiling allowed to capture relevant H3K4me3 enrichment signatures across a mouse brain section, in aggreement with previous studies.

Major molecular players involded in rheumatoid arthritis ca be detected by spatial epigenomics profiling performed in decalcified FFPE mouse paws samples

One of the major limitations of the available spatial transcriptomics technologies is their capacity to interrogate biological material in which the storage condition is not RNA-friendly. Specifically this corresponds to formalin-fixed, paraffin-embedded (FFPE) tissue samples, classically used for clinical sample preservation. While current spatial transcriptomics methods interrogates this type of samples by targeting the remaining RNA by hybridization, this strategy reaches its limits in bone-related samples where an acid-based decalcification procedure - prior FFPE embedding – further destroys the remaining RNA (Duncan et al. 2019).

Considering that acid decalcified FFPE embedded samples are classically used for immunostaining assays after deparafination and rehydration, we thought that our spatial epigenomics strategy should be able to capture histone modification enriched chromatin. To address the performance of our spatial epigenomics methodology in bone-related FFPE samples, we utilized samples issued from a collagen-induced arthritis (CIA) mouse model (Brand et al. 2007). Specifically, CIA mouse paws were fixed with paraformaldehyde, decalcified in presence of formic acid and FFPE embedded (**Figure 3A**). FFPE blocks were cut to obtain thin sections that were deposited on top of our manufactured DNA arrays harboring mosaic sequences. DNA arrays were deparaffined and rehydrated by following standard procedures, followed by spatial epigenomics processing (**Figure 3A**).

Sagittal sectioning of the mouse paws allows to evaluate the ankle and tarsal joints, visualized by the presence of a sagittal plane of the talus, navicular, cuneiform and meta-tarsus bones (**Figure 3B**). To capture potential actively transcribing genes we have

performed a spatial epigenomics profiling targeting the histone modification H3K27 acethylation (H3K27ac) over a DNA array composed of 2,048 unique double-barcoded DNA probes (**Supplemental Figure S3** & **Figure 3C**). Illumina sequencing of the obtained spatial epigenomics library provided > 330 million reads, from wich ~77% presented the Gibson sequence and ~40% presented in addition the dual barcodes (**Figure 3D**). Mapping to the mouse reference genome, allowed to identify ~78 million reads, which were processed as bulk for identifying confident peaks within promoter regions. Demultiplexing of these confident regions gave rise to ~2,000 read counts per SpExel and ~1,800 promoters per SpExel in average (**Figure 3E**).

H3K27ac spatial epigenomic map issued from confident peaks in promoters was quantile normalized and adjusted for variations in tissue density, leading to a digital map representing the ajusted read counts per physical position across the tissue (**Supplemental Figure S3 & Figure 3F**). A t-SNE analysis stratified the corresponding SpExels in 6 clusters, which were then projected on the spatial tissue map (**Figure 3G&H**). Cell type gene marker association analysis allowed to identify major relevant terms corresponding to bone composition (Mesenchymal progenitors: clusters 1&6; Osteocytes: clusters 1,2&3; Osteoblasts: clusters 2&6), cartilage (cartilage progenitor cell: cluster 1&6; mesenchymal progenitors: cluster 6), as well as synovial joint (Mesenchymal stem cells: Synovial fluid: cluster 4) (**Figure 3I**). Furthermore, cell type gene markers associated to an inflammatory response were detected, including the presence of regulatory T cell signatures (FOXP3+ Natural Regulatory T(Treg) cells, Monocytes: cluster 4&5; large granular lymphocytes: cluster 6; Regulatory T cells: cluster 1,4,5&6); as well as markers associated to osteoarthritic cartilage tissue (Mesenchymal progenitors: cluster 4) (**Figure3I**).

To further support this analysis, we have processed a control mouse paw section, devoid of CIA immunization treatment. As illustrated in **Supplemental Figure S4**, the control mouse section deposited on top of a 32×64 (2,048 printed spots) DNA array displayed a thinner shape than the CIA specien, in aggreement with the consequences of the inflammation effect

induced by the immunization. The control paw section has been immunostained with an anti-H3K27ac antibody and exposed to DAPI for nuclei cell density assessment (**Supplemental Figure S4A**). By following the aforementioned double-barcoded spatial epigenomics strategy, an Illumina DNA library has been obtained, and sequenced for a total of > 450 million reads (**Supplemental Figure S4C&D**). The primary bioinformatics analysis allowed to obtain >120 million reads mapped to the mouse genome, which were used for revealing a spatial digital map by applying a quantile normalization procedure and a tissue patterns density processing (**Supplemental Figure S4B-E**). Importantly, a direct comparison between the control sample and the CIA immunized specimen revealed 2,599 promoters specific to the control section; 3,290 common H3K27ac-enriched promoters, and twice as much promoter regions specifically enriched in the CIA sample (6,874 gene promoters; **Supplemental Figure S4G**). Furthermore, a tissue-cell type gene marker association analysis applied to each of these populations confirmed a specific enrichment for the osteoarthritic cartilage tissue (Mesenchymal progenitors), or the presence of macrophage sysnovium signatures in the CIA-specific promoters (**Supplemental Figure S4H**).

Among the various CIA H3K27ac over-enriched promoters, genes coding for proteins like the transmembrane glycoprotein podoplanin (*Pdpn*) – known to participate in osteocyte function, but also as an important sensor for bone damage during inflammatory arthritis (Wehmeyer et al. 2019) – or the Fibroblast growth factor 23 (*Fgf23*) mainly produced by osteocytes and in elevated levels in rheumatoid arthritis (Hu et al. 2024), were observed across the mouse paws sections (**Figure 3J**). Similarly, promoter regions associated to genes like *Dlx5* (Distal-Less Homeobox 5) – a well known transcription factor driving osteoblastogenesis (Samee et al. 2008) , or *Gabbr1* (Gamma-Aminobutyric Acid Type B Receptor Subunit 1), a member of the GABA receptors and known to be over-expressed in rheumatoid arthritis (Song et al. 2024), were also retrieved enriched in H3K27ac signatures across the tissue.

Furthermore, promoter regions associated to genes coding for pro-inflammatory cytokines including *Il1b*, *Il6* or *Il10* – known to drive inflammation and the destructive process of the

tissue (Magyari et al. 2014) – the cytokines *Cd44*, *Cd166*, *Cd248*, the colony-stimulating factor (*Csf1*) - driving the development of the monocyte–macrophage cell lineage, which play an important function in the pathogenesis of rheumatoid arthritis, or the Integrin Subunit Alpha 2 (*Itga2*) – shown to be over-expressed in RA and influencing T cell growth accompanied by an upregulation of pro-inflammatory cytokine expression (He et al. 2021) - were also enriched for H3K27ac across the studied CIA mouse paws (**Figure 3J**). To further validate the H3K27ac enrichment in the aforementioned promoters, a pseudo-bulk profile was generated from the H3K27ac CIA map, and displayed along with public data concerning ChIP-sequencing profiles for the H3K27ac but also RNA sequencing tracks (**Supplemental Figure S5**). These public data were obtained from the platform QC Genomics (Blum et al. 2020; Mendoza-Parra et al. 2016, 2013). Furthermore, the pseudo-bulk profile issued from the H3K27ac CIA maps has been stratified in 6 heatmaps corresponding to the various clusters displayed in Figure 3G; such that the contribution of the various H3K27ac displayed peaks to each of the indicated clusters can be observed (**Supplemental Figure S5**).

Finally, to further confirm the relevance of the obtained H3K27ac over-enrichment signatures in the CIA specimen, we have compared them with public temporal transcriptomics data obtained from a CIA immunization assay and performed in mice paws soft tissue material (**Figure 3K**; (Li et al. 2022)). The reanalysis of their temporal transcriptomes revealed the presence of 7 major gene co-expression paths reflecting the temporal gene expression transitions (**Figure 3L**). Importantly, the comparison of these gene co-expression paths with the H3K27ac promoter enrichment signatures retrieved in the CIA specimen revealed up to 50% of common genes with the strongly expressed paths, and only ~10% in the case of co-expression path 6 presenting a downregulated gene expression behavior, confirming a relatively strong correlation between H3K27ac enrichment at the promoter regions and gene expression response (**Figure 3M**).

Overall, the aforementioned data clearly demonstrate the capacity of the double-barcoded spatial epigenomics strategy to retrieve RA-related signatures from acid-decalcified FFPE mouse paws samples.

Spatially resolving chromatin state signatures by integrating multiple histone modification maps issued from consecutive mouse embryo sections.

To further illustrate the performance of the presented technology, we have performed spatial epigenomics profiling over mouse embryo sections (E11.5). We have manufactured DNA arrays composed by 32×32×2 unique double-barcoded probes (32×32 interstitially printed arrays; 2,048 spots). Mouse embryo sections deposited on top of the DNA arrays were immunolabelled for the histone modification H3K27ac, known to be enriched in active promoters as well as in active enhancer regions. DAPI staining has been used for revealing cell density variability across the tissue section, which has been integrated within the read counts normalization (**Figure 4A**). A t-SNE analysis revealed 6 clusters presenting distinct anatomical locations when projecting this information across the spatial SpExels locations (**Figure 4B**). Tissue-associated Gene Ontology enrichment analysis confirmed the anatomical classification, illustrated by the spinal cord signature retrieved in cluster 2&5; the epicardial progenitor cell hearth association in cluster 1&4; a strong nervous tissue association in cluster 5; the presence of olfactory ensheathing cell signatures in cluster 3, known to be derived from neural crest cells and detectable by the expression of *Sox10* at E11.5 (Perera et al. 2020) ; or a ventral otocyst brain signature in cluster 6&4; corresponding to progenitor cells of the inner ear deteactable at E10.5 thanks to the expression of the transcription factor *Pax2* (Durruthy-Durruthy et al. 2014) (**Supplemental Figures S6,S7** & **Figure 4C**).

In addition to the H3K27ac local enrichment in the promoter regions of *Sox10* or *Pax2*, other genes like L*hx5* (LIM Homeobox 5), known to be broadly expressed in the early embryo but later restricted to the nervous system (Peng and Westerfield 2006); the transmembrane protein 80 (T*mem80*), previously shown to be expressed in the head surface ectoderm

(RNA-seq assays in embryonic mice (E8.5, E9.5, E10.5) (Brunskill et al. 2014)), or two members of the Wnt familly: *Wnt4*, previously shown to be expressed in the spinal cord of E9.5 and E10.5 mouse embryos (Agalliu et al. 2009); or *Wnt11,* shown to be required for proper patterning of the neural tube and somites by regulating notochord formation (Andre et al. 2015) were also enriched for this histone modification (**Figure 4D**). The relevance of finding H3K27ac enrichment in the aforementioned gene promoters was also confirmed in the spatial epigenomics cartography previously described by Deng et al (**Supplemental Figure S6**) (Deng et al. 2022), but also by gene expression data assessed in mouse embryo (E11.5) by the Allen developping mouse brain atlas (**Figure 4E and Supplemental Figure S6**).

While several of the retrieved signatures in the spatial H3K27ac cartography do match with previously described gene expression patterns, this histone modification alone is not a fateful marker of active transcription. Indeed, previous studies demonstrated a correlative relationship between the presence of co-occurent histone modifications at promoter regions with the transcriptional status of the associated gene (Ernst and Kellis 2010, 2012). For instance, promoter regions hosting both H3K4me3 and H3K27ac signatures are known as markers for active transcription, while promoters enriched on H3K27me3 correspond to repressed regions. Furthermore, promoters hosting the H3K27me3 modification and either H3K4me3 or H3K27ac are described as “bivalent promoters”, corresponding to genes devoid of transcription acitivity, but prone to initiate gene expression (Ernst and Kellis 2010, 2012). Herein, we aimed to evaluate this potential combinatorial co-enrichment signatures (previously described as chromatin state analysis(Ernst and Kellis 2010, 2012)) for three major histone modifications: H3K4me3, H3K27ac and H3K27me3. Hence, in addition to the aforementioned H3K27ac spatial cartography, we have also generated spatial epigenomic maps for H3K4me3 and H3K27me3 issued from consecutive sections (**Figure 4F**). Since mouse embryo sections are not located in exactly the same physical position on the manufactured DNA arrays, we have computationally realigned all three spatial epigenomic

maps, leading to the generation of a spatial pseudo-map, used in a second time for interrogating histone enrichment signatures per section, but also across all three overlayed sections (**Figure 4G&H**). The obtained pseudo map has been first used to verify the spatial co-enrichment of the studied histone modifications in the case of the gene *Wnt11*. While all three histone modifications were retrieved over-enriched in *Wnt11* promoter regions (**Figure 4D&F**), they do not systematically overlay in their spatial localizations within the embryo, as revealed when interrogating for contiguous over-enriched SpExels (spatial epigenomics elements) within the pseudo-map. Indeed, distinct spatial localizations within the embryo appeared to be enriched for both H3K4me3 and H3K27ac or the repressive mark H3K27me3 (**Figure 4I**). This is further confirmed when interrogating for *Wnt11* co-enrichment promoter signatures (i.e. other genes presenting the same histone modification erichments in their promoters and spatially colocalized with *Wnt11*). Wnt11 H3K27ac/H3K4me3 (active promoter signature) co-enrichment patterns were associated to the trunk notochord region, while the repressive mark H3K27me3 appears to be enriched to the tail of the embryo, (**Figure 4J**), in aggreement with *Wnt11* expression described in previous studies (**SupplementalFigure S7**) (Andre et al. 2015; Murphy et al. 2022).

Expanding the spatial co-enrichment analysis over all genes captured in the spatial epigenomic maps revealed > 600 promoters presenting H3K4me3/H3K27ac co-enrichment signatures (State 3: active promoters), 4,759 repressed promoters (State 1: enriched for H3K27me3) and > 1,600 bivalent promoters; defined by their enrichment in H3K27me3 and either H3K4me3 (State 4: 681 genes), or H3K27ac (State 2: 974 genes) (**Figure 4K**). Finally, a large fraction of gene promoters were retrieved spatially associated only to H3K4me3 (State 5: 3,500) or H3K27ac (State 6: 5,140), potentially indicating that we failed to properly associate them to a co-enrichment signature due to technical issues (e.g. > 5 contiguous over-enriched SpExels; sequencing depth; antibody performance; etc).

As illustrated for the case of the *Wnt11* promoter (**Figure 4I & 4J**), chromatin states in promoter regions do change across tissue section. To characterize these changes, we have

selected genes based on the number of over-enriched SpExels across the tissue (>40% occupied SpExels when considering all chromatin state combinations; and >10% for at least one of the chromatin states). A total of 750 genes were selected, revealing 291 promoters fluctuating between the active to the repressive state, passing by the bivalent situations. Furthermore, the remaining 459 genes were devoid of the active chromatin state combination (State 3), but presented the repressive mark H3K27me3 (**Figure 4L**). In some cases these promoters shifted to the bivalent states (62 genes) or replaced the repressive mark by either "H3K4me3 alone" or "H3K27ac alone" states (80 genes). Furthermore, in some cases, they fluctuated between one of the bivalent states and the presence of either "H3K4me3 alone" or "H3K27ac alone" states (143 genes hosting the chromatin state S2, 102 genes hosting the S4). Finally, there are 72 genes that do not present the H3K4me3 alone situation, but fluctuates between the repressive state (H3K27me3), the bivalent state S2 (H3K27me3/H3K27ac) or the presence of only the histone modification H3K27ac in their promoter regions (**Figure 4L**). The fact that these 459 gene promoters are devoid of the active chromatin state, but at the same time do present either H3K4me3 or H3K27ac alone, do support the possibility of missing such signature due to technical limitations. For instance, in the case of the *Hoxb4* promoter, our spatial epigenomics profiling clearly revealed a strong repressive signature (H3K27me3) other than the spinal cord region. In contrary, the H3K4me3 and H3K27ac profiling do present a spinal cord enrichment localization (**Figure 4M**). *Hoxb4* gene co-enrichment similarity maps revealed a bivalent chromatin state configuration at the tail of the embryo, and a co-enrichment signature of the histone modifications H3K4me3 and H3K27ac at the medial level of the spinal cord (**Figure 4M**), in agreement with gene expression revealed by in situ hybridization (Allen developping mouse brain atlas; **Figure 4N**).

Overall, this complex combinatorial chromatin state transitions clearly reflects tissue complexity, but also provide means to deeply interrogate histone modification signature transitions across the developping system. It is clear, that having multiple histone

modification profiles and consecutive sections appear essential for further exploiting this molecular cartography.

Discussion

While recent developments in spatially resolved omics are revolutionizing the way in which we can explore tissue complexity, most of the available solutions are only focused in capturing the presence of RNA. Previous studies indicated that the half-life of RNA molecules within the cell can range from 40 minutes to 9 hours, notably based on its physiological function (Tani et al. 2012), suggesting that a transcriptomics profiling assay does not directly reflect the transcribing status of the cell at the time of its measurement. Furthermore, per definition a transcriptomics analysis can only identify readouts of expressed genes but it cannot distinguish between genes being repressed or kept in a prone transcriptional state, since both are transcriptionally inactive.

In bulk assays, the use of chromatin immunoprecipitation strategies for interrogating the presence of histone modifications across the genome became the optimal way to interrogate the transcription regulatory status at the time “t” at which samples are collected (Bannister and Kouzarides 2011). Similarly, the recent development of the enzymatic strategy “Cut&tag” (based on the use of a recombinant protein A and the Transposase Tn5 (Kaya-Okur et al. 2019)) allowed to gain in sensibility by reducing the number of required cells from millions to few thousands, and even reaching the single-cell level (Bartosovic et al. 2021).

In this study, we present a new method allowing to capture histone modification signatures in a spatially resolved manner by using a double-barcoded DNA array able to capture the ProteinA-Tn5. In this manner, cleaved chromatin is immobilized by creating a physical tagmentation with the DNA probes of the DNA array. While our technology is not as resolutive as the methodology presented by Deng et al. (Spatial-Cut&Tag: 20 microns microfluidic channels (Deng et al. 2022)), the flexibility for manufacturing DNA arrays covering custom-based surfaces at a cost-effective manner – thanks to the combinatorial

number of required probes – makes of our approach extremely powerful when aiming to interrogate large number of tissue sections even when presenting large dimensions (**Supplemental Figure S8**). For instance, major efforts, like the one performed by Ortiz et al, for establishing a molecular atlas of the adult mouse brain by spatially-resolved transcriptomics assays in 75 sections (Ortiz et al. 2020), could be expanded to the level of the histone modification signature transitions regulating gene transcription.

Indeed, our technology can be considered as a low resolution high-throughput screening strategy (i.e. retrieve novel targets thanks to the unbiased power provided by Illumina sequencing), which can be combined with more resolutive methods which require to identify potential list of probe panels for interrogating the tissue of interest by in situ hybridization (e.g. Epigenomic MERFISH (Lu et al. 2022)). It is worth also to note that denoising strategies, initially used for image treatment but now getting applied to spatial omics approaches (Jiao et al. 2024), can be combined with our approach, for obtaining high resolutive outcomes at a cost effective manner.

One main limitation of our current technology is that, despite the use of large sequencing-read coverage, a rather small fraction of reads did contribute to the spatial profile, thus leading to rather sparse view across the analyzed sections. By performing multiple assays (fresh-frozen tissues, but also FFPE samples) we have observed that the primary bioinformatics treatment, focused in detecting the dual barcode sequences, represent the current bottle-neck of the assay. Hence, instead of investing resources for providing further sequencing depth; we anticipate the elaboration of an enhanced version of our DNA arrays, where the length of the dual barcodes will be expanded to improve the recovery during the primary steps of the bioinformatics pipeline.

Finally, in this study we have also highlighted the necessity of collecting consecutive sections for acquiring distinct histone modification signatures, which can then be interrogated across sections for revealing chromatin state signatures in a spatially resolved manner. We speculate that this strategy will further gain in power in the following years with the decrease

in sequencing costs, thus allowing to collect multiple consecutive sections for reconstructing volumetric views of the chromatin state signatures driving tissue architecture.

Materials and Methods

DNA array manufacturing

DNA arrays for spatial epigenomics experiments were produced by depositing two types of complementary oligonucleotides, namely “Barcode for rows (BCr)” and “Barcode for columns (BCc)” hosting sequences. Like in our previous study describing the use of double-barcoded DNA arrays for spatial transcriptomics (Lozachmeur et al. 2023), we have used BCr oligonucleotides presenting: an amino C6 linker at the 5’ extremity, followed by four G or C nucleotides (S), a T7 promoter sequence (GACTCGTAATACGACTCACTATAGGG), a unique molecular idenfitier (UMI: WSNNWSNNV), a molecular barcode (8 nucleotides) associated to the printed row in the DNA array and a 30 nucleotides long adapter sequence with a GC-content of 40%, (here named as Gibson sequence: “ACATTGAAGAACCTG-TAGATAACTCGCTGT”).

The BCc oligonucleotide has been modified for spatial epigenomics applications as following: At its 5’ extremity, it presents a “CTGTCTCTTATACACATCT” (herein called as Mosaic sequences or MOS), corresponding to the sequence recognized by the transposase Tn5 (Reznikoff 2003). Then similarly to previous study, BCc contains a unique molecular identifier (UMI: WSNNWSNNV), a molecular barcode (8 nucleotides) associated to the printed column in the DNA array, and a complementary Gibson sequence.

Oligonucleotides were diluted to 5 µM in presence of sciSPOT Oligos B1 solution (Scienion CBD-5421-50) in a 96 wells plate and stored at -20°C for long storage. DNA arrays were printed onto Superfrost Plus Adhesion Microscope Slides (J7800AMNZ Epredia).

For mouse embryo assays, we have manufactured DNA arrays composed by 32×32×2 interstitially printed probes (2,048 different probes in total, 177 μm pitch distance; ~100 μm printed spot). For this, 32 BCr oligonucleotides presenting a unique molecular barcode were printed per row, by depositing ~250 pL (250 μm between contiguous spots) with a DNA pico-liter spotter (Scienion sciflexarrayer S3). Similarly, 32 different BCc oligonucleotides were printed per column, by spotting ~250 pL of each of them on top of the previoulsy prinded BCr nucleotides. Then, the same 32 BCr oligonucleotides were printed per row with a shifted position of 125 μm in both axis (i.e. interstitial printing), followed by the deposition of 32 different BCc oligonucleotides printed per column. In that manner a total 32 unique BCr (printed twice) and unique BCc oligonucleotides are required for reaching an array presenting 2,048 unique double-barcoded probes.

For mouse brain sections, we have generated 64×32×2 interstitially printed arrays (4,096 different probes), where 32 BCr oligonucleotides presenting a unique molecular barcode were printed per row, other 32 different BCr oligonucleotides were printed per row with a shifted position of 125 μm in both axis (i.e. interstitial printing), followed by the deposition of 64 different BCc oligonucleotides printed per column. Finally, for mouse paws studies, a simpler DNA array printout was prodruced (64×32 DNA arrays) by depositing 32 BCr probes per row and 64 BCc probes per column (250 microns pitch distance, 100 microns spot diametter).

In all printout designs, fiducial borders were printed per DNA array, by adding three row/columns of spots with the used pitch distance (250 μm) by printing a BCr1 oligonucleotide presenting a Cy3-label at the 3'-extremity. During fiducial manufacturing, some spots were skipped in purpose for creating asymetry which will help for image scanning.

UV irradiation (254 nm; 5 min) was applied for crosslinking, followed by T4 DNA polymerase elongation (depending on the size of the printing area: 15 μL to 40 μL μL /printed region with

coverslips: 0.06 U/µL T4 DNA Polymerase (New England Biolabs M0203L); 0.2 mM dNTPs (Life technologies R0141, R0151, R016, R0171) during 1 hour at 37°C). In order to evaluate the double-strand DNA elongation performance, a quality control assay was made within every batch production (18 slides per batch; 2 printed regions per glass slide for size areas of 8×16 mm (32×64 either linear or interstitially printed spots), or 3 printed areas for 8×8 mm arrays (2,048 spots)), by incorporating a dCTP-Cy3 nucleotide during the elongation (dATP/GTP/TTP at 0.2 mM, dCTP at 0.005 mM and dCTP-Cy3 at 0.002 mM (Cytiva PA53021)), and analyzed by fluorescent microscopy. After elongation, slides were washed with 0.1×SSC buffer (Merck S6639), then ddH_2O and finally spin dried (Labnet slide spinner C1303-T-230V, 4800 rpm). DNA array slides have been stored at +4°C in a sealed container.

Fresh-frozen tissue cryosectioning

Tissue samples were cryosectionned (16 µm sections for mouse embryo tissue and and 20 µm sections for mouse brain WT tissue), collected between -15°C and -12°C then deposited on the DNA-arrays defined regions. Tissue sections were fixed with 1% Para-formaldheyde (PFA) (Life technologies 28908) in 1×PBS (Life technologies 70011-036) at room temperature during 30 minutes, then washed three times with 1×PBS.

Mouse paws decalcification, formalin-fixed embedding, and sectioning

Mouse paws have been collected and fixed in 4% buffered formol in 1×PBS for 24 hours. Then paws have been decalcified for 7 consecutive days in an +4°C refrigerated room on agitation. A maximum of 5 paws, each inside a cassette, are decalcified in 1L solution 3% formic acid in PBS. At mid incubation time, decalcification solution has been renewed and paws have been cut in half and put back into the cassette. Once decalcification is finished, paws have been rinsed in ddH_20 to then be dehydrated and infiltrated with paraffin using an automated device (ASP300 Leica microsystems) following this standard protocol: 70% ethanol, 2 washes of 70% ethanol, 96% ethanol, 100% ethanol, half and half mix of 100% ethanol with 100% xylene, empty the incubation chamber and final step of 100% xylene,

three baths of paraffin to allow paraffin inflitration. Afterward, each half paw has beenembedded in paraffinblocks. All blocks of mouse paw FFPE can be stored at room temperature.

4µM thick sections around the Talus bone have been collected from those FFPE mouse paw blocks using a microtome. Paw FFPE sections were deparaffined and rehydrated following standard protocols on an automated device (BondRx, Leicamicrosystems).

Histone modifications associated chromatin capturing in situ

Tissue sections were permeabilized during 1 hour at room temperature with 1% Triton X-100 (Merck 93443) in 1×PBS solution, then washed three times with 1×PBS followed by 5 minutes incubation at room temperature with 0.1M HCl for chromatin loosening as described previously (Lu et al. 2022), and washed three times with 1×PBS.

Afterwads, sections were covered during 1 hour at room temperature with a blocking and permeabilizing mix (10% Normal Donkey serum (NDS) (Merck D9663), 0.25% Triton X-100 (Merck 93443), 0.1% Bovine Serum Albumin (BSA) (Merck A9418), 2 mM EDTA (Life technologies AM9260G), 0.5 mM Spermidine (Life technologies 132740010), 1 tablet EDTA-free Protease Inibitor Cocktail (PIC) (Merck 11873580001)) prepared in 1×PBS. Then tissue were washed twice with 150 mM NaCl Buffer (20 mM Hepes-KOH 1M pH=7.2 (Merck 391338), 150 mM NaCl; 0.1 % BSA, 0.5 mM Spermidine, 1 tablet PIC).

After that, sections were incubated with a solution containing 2 µg of a selected primary Antibody (anti-Histone H3 (tri methyl K4) antibody (Abcam ab213244); anti-Histone H3 (acethyl K27) antibody (Abcam ab177178); or anti-Histone H3 (tri methyl K27) antibody (Abcam ab 192985)), in 150 mM NaCl Buffer at room temperature during 1 hour, followed by washing twice with 150 mM NaCl Buffer. Then tissues were incubated at room temperature with a pre-incubated mix (2 hours) containing 1:1 molar ratio of unloaded fusion protein pA-Tn5 (in-house production (9.5 µM) prepared using the plasmid Addgene #124601 by following previous protocols (Kaya-Okur et al. 2019; Li et al. 2021), and the corresponding

secondary Antibody (donkey anti-rabbit IgG antibody (Merck SAB3700932). The pre-incubated mix was complemented with either a conjugated Alexa-488 secondary Antibody (donkey anti-rabbit IgG antibody, Alexa Fluor 488 (Life technologies A-21206)) or a conjugated Alexa-555 secondary Antibody (donkey anti-rabbit IgG antibody, Alexa Fluor 555 (Abcam ab 150062)), added at 1/1,000 dilution.

The pre-incubated mix was used in a 5 times molar excess relative to the number of moles of the MOS sequence deposited per DNA array: for 2,048 spots it corresponds to $2{,}56.10^{-12}$ moles; for 4,096 spots it corresponds to $5{,}12.10^{-12}$ moles.

After 3 hours of incubation with the assembled mix, tissue sections were washed twice with 300 mM NaCl Buffer (20 mM Hepes-KOH 1M pH=7.2, 300 mM NaCl; 0.1 % BSA, 0.5 mM Spermidine, 1 tablet PIC) to remove PA-Tn5-antibodies in excess. Then, tissue sections were incubated during 5 minutes with DAPI (Life technologies 62248) 1/5,000 dilution in 1×PBS, then washed twice with 1×PBS.

DNA arrays were scanned under the TRICT filter to reveal the presence of the fiducial borders as well as the physical position of tissue sections on top of the printed DNA arrays. In case of the use of the conjugated Alexa-488 secondary Antibody, an additionnal scanning using FITC filter was done to reveal the localisation of the studied modification of histone. Finally, an additional scanning of the tissue using DAPI filter was done to observe nuclear localisation and to get cell density information. After imaging, sections were incubated overnight at 37°C with 10 mM $MgCl_2$ (Life technologies AM9530G) in 300 mM NaCl Buffer for *in situ* tagmentation.

Next day, tissue sections were incubated with a mix containing 2.25 µL of EDTA 0.5 M, 2.75 µL of 10% Sodium dodecyl sulfate (SDS) (Merck 71736) and 0.5 µL of 20 mg/mL Proteinase K (Life technologies 25530049) in 50 µL of ddH_2O overnight at 37 °C. The mix volume was ajusted depending of the slide printing area.

Finally, slides were washed under agitation (300 rpm) in containers containing: 100 mL of preheated buffer 1 (2×SSC and 0.1% SDS) during 15 minutes at 50°C, then 10 minutes with buffer 2 (0.2×SSC) at room temperature and 10 minutres with buffer 3 (0.1×SSC) at room temperature. At last the slides were washed in ddH_2O and spin-dried.

After that, DNA arrays were incubated with 0.1 M NaOH solution during 10 minutes at room temperature and then neutralized after removing liquid with a mix containig 100 µL of 0.1 M NaOH, 11.8 µL of 10×TE and 6.5 µL of 1.25 M acetic acid during 2-3 minutes. Finally, slides were washed within a falcon containing buffer 3 (0.1×SSC), then into a falcon containing ddH_2O and spin-dried.

cDNA's complementary strand synthesis

DNA arrays were incubated with a poly (C) tailing mix containing 0.6 U/µL terminal transferase (New England Biolabs M0315), 1×terminal transferase buffer (New England Biolabs BO315), 0.3×$CoCl_2$ (New England Biolabs B0252), 0.2 mM dCTP and 0.01 mM ddCTP during 35 minutes at 37°C, then 20 minutes at 70°C, and cooled at 12°C. To avoid evaporation, each printing area was closed by using sealing chambers (Life technologies AB-0576 (size 25 µL) or AB-0577 (size 65 µL)). Finally, slides were washed within a falcon containing buffer 3, and a falcon containing ddH_2O, prior spin-dry.

DNA arrays were then incubated with a mix containing: 0.1 U/µL Klenow exo- (New England Biolabs M0212), 0.2 mg/mL BSA, 1×NEB2 buffer (New England Biolabs B7002), 0.5 mM dNTPs and an oligonucleotide (1 µM) having a complementary sequence for the poly (C) tailed sequence, as well as a 5'-extremity providing an adapter sequence ("GTTCAGACGTGTGCTCTTCCGATCTGGGGGGGGGH"). As in the preceding step to avoid evaporation, the reaction was performed by using sealing chambers.The following conditions were used: 47°C during 5 minutes (primer annealing), 37°C during 1 hour (extension), 10 minutes at 70°C (enzyme deactivation) and cooled et 12°C. In order to recover the synthesized complementary DNA strands, DNA arrays were incubated with 100

µL of 0.1 M NaOH solution during 10 minutes, then the liquid was collected and a second incubation with 0.1 M NaOH solution was repetead. The two fractions were collected together and neutralized (for 100 µL 0.1 M NaOH solution add 11.8 µL of 10×TE and 6.5 µL of 1.25 M acetic acid). DNA arrays were also neutralized with 100 µL 0.1 M NaOH solution mixed with 11.8 µL of 10×TE and 6.5µL of 1.25 M acetic acid, then washed with buffer 3 and ddH_2O and stored at 4 °C for eventual supplementary experiments.

The neutralized collected solution was precipitated with cold 100% ethanol (2×volume), 200 mM NaCl and 1 µL of 20 mg/mL glycogen (Merck 10901393001) and conserved at -20°C for at least 1 hour, then centrifuged at 12,500 rpm during 30 minutes, washed once with cold 70% ethanol and resuspended (after 30 minutes air drying) in 23 µL of ddH_2O.

Ilumina sequencing library preparation

Half of the resuspended cDNA's complementary-probe material (12ul) was adjusted to 23ul with water, then mixed with: 25 µL of Q5 hot start high fidelity 2×master mix (New England Biolabs M0464L), 1 µL of adapter seq1 primer (0.02 µM (GTTCAGACGTGTGCTCTTCCGATCT)) and 1 µL of adapter seq2 primer (0.02 µM, (TACACTCTTTCCCTACACGACGCTCTTCCGATCTGACTCGTAATAC, corresponding to a part of the T7-promoter sequence)). This mix was amplified by PCR (98°C, 30 s; 15 cycles: 98°C 10 s; 65°C 75 s; then 65°C 5 min; 12°C hold) and then cleanned at 0.9×ratio with SPRIselect beads (Beckman Coulter B23318) and resuspended in 24 µL of ddH_2O.

A second PCR amplification was done by mixing the 24 µL of adaptors linked cDNA's complementary-probe strand with: 25 µL of Q5 hot start high fidelity 2×master mix, 0.5 µL of universal primer and 0.5 µL of index primer (0.1 µM final concentration NEBNext multiplex oligos for Illumina index primers set 1 E7335) by using the following cycling conditions: 98°C for 30 s, 15 cycles of 98°C for 10 s and 65°C for 75 s; then final extension at 65°C for 5 min and hold at 12°C.

Then sample was cleanned at 0.9×ratio with SPRIselect beads. Finally, the library was used for Illumina sequencing (150 nts paired-ends sequencing; NovaSeq 6000; final loading concentration: 70 pM; Pool loading concentration: 0.35 nM).

Bioinformatics processing

Primary analysis has been performed with our in-house developped tool SysISTD (SysFate Illumina Spatial transcriptomics Demultiplexer: https://github.com/SysFate/SysISTD).

SysISTD takes as entry paired-end sequenced reads (FASTQ or FASTQ.gz format), and two TSV files, the first one containing the sequence of the molecular barcodes associated to the rows or column in the printed arrays and the second file presenting the physical position architecture of the spatial barcodes. SysISTD search for the Gibson sequence (regex query), then for two neighboring barcodes. Paired-reads presenting these features were aligned to the mouse (mm9) genome with Bowtie 2. In this manner, SysISTD generates SAM files per spatial coordinate, which are reprocessed with the help of featureCounts (Liao et al. 2014) and an in-house script for detecting reads counts per gene promoter (+/-5kb) and per spatial coordinate (SysISPD (SysFate Illumina Spatial Promoters Demultiplexer: https://github.com/SysFate/SysISPD). While the mouse reference genome mm10 is already available since December 2011, we are still working with the mm9 version because we systematically perform a direct comparison of our findings with the NGS-QC database, composed by > 100 thousand publicly available bulk ChIP-sequencing data built on the mouse mm9 / human hg19 reference genomes. Furthermore, >90% of the genes retrieved in the mm10 version are already available in the mm9 genome assembly.

As outcome, SysISPD generates a matrix presenting read counts associated to physical coordinates in columns and known transcripts ID associated to the evaluated promoters in rows. To focus the downstream analysis to the physical positions corresponding to the

analyzed tissue, we used an in-house R script taking as entry an image of the DNA array scanned with the TRICT filter, releaving the presence of the fiducial borders.

Specifically, we upload to R (R Core Team (2020)) a croped image within the fiducials (imager package) and we use the "px.flood" function to retrieve the pixels associated to the tissue. Finally we applied a pixel to gexel coordinates conversion prior to cross this information with the outcome of SysISPD.

In parallel, all SAM files per coordinate were merged into a single-file and processed with the peak caller MACS2 (Zhang et al. 2008). The obtained confident peaks (*P*-value < $1x10^{-5}$) were annotated to their proximal coding regions. Finally, the spatial coordinates/promoters matrix provided by SysISPD was filtered by selecting transcripts associated to promoter regions presenting confident peaks.

The "confident promoter peaks tissue-focused" matrix was quantile normalized with our previously described tool MULTILAYER (Moehlin et al. 2021b, 2021a). Then, the normalized matrix was further processed for revealing cell/tissue density contrast information by integrating normalized counts with pixel value levels retrieved in either the DAPI or the histone modification immunostaining image. The obtained normalized and pixels intensity adjusted matrix was used for all downstream analysis.

T-distributed stochastic neighbor embedding (t-SNE) analysis has been performed within R with the help of the Rtsne package. Projection of the stratified clusters (*k*-means metrics) was performed with the help of an in-house R script. Genes presenting over-represented counts per promoter and per cluster were used for performing a cell/tissue type enrichment analysis with Enrichr (Chen et al. 2013) and by using the Mouse gene Atlas, *Tabula Muris*, ARCHS4 (Lachmann et al. 2018), CellMarkerAugmented (Zhang et al. 2019) and Panglao (Franzén et al. 2019) databases.

Mouse embryo sections alignment and chromatin state transitions detection

SpExel physical positions for each mouse embryo section were realigned relative to the tail of the embryo, allowing to associated the same pseudo-coordinates for all evaluated sections.

The associated histone modification identifier was added to the name of each transcript within the aligned matrices and concatenated into a single matrix allowing to interrogate for contiguous SpExels within sections and enrichment patterns similarity across sections, as described previously in MULTILAYER (Moehlin et al. 2021b, 2021a). Contiguous promoter patterns and promoter co-enrichment patterns were captured by MULTILAYER over the embryo pseudo-map hosting data of all three histone modifications. Chromatin state transitions were computed by extracting promoter co-enrichment patterns similarity over all three sections with at least 5 consecutive SpExels.

Data access

All datasets generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE286251.

Furthermore, processed spatial matrices as well as the corresponding scanned images are available in Mendeley Database: https://data.mendeley.com/datasets/d73mg855tw/4

Acknowledgements

We thank all members of the team SysFate for contributing to the discussion of this project and the Genoscope sequencing platform for their technical support. Notably we are extremely thankful for the effort performed by Ms. Sandrine Lebled, Isabelle Bordelais, and Alice Moussy from the Genoscope sequencing platform, for finding the optimal conditions for the deposition of Illumina libraries within the sequencing instrument. This work was supported by the institutional bodies CEA, CNRS, Université d'Evry-Val d'Essonne. M.G.M.F

has been funded by the Institut National du Cancer (INCa: Funding N° 2020-124; N° 2022-078), G.L by the ANRT doctoral fellowship CIFRE N° 2021/1899 in collaboration with Novalix. M. D is supported by EU grant HORIZON N° 101070740. Mouse embryos were kindly provided by Wojciech Krezel (IGBMC, Strasbourg; France). Mouse brain samples were kindly provided by Jean-Pierre Mothet (LuMin; ENS Paris Saclay; France). Mouse paws were provided by Novalix.

Author contributions

MG Mendoza-Ferri: formal analysis, investigation, and methodology. G Lozachmeur: resources and methodology. M Duvina: data curation, software, and formal analysis. L Perret: resources and methodology. D Merciris: resources and methodology. A Gigout: supervision, review and editing. MA Mendoza-Parra: conceptualization, formal analysis, supervision, funding acquisition, and writing—original draft.

Conflict of interest

Patent application pending (MENDOZA-PARRA, Mr. Marco Antonio; Application number: EP24307128; date of submission: 17 December 2024).

References

Agalliu D, Takada S, Agalliu I, McMahon AP, Jessell TM. 2009. Motor Neurons with Axial Muscle Projections Specified by Wnt4/5 Signaling. *Neuron* **61**: 708–720.

Andre P, Song H, Kim W, Kispert A, Yang Y. 2015. Wnt5a and Wnt11 regulate mammalian anterior-posterior axis elongation. *Development* **142**: 1516–1527.

Bannister AJ, Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell Research* **21**: 381–395.

Bartosovic M, Kabbe M, Castelo-Branco G. 2021. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat Biotechnol* **39**: 825–835.

Blum M, Cholley P-E, Malysheva V, Nicaise S, Moehlin J, Gronemeyer H, Mendoza-Parra MA. 2020. A comprehensive resource for retrieving, visualizing, and integrating functional genomics data. *Life Sci Alliance* **3**: e201900546.

Brand DD, Latham KA, Rosloniec EF. 2007. Collagen-induced arthritis. *Nat Protoc* **2**: 1269–1275.

Brunskill EW, Potter AS, Distasio A, Dexheimer P, Plassard A, Aronow BJ, Potter SS. 2014. A gene expression atlas of early craniofacial development. *Dev Biol* **391**: 133–146.

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**: 128.

Chen K, Chen Z, Wu D, Zhang L, Lin X, Su J, Rodriguez B, Xi Y, Xia Z, Chen X, et al. 2015. Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat Genet* **47**: 1149–1157.

Deng Y, Bartosovic M, Kukanja P, Zhang D, Liu Y, Su G, Enninful A, Bai Z, Castelo-Branco G, Fan R. 2022. Spatial-CUT&Tag: Spatially resolved chromatin modification profiling at the cellular level. *Science* **375**: 681–686.

Duncan I, Danziger N, Duncan D, Hemmerich A, Edgerly C, Huang R, Vergilio J-A, Elvin JA, He J, Britt N, et al. 2019. Acid-Based Decalcification Methods Compromise Genomic Profiling from DNA and RNA. *Blood* **134**: 4659.

Durruthy-Durruthy R, Gottlieb A, Hartman BH, Waldhaus J, Laske RD, Altman R, Heller S. 2014. Reconstruction of the mouse otocyst and early neuroblast lineage at single-cell resolution. *Cell* **157**: 964–978.

Ernst J, Kellis M. 2012. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**: 215–216.

Ernst J, Kellis M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature Biotechnology* **28**: 817–825.

Franzén O, Gan L-M, Björkegren JLM. 2019. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database* **2019**: baz046.

He P, Wang B-H, Cao R-R, Zhu D-C, Ge B, Zhou X, Wu L-F, Lei S-F, Deng F-Y. 2021. ITGA2 protein is associated with rheumatoid arthritis in Chinese and affects cellular function of T cells. *Clinica Chimica Acta* **523**: 208–215.

Hu S-L, Thadevoos LA, Ho T-L, Lin Y-Y, Chen H-T, Huang C-C, Su C-M, Tang C-H. 2024. FGF23 facilitates IL-1β synthesis in rheumatoid arthritis through activating PI3K, Akt, and NF-κB pathways. *Environ Toxicol* **39**: 3283–3291.

Jiao S, Lu D, Zeng X, Wang T, Wang Y, Dong Y, Peng J. 2024. DiffuST: a latent diffusion model for spatial transcriptomics denoising. *bioRxiv* 2024.06.19.599672.

Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K, Henikoff S. 2019. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**: 1930.

Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, Silverstein MC, Ma'ayan A. 2018. Massive mining of publicly available RNA-seq data from human and mouse. *Nature Communications* **9**: 1366.

Li L, Freitag J, Asbrand C, Munteanu B, Wang B-T, Zezina E, Didier M, Thill G, Rocher C, Herrmann M, et al. 2022. Multi-omics profiling of collagen-induced arthritis mouse model reveals early metabolic dysregulation via SIRT1 axis. *Scientific Reports* **12**: 11830.

Li Y, Nakka K, Olender T, Gingras-Gelinas P, Wong MM-K, Robinson DCL, Bandukwala H, Palii CG, Neyret O, Brand M, et al. 2021. Chromatin and transcription factor profiling in rare stem cell populations using CUT&Tag. *STAR Protocols* **2**: 100751.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.

Lozachmeur G, Bramoulle A, Aubert A, Stüder F, Moehlin J, Madrange L, Yates F, Deslys J-P, Mendoza-Parra MA. 2023. Three-dimensional molecular cartography of human cerebral organoids revealed by double-barcoded spatial transcriptomics. *Cell Rep Methods* **3**: 100573.

Lu T, Ang CE, Zhuang X. 2022. Spatially resolved epigenomic profiling of single cells in complex tissues. *Cell* **185**: 4448-4464.e17.

Magyari L, Varszegi D, Kovesdi E, Sarlos P, Farago B, Javorhazy A, Sumegi K, Banfai Z, Melegh B. 2014. Interleukins and interleukin receptors in rheumatoid arthritis: Research, diagnostics and clinical implications. *World J Orthop* **5**: 516–536.

Mendoza-Parra MA, Saleem M-AM, Blum M, Cholley P-E, Gronemeyer H. 2016. NGS-QC Generator: A Quality Control System for ChIP-Seq and Related Deep Sequencing-Generated Datasets. In *Statistical Genomics: Methods and Protocols* (eds. E. Mathé and S. Davis), *Methods in Molecular Biology*, pp. 243–265, Springer New York, New York, NY https://doi.org/10.1007/978-1-4939-3578-9_13 (Accessed October 10, 2018).

Mendoza-Parra M-A, Van Gool W, Mohamed Saleem MA, Ceschin DG, Gronemeyer H. 2013. A quality control system for profiles obtained by ChIP sequencing. *Nucleic Acids Res* **41**: e196.

Moehlin J, Koshy A, Stüder F, Mendoza-Parra MA. 2021a. Protocol for using MULTILAYER to reveal molecular tissue substructures from digitized spatial transcriptomes. *STAR Protoc* **2**: 100823.

Moehlin J, Mollet B, Colombo BM, Mendoza-Parra MA. 2021b. Inferring biologically relevant molecular tissue substructures by agglomerative clustering of digitized spatial transcriptomes with multilayer. *cels* **12**: 694-705.e3.

Murphy P, Armit C, Hill B, Venkataraman S, Frankel P, Baldock RA, Davidson DR. 2022. Integrated analysis of Wnt signalling system component gene expression. *Development* **149**.

Ortiz C, Navarro JF, Jurek A, Märtin A, Lundeberg J, Meletis K. 2020. Molecular atlas of the adult mouse brain. *Sci Adv* **6**: eabb3446.

Peng G, Westerfield M. 2006. Lhx5 promotes forebrain development and activates transcription of secreted Wnt antagonists. *Development* **133**: 3191–3200.

Perera SN, Williams RM, Lyne R, Stubbs O, Buehler DP, Sauka-Spengler T, Noda M, Micklem G, Southard-Smith EM, Baker CVH. 2020. Insights into olfactory ensheathing cell development from a laser-microdissection and transcriptome-profiling approach. *Glia* **68**: 2550–2584.

Reznikoff WS. 2003. Tn5 as a model for understanding DNA transposition. *Mol Microbiol* **47**: 1199–1206.

Samee N, Geoffroy V, Marty C, Schiltz C, Vieux-Rochas M, Levi G, de Vernejoul M-C. 2008. Dlx5, a Positive Regulator of Osteoblastogenesis, is Essential for Osteoblast-Osteoclast Coupling. *Am J Pathol* **173**: 773–780.

Song S, Wang Y, Fang Y, Teng F, Yue T, Chen L, Wan X, Hong R, Xue M. 2024. Glucocorticoids promote joint microenvironment alteration of GABBR1 expression associated with mitigating rheumatoid arthritis. *Cell Mol Biol (Noisy-le-grand)* **70**: 73–77.

Tani H, Mizutani R, Salam KA, Tano K, Ijiri K, Wakamatsu A, Isogai T, Suzuki Y, Akimitsu N. 2012. Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res* **22**: 947–956.

Wehmeyer C, Naylor A, Moeller K, Poologasundarampillai G, Pap T, Buckley C. 2019. P124/O27 Osteocyte-derived podoplanin is an important regulator of bone remodelling in the KBxN serum transfer model of rheumatoid arthritis. *Ann Rheum Dis* **78**: A55.

Zhang D, Deng Y, Kukanja P, Agirre E, Bartosovic M, Dong M, Ma C, Ma S, Su G, Bao S, et al. 2023. Spatial epigenome–transcriptome co-profiling of mammalian tissues. *Nature* **616**: 113–122.

Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, Luo T, Xu L, Liao G, Yan M, et al. 2019. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Research* **47**: D721–D728.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137.

Figure Legends

Figure 1. Principle behind the double-barcoded spatial epigenomics technology. (A) DNA arrays are printed by depositing first probes in a row manner, followed by deposition of a second type of probes in a column manner. The first type of probes is composed by a T7 promoter (T7p) sequence, a unique molecular barcode associated to the row (BCr1, …BCri) and bridge sequence called “Gibson”. The second type of probes present a complementary Gibson sequence, a unique molecular barcode associated to the column (BCc1, …BCcj) and a complementary mosaic sequence (MOS’). **(B)** Since both probes were combined during printing, after UV irradiation for immobilization, hybridized probes are elongated with the T4 DNA polymerase, leading to a double-strand molecule presenting the mosaic sequence (MOS) at the free end of the probe. **(C)** After depositing a tissue section on top of the manufactured DNA array, a 1st antibody against the protein of interest is incubated, followed by a second antibody against the 1st antibody and finally incubated with the recombinant protein A-transposase PA-Tn5. **(D)** Scheme illustrating the loading of the Tn5 into the Mosaic sequence retrieved on the printed probes on the DNA array, followed by the chromatin

cleavage induced by Magnesium chloride, leading to the formation of hairpin-like DNA structures. **(E)** Scheme illustrating the molecular biology steps required for generating a copy of the genomic DNA captured by the DNA probes, followed by their amplification for Illumina massive parallel DNA sequencing.

Figure 2. Spatial epigenomics profiling over a whole mouse brain section. (A) Micrograph displaying the scanning of a mouse brain coronal section deposited on top of a DNA array composed of 64×32 interstitial printed probes harboring the Mosaic sequence. Three rows of Cy3-labelled probes are visible at the border of the micrograph corresponding to fiducials used for defining the physical position of the tissue section across the 4,096 printed probes. Notice that the DNA array covers a surface of 16×8 millimeters. **(B)** micrograph displaying a bright-field scanning of a DNA array printed area where each of the printed probes are visible thanks to salt deposition. Thanks to the interstitial printing, the pitch distance resolution of the array is ~177microns. **(C)** Immunolabelling of the section displayed in (A) revealing the histone modification H3K4me3. The Inset magnification area reveals the nuclear staining visible through this labelling, as well as the differences in cell density across different parts of the mouse brain section. **(D)** Electropherogram displaying the spatial epigenomics (SpE) Illumina sequencing library obtained from the mouse brain section. **(E)** Number of total sequenced reads, as well as those recovered at different stages of the primary bioinformatics analysis. **(F)** Violin plots displaying the number of read counts or promoter regions per SpExel (Spatial Epigenomics element, in analogy to Pixels: Picture elements). **(G)** Digital display corresponding to the number of adjusted read counts associated to the physical positions in the mouse brain section (heatmap displayed in logarithmic scale). **(H)** Venn-diagram displaying the number of promoters presenting H3K4me3 peaks in common between the spatial epigenomics map (SpE) and two bulk H3K4me3 ChIP-sequencing profiles (GSM1656749; GSM1000095). For this comparison, all mapped reads across the tissue were collapsed to generate a pseudo-bulk profile and processed with the peak caller MACS2 (*P*-value $<1x10^{-5}$). **(I)** Peak-centered heatmap for the H3K4me3 peaks retrieved in common between the spatial epigenomics map (SpE) and both bulk datasets. Insets display average peak density retrieved in 9 peak classes issued from peak clustering based on their shape. **(J)** Common peaks retrieved between the spatial H3K4me3 profile and the spatial data published by Deng. et al. (GSM5622964). Notice that more than 5,000 to 8,000 peaks are retrieved in common, on the grounds of the confidence threshold used for the comparison (MACS2 peak calling applied on the corresponding pseudo-bulk data). **(K)** Genome browser views assessed around the promoter region associated to the gene *Adcy5* (Adenyl cyclase). Top panel: Bulk ChIP-sequencing data

issued from mouse brain samples and visualized within the QC Genomics genome browser. Heatmap on the bottom of each profile corresponds to a local QC enrichment assessment (Mendoza-Parra et al; NAR 2013). Bottom: Pseudo-bulk enrichment profiles issued either from the public dataset generated by Deng. et al. (GSM5622964; Blue), or from the H3K4me3 enrichment map generated in this study (black). **(L)** Spatial H3K4me3 enrichment signatures retrieved in the promoter region associated to the gene *Adcy5*. **(M)** Same as (L) but expressed in a differential enrichment context relative to the average levels across the tissue. **(N)** Mouse brain tissue section analyzed by Deng et al; and the corresponding H3K4me3 enrichment spatial signature associated to the gene *Adcy5*. Image adapted from Deng et al. Science 2022. **(O)** In situ hybridization (ISH) on a mouse brain section for the gene *Adcy5* (Allen mouse brain atlas).

Figure 3. Spatial epigenomics assay in decalcified FFPE embedded tissues. (A) Strategy for analyzing mice paw tissues issued from a collagen-induced arthritis (CIA) model. **(B)** Representative mouse paw tissue section revealing the presence of the tibia, talus, navicular, cuneiform and meta-tarsus bone. **(C)** H3K27-acethylation (H3K27ac) spatial epigenomics profiling performed on a mouse paw tissue section deposited on a double-barcoded DNA array (64×32 printed probes). **(D)** Number of total sequenced reads and those recovered at different stages of the primary bioinformatics analysis. **(E)** Violin plots displaying the number of read counts or promoter regions per SpExel (Spatial Epigenomics element, in analogy to Pixels: Picture elements). **(F)** Digital view of the normalized and adjusted number of read counts associated to the physical location (SpExel) of the tissue section. **(G)** T-distributed stochastic neighbor embedding dimensionality reduction analysis (t-SNE) displaying the presence of 6 clusters of SpExels; **(H)** Clusters' projection on top of the physical position within the tissue. **(I)** Cell types gene marker association analysis performed per clusters identified in (G). Relevant cell types retrieved enriched per cluster are highlighted (green boxes). **(J)** Example of gene promoters found enriched in H3K27ac read counts and associated to the aforementioned clusters. Cell type annotations indicated aside are collected from the CellMarker augmented database used for this analysis. **(K)** Scheme of the Collagen-induced arthritis (CIA) immunization assay performed by Lingzi et al. allowing to collect bulk-transcriptomics data from mice paws soft tissue at different time-points. **(L)** Heatmap revealing the differential gene expression signatures issued from the soft-tissue transcriptomes, and classified in 7 gene co-expression paths. Along the public transcriptome, the detection of H3K27ac over-enrichment across the decalcified FFPE mouse paw section (number of over-enriched SpExels) is displayed (blue heatmap). **(M)** Fraction of common genes between those associated to one of the gene co-expression paths in (L) and those

presenting a H3K27ac enrichment in the spatial epigenomics map. Notice that only H3K27ac over-enriched promoters (LFC: Log-fold change >1) being present in more than 1, 5, or 10% of total SpExels across the section were considered in this analysis.

Figure 4. Inferring chromatin state transitions across the mouse embryo by integrating consecutive spatial epigenomic landscapes. (A) Left panel: Scan of a DNA array (TRITC filter) hosting a cryosection of a mouse embryo (E11.5) and immuno-stained with an antibody targeting the histone modification H3K27-acethylation (H3K27ac). Notice the presence of Cyanin-3 (Cy3)-labelled DNA probes delimiting the DNA array composed of 32×32 interstitially printed probes. Middle panel: DAPI staining; Right panel: Digital map displaying the normalized read counts captured per physical position (SpExel) across the mouse embryo section. **(B)** t-SNE analysis allowing to stratify SpExels in 6 different clusters. Right panel: Projection of the stratified clusters within the digital map. **(C)** Tissue-cell types gene marker association analysis performed per clusters identified in (B). **(D)** Local enrichment signatures associated to 6 gene promoters presenting H3K27ac over-represented counts. **(E)** In situ hybridization (ISH) and gene expression data (Allen mouse brain atlas) for the gene *Sox10* and *Lhx5*, revealing its spatial signature, which is coherent with the H3K27ac digitized view displayed in D. **(F)** From left to right: Histone modifications immunostaining (TRITC filter: H3K4me3, H3K27me3); nuclei revealed by DAPI; digital map of the normalized read counts across the tissue section; local read counts enrichment associated to the gene Wnt11. **(G)** Strategy for interrogating changes in chromatin histone modification signatures across the embryo: (i) Overlay of all three histone modification maps; (ii) generate a pseudo map where SpExels of each digitized map are allocated to common pseudo-coordinates; (iii) Interrogate for contiguous enrichment patterns similarity within tissue section, but also across all three aligned sections. **(H)** Pseudo-map obtained from the overlay of all three-histone modification digital maps. **(I)** Promoter enrichment patterns associated to *Wnt11* gene (>5 contigous SpExels) retrieved in either of the histone modification maps. **(J)** Spatial gene promoter's co-enrichment analysis for *Wnt11* in the H3K27ac or H3K27me3 digital map. SpExels colored in red correspond to the location of *Wnt11*, while others correspond to other gene promoters sharing a similar spatial pattern (Tanimoto similarity index). Notice that these two spatial gene promoter's co-enrichment maps present distinct spatial localizations. **(K)** Heatmap displaying the co-occurring promoter enrichment patterns between all three histone modifications. Six chromatin co-occurring states were identified, and functionally associated to either active, repressed, or bivalent promoters, as described by (Ernst and Kellis 2010, 2012). Notice the presence of two states for which no functional association has been attributed: H3K4me3, or H3K27ac alone). **(L)** Gene promoters presenting different chromatin

co-occurring states across the tissue. **(M)** Top: Local read counts promoter enrichment for the gene *Hoxb4* in either of the histone modification maps within the pseudo map. Bottom: *Hoxb4* co-enrichment patterns revealing the presence of either bivalent (H3K27me3/H3K27ac or H3K27me3/H3K4me3) or promoter active regions (H3K27ac/H3K4me3). **(N)** In situ hybridization (ISH) and gene expression data (Allen mouse brain atlas) for the gene Hoxb4, revealing its spatial signature coherent with the H3K27ac digitized view displayed in M.

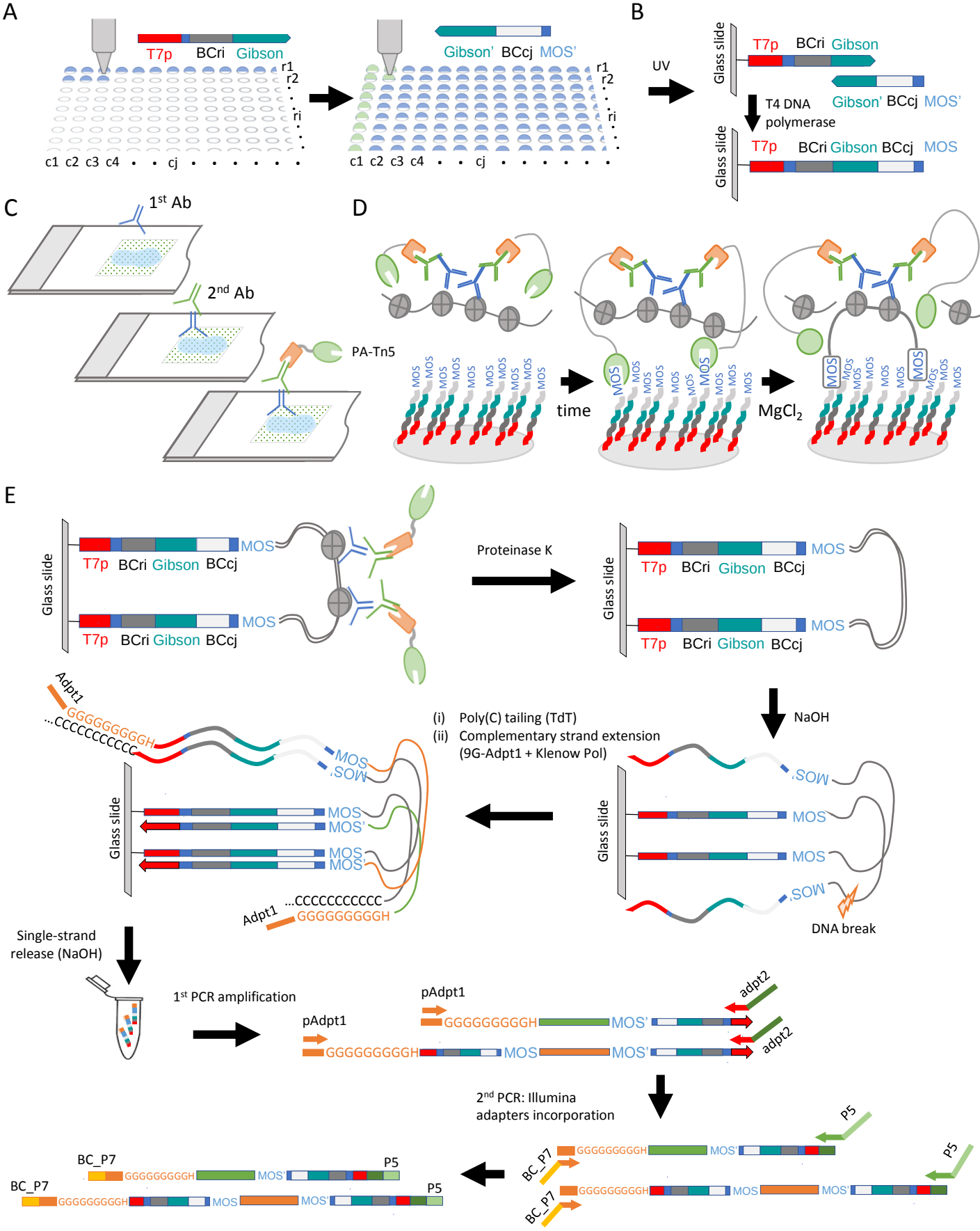


Figure 1

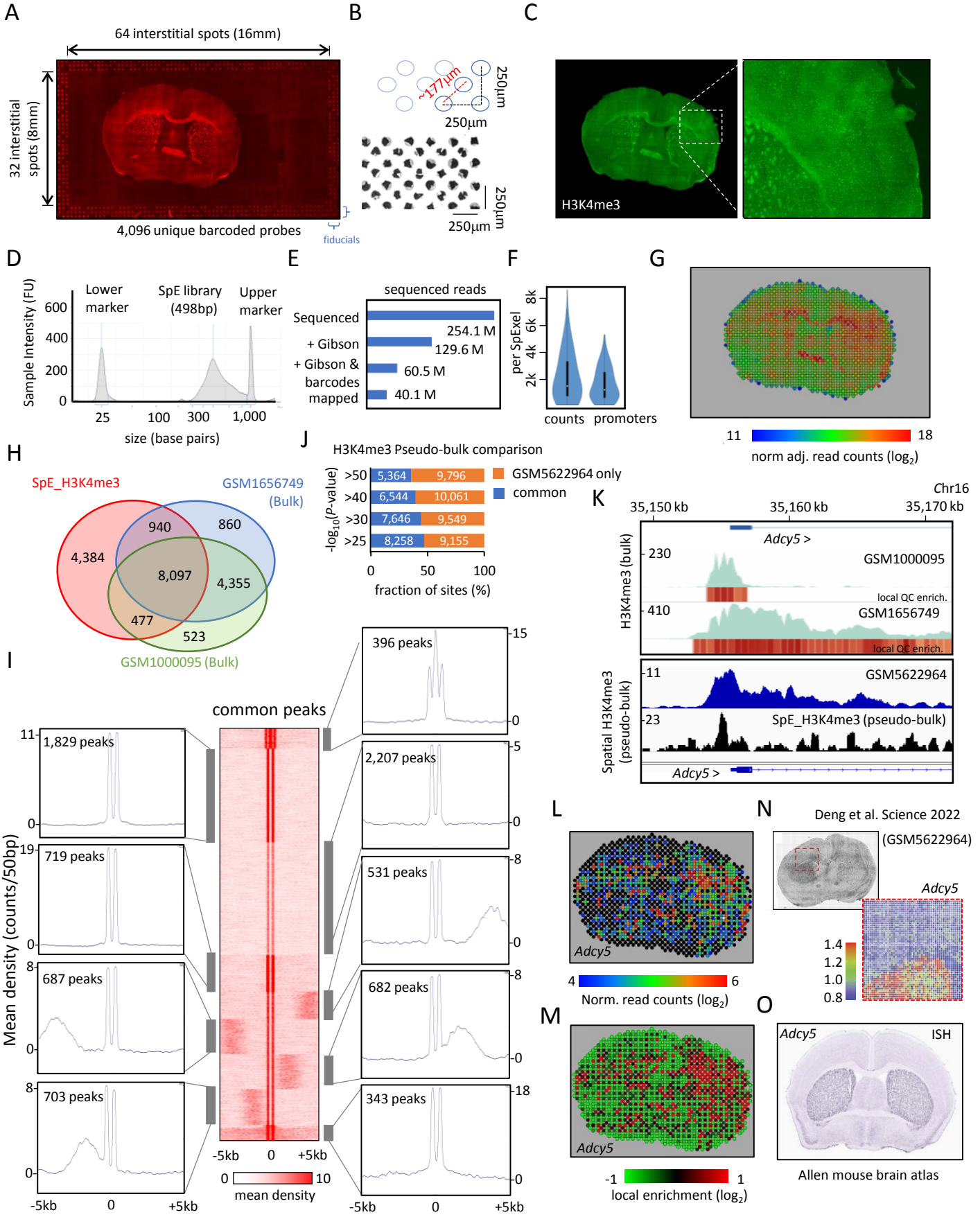
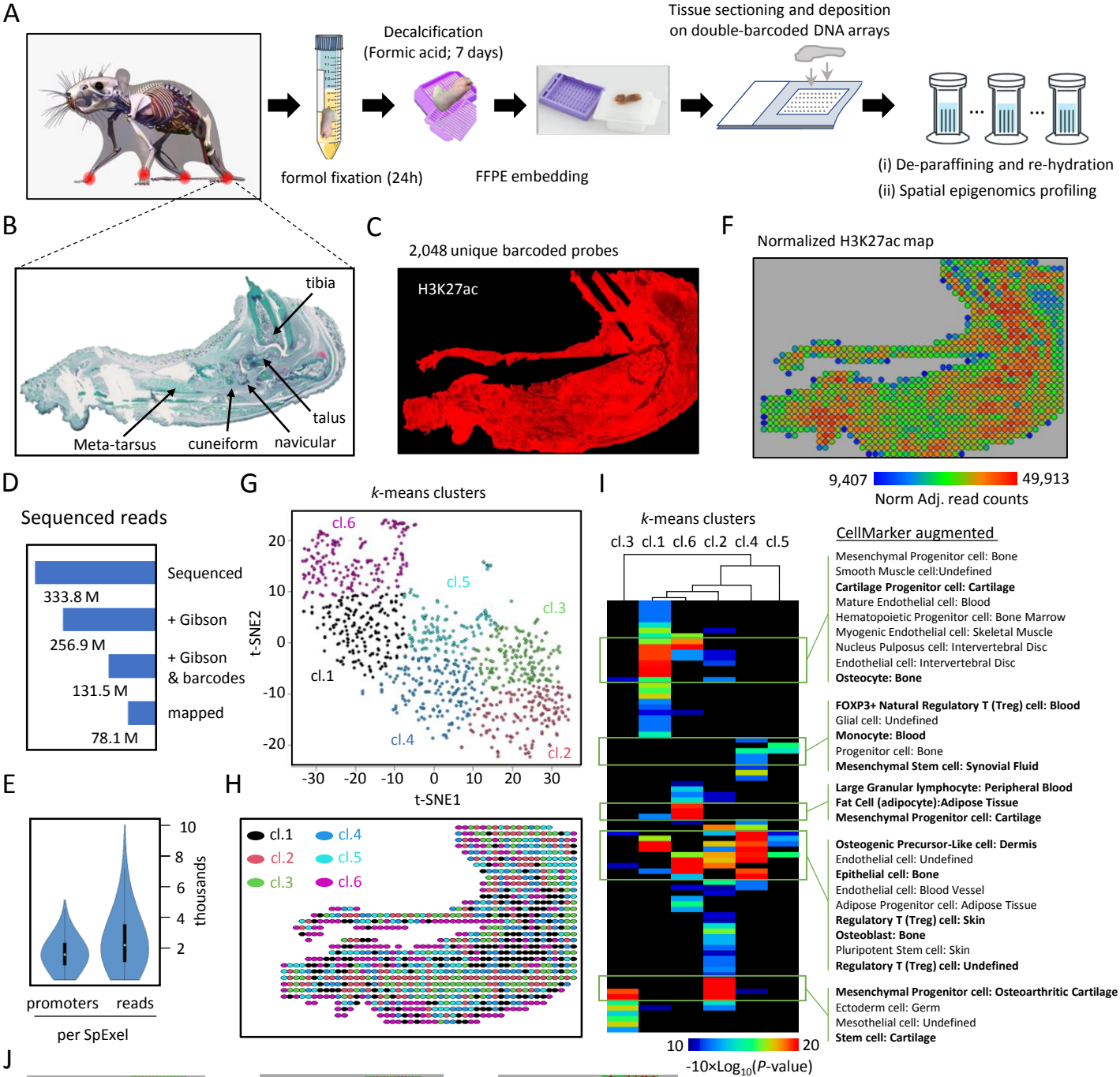
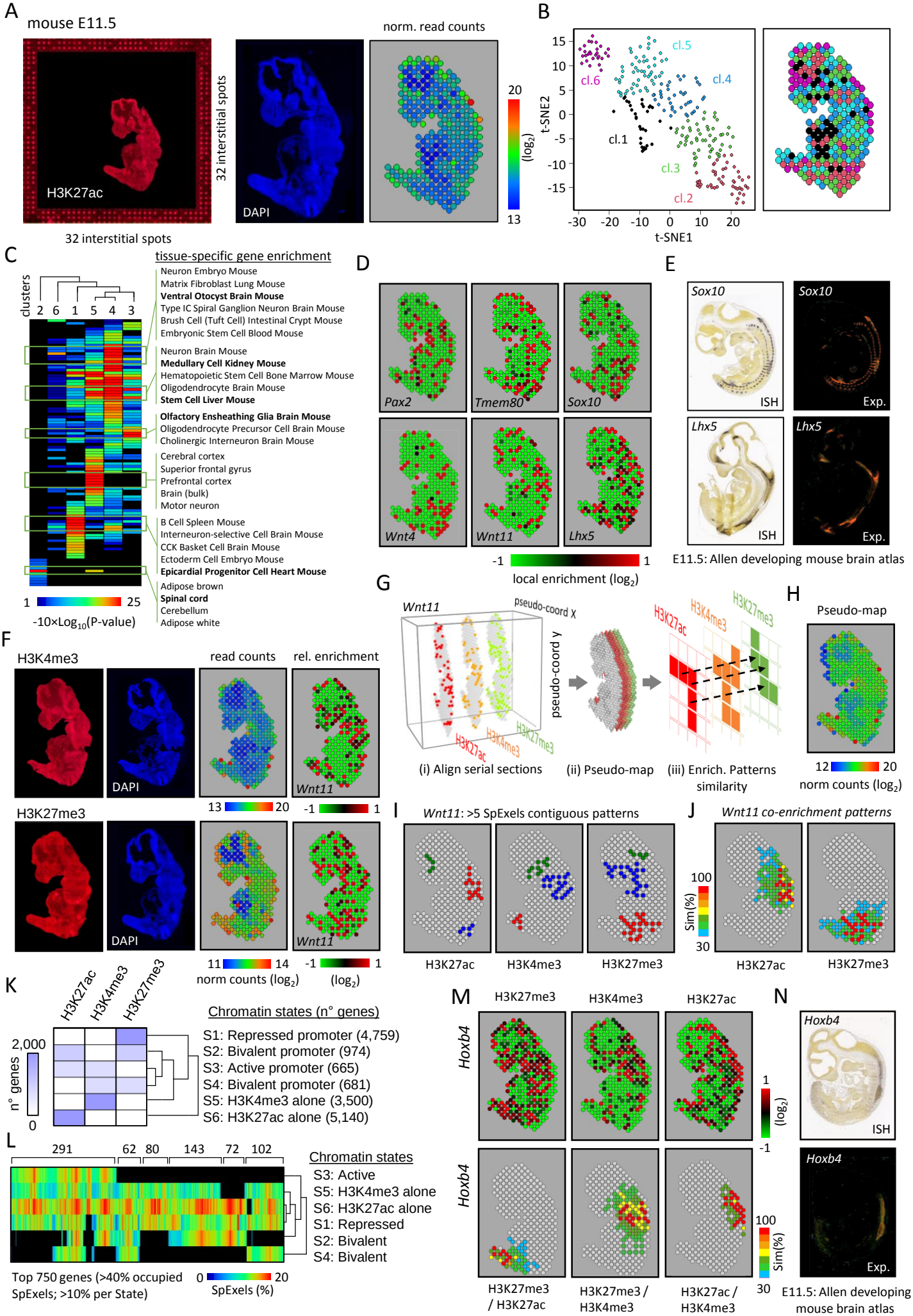


Figure2





Tissular chromatin states cartography based on double-barcoded DNA arrays capturing unloaded PA-Tn5 Transposase

Maria Grazia Mendoza-Ferri, Gwendoline Lozachmeur, Maximilien Duvina, et al.

Genome Res. published online May 13, 2025
Access the most recent version at doi:10.1101/gr.280305.124