ProA and ProB repeat sequences shape genome organization, and enhancers open domains

Authors
Konstantinn Acen Bonnet*[1], Nicolas Hulo*[2], Raphaël Mourad[3], Adam Ewing[4], Olivier Croce[5], Magali Naville[6], Nikita Vassetzky[7], Eric Gilson[5], Didier Picard[8], & Geneviève Fourel[9,#]

Affiliations
[1] Institute for Medical Biometry and Bioinformatics, Moorenstraße 5, 40225 Düsseldorf, Germany
[2] Institute of Genetics and Genomics of Geneva, University of Geneva, Geneva, Switzerland
[3] Molecular, Cellular and Developmental Biology department, Centre de Biologie Intégrative, University of Toulouse, CNRS, UPS, Toulouse, France
[4] Mater Research Institute, University of Queensland, Woolloongabba, QLD 4102, Australia
[5] Institut de Recherche sur le Cancer et le vieillissement, UMR7284, INSERM U1081, Faculté de Médecine de Nice, Université Côte d'Azur, Nice, France
[6] Institut de Génomique Fonctionnelle de Lyon, Ecole Normale Supérieure de Lyon, UCBL1, CNRS UMR 5242, Lyon, France
[7] Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 119991 Moscow, Russia
[8] Department of Molecular and Cellular Biology, University of Geneva, Geneva, Switzerland
[9] Laboratoire de Biologie et Modélisation de la Cellule, Ecole Normale Supérieure de Lyon, CNRS, UMR5239, Inserm U1293, Université Claude Bernard Lyon 1, Lyon, France

* These authors contributed equally to this work

lead contact: genevieve.fourel@ens-lyon.fr (G.F.)

SUMMARY

There is a growing awareness that repeat sequences (RepSeq) - the main constituents of the human genome - are also prime players in its organization. Here we propose that the genome should be envisioned as a supersystem with three main subsystems, each composed of functionally redundant, cooperating elements. We define herein ProA and ProB RepSeqs as sequences that promote either the A/euchromatin or the B/heterochromatin compartment. ProA and ProB RepSeqs shape A/B partitioning, such that the relative proportions of ProA and ProB RepSeqs determine the propensity of a chromosome segment to adopt either an A or a B configuration. In human, core ProA RepSeqs are essentially made of Alu elements, whereas core ProB RepSeqs consist of young L1 and some Endogenous Retroviruses (ERVs) as well as a panel of AT-rich microsatellites and pericentromeric and telomeric satellites. Additionally, RepSeqs with more indefinite character and, importantly, their derivatives known as "transcriptional enhancers", can shift between ProA and ProB functions and thus act to open or close specific chromatin domains depending on the cellular context. In this framework, genes and their promoters appear as a special class of RepSeqs that, in their active, transcribed state, reinforce the openness of their surroundings. Molecular mechanisms involve cooperativity between ProB elements, presumably underpinned by the condensate-like properties of heterochromatin, which ProA elements oppose in several ways. We provide strong arguments that altered CpG methylation patterns in cancer including a marked loss in the B compartment, result primarily from a global imbalance in the process of CpG methylation and its erasure. Our results suggest that the resulting altered methylation and impaired function of ProB RepSeqs globally weaken the B compartment, rendering it more plastic, which in turn may confer fate plasticity to the cancer cell.

SUMMARY 1
INTRODUCTION 3
RESULTS 3
Genome organization is largely inscribed in the DNA sequence 3
The identification of ProA and ProB elements in the human genome points to a key role of repeat sequences in genome organization 4
Identifying cis-determinants of A/B partitioning: ProA and ProB elements 4
Knowing RepSeq composition and density suffices to accurately predict HiC EV 6
Identifying core sets of ProA and ProB elements and mechanisms involved 6
Compositional genomics reveals general trends for the local densities of ProA and ProB elements, and a marked heterogeneity between functionally similar territories 7
DNase I Hypersentitive Sites oppose B compartment to open domains 8
DHS/enhancers are of RepSeq origin, switch from a ProB to a ProA state upon induction, and can globally unfold the genome 9
DHS/enhancers are of RepSeq origin 9
DHS/enhancers are dual-function elements that switch from a ProB to a ProA state upon induction 9
A scenario for domain opening that incorporates alteration of CpG methylation 10
Viruses drive global genome opening by switching or tuning selected sets of DHS/enhancers towards ProA 10
RepSeq ProB function is impaired in cancer by both specific and global mechanisms, resulting in genome opening and plasticity 11
DISCUSSION 13
BOX 1: Life history of retrotransposable elements 16
BOX2: CpG methylation in repression of L1s and ERVs, and in activation of Alu elements: mechanisms and functions. 17
BOX3: RepSeq transcription is the Achilles' heel of cancer cells 19
RESOURCES & METHODS 19
Resources 19
AorB vector 20
RepSeq insert count and copy count in the human genome 20
Correlation with HiC EV 20
Assessing feature density or summed signal in 100 kb bins 20
Multivariate linear model of HiC EV profile 21
GO analysis 21
Identification of "outlier bins": HUVEC.A(vecB), HUVEC.B (vecA) 22
Analysis of CpG methylation 22
Fig. 1: High conservation of the A/B compartment profile points to the existence of pivotal DNA determinants. 23
Fig. 2: A/B compartment toggle-switch points to ProA and ProB determinants as both enriched in the cognate compartment and correlating with HiC EV value. 24
Fig. 3: Identifiying ProA and ProB RepSeq in the human genome. 26
Fig. 4: CorrA and CorrB RepSeq local densities predict HiC EV. 27
Fig. 5: Compositional genomics reveals general trends for the local densities of ProA and ProB elements, and an outlier composition for outlier bins. 28
Fig. 6: Evidence that DHS open domains. 31
Fig. 7: The function of ProB RepSeqs is impaired in cancer, which drives genome opening. 33
Fig. 8: General principles for how densities in ProA and ProB elements determine A/B partitioning 35
REFERENCES 37

INTRODUCTION

The genome of higher eukaryotes is organized in the space of the nucleus essentially in two main compartments, heterochromatin and euchromatin, which are thought to correspond approximately to B and A compartments, respectively, as defined experimentally by HiC, and within which interactions between distant loci are preferentially detected (1, 2). B is significantly more compact and is unfavorable to gene expression such that the vast majority of expressed genes are in A. In a browser view, this translates into a succession of A and B domains as indicated by the HiC eigenvector profile (HiC EV) (1), in a characteristic pattern for a given cell type (Fig. 1). The A and B regions may be further divided into subdomains that constitute the elementary unit of unfolding/refolding of the genome (3). In vertebrates, these domains are called topologically associated domains (TADs) and are often delimited by boundary elements associated with DNA-binding factors, and in a high proportion of cases the transcription factor (TF) CTCF (4, 5). In a living cell, the chromatin fiber is pervasively set in motion due to the action of "active effectors" (6, 7), in particular cohesin rings that relentlessly extrude DNA, forming chromatin loops that enlarge until the cohesin ring detaches or arrests. This occurs preferentially when a cohesin ring encounters a bound CTCF molecule and is responsible for TADs appearing as triangles with a marked apex in contact maps (8-10). Furthermore, such pervasive chromatin loop extrusion by cohesins behaves as a force which globally opposes B compartment formation, and therefore insufflates dynamics into A/ B partitioning (11, 12).

While the genome itself is generally thought to be essentially identical between all or almost all cells of an organism and throughout development, A/B compartmentalization is largely undetectable in the early embryo and in the early G1 phase just after mitosis. Compartments arise gradually in both cases alongside with the emergence and reconstitution, respectively, of the H3K9me3/HP1α -based heterochromatin system (13-16), pointing to a key role of heterochromatin and associated coalescence forces in B compartment formation, and more generally in genome compartmentalization (17, 18), as supported by modeling (19, 20). The mechanisms responsible for the chromatin states underlying A and B compartments are now well understood. For each of A and B, they consist of a system of histone marks/chromatin modifier/chromatin binding factors, with the A and B systems being symmetrical but incompatible ("mutual opposition") (21). Moreover, interwoven feed-forward loops animate each of the A and B systems, enabling a memory of chromatin states at distinct time and physical scales, which is responsible for the apparent stability of A and B states (21-24). A/B partitioning is thus governed by a toggle-switch scheme as described for many metastable 2-state systems, with mutual opposition dictating the existence of two states, and the strength of arrows and lines dictating the dynamics of alternation between states (Fig. 2). However, what causes chromatin states A or B to be found at a given location largely remains a conundrum.

Epigenomics has produced a wealth of data over the last 20 years depicting chromatin fiber composition and long-range interactions throughout the human genome. Epigenomics is unbiased as to the nature of the underlying sequences, however interpretation often turns out to be gene-centered, with more attention being paid to active chromatin states, associated with gene activity, than to inactive chromatin states that actually occupy most of the genome (25, 26). In addition, repeat sequences (RepSeqs) have largely been left out due to technical hurdles. In this article, we re-analyze with fresh eyes a set of epigenomic data in an attempt to identify general rules and key players responsible for human genome organization. We first show that genome organization is largely inscribed in the DNA sequence, and set out to identify the DNA elements involved. We incriminate RepSeqs, providing strong evidence that they are responsible for the formation of a default chromatin landscape largely invariant between cell types, which predetermines some domains for compartments A or B, and dictates the propensity of other regions to shift between the A and B states. In these latter regions, DNase I hypersensitive sites (DHS), which are DNA elements that bind transcription factors and for a majority correspond to the so-called transcriptional "enhancers", hitherto considered to activate gene transcription only in a direct manner, appear in fact responsible for domain opening.

RESULTS

Genome organization is largely inscribed in the DNA sequence

Let's observe the HiC EV profile obtained with a 100 kb resolution along a representative 25 Mb region of chromosome 10 (Fig. 1A and Extended Data Fig. 1A). In Fig. 1A, the HiC EV is aligned for 8 ENCODE reference cell lines. HUVEC shown at the top is a primary cell "line" (cord blood endothelial cells, expanded in culture), and it will serve as our reference cell in this study. Most striking is the general conservation of the profile. In particular, one can distinguish regions that belong to the same compartment in all 8 cell lines and cover a majority of the 25 Mb region. These regions will be referred to as

"AlwaysA/AlwaysB" (Fig. 1A, bottom part). Conversely, "AorB" are regions that are either A or B depending on the cell line (Fig. 1A, bottom part), and of those, regions that are A (respectively B) in our reference line HUVEC will be referred to as AorB:A (respectively AorB:B). The left half of the chromosome segment shown in Fig. 1A is enriched in AlwaysA/AlwaysB regions whereas the right half is enriched in AorB regions (Extended Data Fig. 1B). In HUVEC, AlwaysA, AlwaysB, AorB:A, AorB:B regions cover each about 25% of the genome (Fig. 1C). Intriguingly, the absolute values of the HiC EV appears lower in the AorB regions when compared to the AlwaysA/AlwaysB regions where it is fairly uniformly high (Fig. 1A and E).

In order to better take into account genome heterogeneity, we further partitioned the hg19 version of the human genome into 24 subclasses of 100 kb bins according to gene density and whether histone marks typically found at active promoters are detected in the bin ("active" or "inactive" bins) (Fig. 1C; Extended Data Fig. 1B; see Resources & Methods section). As anticipated, a majority of active bins are found in A subclasses (orange bars), whereas a majority of gene-empty bins are found in B subclasses. The counts for the latter would be even higher if one used more recent versions of the human genome than hg19, which contains limited portions of the pericentromeric sequences. That some gene-empty, mostly inactive bins harbor active promoter marks is due to the fact that such marks may be found outside promoters, at potent enhancers in particular. In the following, we will mainly focus on 13 major subclasses, each represented by more than 400 bins (Fig. 1C). The case of B/AlwaysB bins featuring active promoter marks and/or a high gene density, and which largely overlap with so-called "gene complexes", will be treated separately. Approximately half of the genes in the human genome are contained within AorB regions (Extended Data Fig. 1C). AorB regions are A or B depending on the tissue type or developmental stage, or on alterations associated with immortalization or carcinogenesis for a particular cell line, and are accordingly enriched in genes involved in evolutionarily specialized but diverse functions, by opposition to housekeeping genes which are more enriched in AlwaysA regions, as revealed with a Gene Ontology (GO) analysis (Extended Data Fig. 2). AorB bins are found in approximately equal numbers in two situations: (i) at junctions between A and B domains, where the HiC EV curve crosses the value 0 (Fig. 1A and C, pink curve); (ii) in relatively small domains that contain mostly AorB bins (Fig. 1A, compare right and left parts of the 25 Mb region; Fig. 1C, pink curve; Fig. 1D). Thus, whereas AorB regions found at A/B transitions may rightfully appear as the place where A and B influences from adjacent regions confront each other, these observations suggest that AorB regions also represent a specific habitat, favorable for genes whose expression must be strongly regulated, as independently supported by a previous report (27).

To be more nuanced than just mentioning AorB, we created an indicator referred to as "AorB vector" (AorBvec) scoring the A or B status for each 100 kb bin in the 8 cell lines (Fig. 1B and E). The value of the AorB vector is +4 when the considered bin is A in all 8 lines, i.e. AlwaysA +3 when the considered bin is A in 7 lines and B in 1 line, and so forth down to -4 for AlwaysB. The AorB vector is therefore a sort of probability function that a bin is A or B in a given cell line, which could be rendered more accurate by including a larger number of cell lines. AorB regions therefore display AorBvec values between -3 and +3. The AorB vector profile is overall strikingly similar to the HiC EV profiles of the cell lines and in particular to our reference HUVEC HiC EV (Fig. 1A, B, and E), which has several implications. First, AorB regions, which alternate between A and B compartments between cell lines, largely coincide with regions displaying low absolute values of the HiC EV, i.e. low apparent contact preference between A and B compartments in a given cell line ("low A character" or "low B character"). AorB regions therefore display high chromatin "plasticity" both within an individual cell and between tissues. AlwaysA and AlwaysB regions instead contribute to the similarity of the HiC EV and AorBvec profiles by featuring high absolute values for both parameters, indicative of low plasticity. Second, an AorB:A domain in HUVEC is more likely to be AorB:A than AorB:B in another cell line, especially when it contains genes: there is an "A trend" - and conversely an AorB:B domain in HUVEC will most often be B in another lineage. This translates into identical signs for HiC EV and AorBvec (Fig. 1E) ("directionality"). In conclusion, the striking conservation of the HiC EV profiles between cell lines indicates that major determinants of HiC EV and altogether of A/B partitioning are to be found in the DNA sequence as suggested earlier (28, 29), determining both compartment directionality and the degree of plasticity.

The identification of ProA and ProB elements in the human genome points to a key role of repeat sequences in genome organization

Identifying cis-determinants of A/B partitioning: ProA and ProB elements

Since we showed that genome organization appears largely inscribed in the DNA sequence, we decided to identify sequence determinants of A/B compartmentalization. One may think essentially of two types of DNA features: GC content (%GC), the density of different types of repeat sequences (referred to as RepSeqs in the following), and genes, which can be viewed as a particular type of repeat sequences. It is in fact already known that the genome partitions into domains of varying %GC that

closely coincide with the A and B domains (see Fig. 4A), with a higher %GC found in the A compartment. Therefore, %GC is a good predictor of A/B partitioning (30). Strikingly, high %AT appears to promote the spreading of the recently described heterochromatin in prokaryotes (31), and there is evidence suggesting that this may also be true in eukaryotes (32-34). Thus, one reason for the presence of high %AT, and hence low %GC, in the B compartment might be because it may promote the binding of structural heterochromatin proteins to DNA or nucleosomes by mechanisms which await characterization. Notably, DNA binding is regarded as a preliminary step before structural heterochromatin proteins bind nucleosomes (35, 36). AT-binding factors, both structural proteins and RNA as well as remodeler machineries may also be involved (37, 38).

Considering this background knowledge about %GC, we will focus on "interspersed" repeats in the following to try to identify cis-determinants of A/B partitioning. In other words, we aim here at identifying what will be called in the following ProA and ProB DNA elements, defined as elements that locally promote the formation of A and B compartments, respectively. We reasoned that the density in individual determinants of HiC EV should correlate (or anti-correlate) with the HiC EV value, and conversely, knowing such densities should allow one to predict the HiC EV profile (Fig. 2). We therefore first assessed the copy count in 25 kb bins for each of the 1395 RepSeq subfamilies of the RepBase library, and classified these subfamilies according to their degree of Spearman correlation with the HiC EV in HUVEC (Fig. 3A; Table 1; see Resources & Methods section). We arbitrarily retained families scoring above 0.01 and below -0.01, for which the calculated correlation value is highly significant ($P < 0.01$), and called these two sets CorrA RepSeq and CorrB RepSeq, containing 250 and 367 subfamilies, respectively (Fig. 3A; Table 1). As expected, considering the procedure by which CorrA and CorrB RepSeq sets were defined, their density profiles approximately follow the HiC EV profile (see below). For each RepBase subfamily, we further assessed the correlation with the HUVEC ChIP signals for H3K9me3 and H3K27me3, which are typifying histone marks for the HP1α -based and Polycomb-based heterochromatin systems, respectively (Table 1). We also assessed the enrichment in A or B compartments, and in AlwaysA or AlwaysB regions (Table 1). A UMAP analysis of the RepSeq subfamilies using these parameters as well as copy count and estimated evolutionary age shows a butterfly image with trends expected for the B compartment (such as high correlation with H3K9me3 and low correlation with HiC EV) in the left wing, and conversely, trends expected for the A compartment in the right wing (Fig. 3B). A major part of RepSeq subfamilies building the left wing were independently shown to be enriched in genome fractions biochemically characterized as being embedded in heterochromatin (39) (Fig. 3B). Thus, RepSeq subfamilies can be neatly partitioned into those associated with A and those associated with B. The fact that the CorrA and CorrB RepSeq sets each form a distinct cluster on the outer edges of the cognate wings confirms a strong association (Fig. 3B). Correlation with HiC EV is largely conserved between cell lines for RepSeq for values below -0.01 or above 0.01 such that CorrA and CorrB sets would only slightly differ had a different cell line been selected as a reference (Extended Data Fig. 3; Table 2).

CorrA RepSeq subfamilies primarily consist of transposable elements (TEs) covering more than two thirds of the TE inserts in the genome, and contain the vast majority of short interspersed elements (SINE) subfamilies (Fig. 3C). CorrB TE RepSeq subfamilies mainly consist endogenous retroviruses (ERVs) and long interspersed elements (LINE) L1 (L1) (Fig. 3C). Together, CorrA RepSeq and CorrB RepSeq contain 90% of the 3.4 million TE inserts in the human genome, covering 37.6% of the hg19 length. The TE subfamilies absent from both lists (and therefore more evenly distributed in the genome) consist in large part of relatively young ERV subfamilies with low insert counts ("BetW", Fig. 3C-D; Extended Data Fig. 4 and 5; Table 1). The CorrB RepSeq set further contain simple repeats (also called "microsatellites", or "STR, for short tandem repeats"), and low complexity sequences that represent about one third of the CorrB RepSeq copies (Fig. 3A; Extended Data Fig. 5). A salient feature of the CorrB non-TE RepSeqs is their predicted propensity to adopt a non-B DNA structure: (i) CorrB simple repeat sequences consist in alternating purine and pyrimidine, with only C (or G) on one strand when not made only of A and T, which has a predicted capacity to adopt a Z-DNA conformation (40, 41); (ii) such A-rich and AT-rich low complexity sequences can form triplexes with RNAs via Hoogsteen pairing, in particular with the lncRNA KCNQ1OT1 (42, 43) (Table 1). In addition, simple repeats are commonly transcribed and high GC skewing of CorrB non-TE RepSeq suggests that they are prone to forming R-loops upon transcription (44-46). Simple repeats are known to specifically bind a host of TFs (47). Both Z-DNA and R-loop forming potential has also been demonstrated for TEs, especially SINEs and ERVs (46, 48-50). Finally, satellite sequences building up constitutive heterochromatin at (peri)centromeres and telomeres, which by definition fall into AlwaysB and are poorly represented in hg19, are also part of CorrB RepSeq (Extended Data Fig. 5). R-loop formation is a conserved feature of pericentromeric satellites, and thought to serve a physiological, regulatory role in promoting specific DNA transactions (50-53).

Knowing RepSeq composition and density suffices to accurately predict HiC EV

As anticipated based on our definition of CorrA and CorrB elements, using CorrA and CorrB RepSeq sets together with gene density in a linear model approach gives a very good score for predicting the HiC EV, the predicted profile most closely resembling that of AorBvec (Fig. 4B, line 12). This score is better than that obtained using %GC as the only parameter (Fig. 4B, line 15), suggesting that RepSeqs have a contribution to the prediction beyond their GC content. Furthermore, this score is in the same range as the ones based on chromatin features known to be associated with gene activity and highly enriched in the A compartment (Fig. 4A; Extended Data Fig. 1A). This is the case in particular for the count of DNase I hypersensitive sites (DHS), which has a very strong predictive power as already reported (54, 55) (Fig. 4B, line 11), or the GROseq signal, which reflects gene transcription activity itself (Fig. 4B, line 5). As also expected, in the linear models, the multiplicative coefficients associated with the CorrA RepSeq density, gene density and chromatin features associated with gene activity are positive, while the coefficients are negative for the CorrB RepSeq counts, which can be interpreted as the former seemingly promoting A character whereas the latter would seemingly promote B character. Furthermore, combining chromatin features with DNA features improves the prediction score. Intriguingly, in this case the features associated with gene promoter activity (H3K4me3, ChromHMM strong Promoter) contribute to the models with negative scores (Fig. 4B, line 1-2), which suggests that while DHS and enhancers are clearly ProA features, gene promoters in an active state behave to some extent as ProB features. In conclusion, it appears that the distribution of most RepSeq subfamilies in the human genome is sufficiently heterogeneous with respect to compartmentalization such that one can define two sets of RepSeq on the basis of correlation with the HiC EV, together covering roughly 90% of RepSeq inserts in the genome, and predict the organization of the genome into A and B compartments just by taking this into account.

Identifying core sets of ProA and ProB elements and mechanisms involved

Can it be claimed however that CorrA and CorrB RepSeq specifically promote A and B compartments, respectively? ProA and ProB elements are expected not only to locally promote the cognate character in the cognate compartment, but also to antagonize the opposite character according to the toggle-switch scheme (Fig. 2), and therefore to be depleted in the opposite compartment. In other words, ProA and ProB elements should be significantly enriched in the A and B compartments, respectively, which is clearly not the case for all CorrA and CorrB RepSeq subfamilies (Fig. 3D). We therefore decided to apply an additional filter by retaining only those RepSeq CorrA (or CorrB) that show enrichment above 1.17 in the A compartment (or B compartment) (Fig. 3D; Extended Data Fig. 5). The density of the corresponding RepSeq CorrA.enrichA and CorrB.enrichB subfamilies more closely follow the HiC EV curve than observed for the complete CorrA and CorrB sets (Fig. 4A). These two "Corr.enrich" RepSeq sets now clearly fulfill expectations for sequences that promote A and B character, respectively. This does not preclude that additional RepSeq subfamilies may have a role in genome organization. The following two putative RepSeq profiles would clearly evade our correlation/enrichment-based selection: (i) "Compartment mitigators", such as RepSeqs that would have some specialization in promoting B character within an A context, and vice-versa; (ii) "Chameleons" that would tend to behave as ProA in an A context and as ProB in a B context. In addition, some RepSeq subfamilies may play a role through only some of their members without reaching our thresholds when analyzed as a group. The CorrA.enrichA and CorrB.enrichB RepSeq should therefore be understood as archetypal ProA and ProB RepSeq that display both pure and pronounced ProA and ProB character, respectively. These are clearly the types of profiles that should be considered in further endeavours to try to understand the molecular mechanisms by which RepSeqs are involved in genome compartmentalization. CorrA.enrichA and CorrB.enrichB TEs represent 1.3 and 0.3 million inserts in total, each covering approximately 12% of the hg19 genome length, respectively, and are essentially composed of Alu and SVA sequences for ProA and of ERV and L1 elements for ProB (Fig. 3C; Extended Data Fig. 4 and 5). By comparison with CorrA and CorrB sets, both CorrA.enrichA and CorrB.enrichB RepSeq sets are slimmed down with respect to relatively old subfamilies, and are therefore correspondingly enriched in relatively young subfamilies (Fig. 3C; Extended Data Fig. 5A). This is consistent with a model in which ProA and ProB elements may promote the A and B compartments, respectively, via their cis-regulatory elements, which degenerate and loose functionality with age (Box1). Thus, similarly to the cis-regulatory elements of genes known as promoters and enhancers, RepSeq harbor cis-regulatory elements, which bind both activators and repressors, locally anchoring transcriptional activating and repressing systems, respectively (56-64). Repression of young, potentially active TEs appears especially crucial to maintain genome stability by inhibiting both transcription and recombination, which chiefly involves HP1α /H3K9me3-based heterochromatin in somatic cells (4, 56, 65) (Box2). As the most prominent chromatin marker in the B compartment is precisely the H3K9me3 mark (Fig. 4A), this striking coincidence suggests a capacity for any newly inserted TEs to act as a ProB element, without making any assumptions about the exact underlying mechanism. While TE cis-regulatory sequences decay and lose functionality as the TE sequence overall drifts with evolutionary time, heterochromatin marks may still be observed long after a TE has seemingly lost any transcription potential (56, 66) (Box1).

L1 inserts meet these expectations as the youngest L1 subfamilies are enriched among the RepSeq scoring best as ProB, with a trend by which scores tend to diminish with age (Fig. 3D, Extended Data Fig. 5A). By contrast, the youngest Alu elements do not feature marked polarity, and only young Alu subfamilies that have already significantly drifted score as CorrA.enrichA (Fig. 3D, Extended Data Fig. 5A). This is consistent with a scenario in which new L1 inserts are born as strong ProB. New Alu inserts may instead feature both ProA- and ProB-type potential and progressively lose the latter through sequence drift, being eventually selected against in the B compartment and selected for in the A compartment as ProA, as previously suggested by others (67). While a majority of ERV subfamilies score on the B side, no such clear trends as for Alu and L1 are found for the ERV1 and HERVK subfamilies, with an abundance of young subfamilies displaying a mixed character, similar to young Alu elements (Extended Data Fig. 5A). The LTRs of young ERVs are GC-rich sequences (Extended Data Fig. 6), and can generally behave as *bona fide*, potent enhancers of gene expression, as demonstrated both in classical reverse genetics assays and by *in situ* inactivation of individual inserts, with subfamilies displaying distinctive tissue-specificity associated with the binding of tissue-specific transcription factors (59, 60, 68-72). Young ERVs inserts are therefore markedly bipotential, having strong enhancer capacity while being also prominent targets for heterochromatin-mediated repression (60, 65, 73). Correspondingly, a very large number of ERV LTRs are found associated with either an active or a repressed chromatin state in various tissues (26). This suggests a role for ERVs as compartment mitigators or chameleons, with the large number of ERV subfamilies being suggestive of specialization in specific tissues, stage of development, or signaling pathways. The idiosyncratic evolution of individual TE inserts belonging to other classes, in particular L1 and Alu, may allow the emergence of elements with high enhancer activity and therefore similar bipotentialities as the LTR of ERV inserts, as indeed observed (62, 63, 72, 74).

In conclusion, core sets of TEs featuring more pronounced ProA or more pronounced ProB character are enriched in and shape the cognate compartment. Linear models suggest that molecular mechanisms responsible for active and repressed chromatin states can underlie the ProA and ProB functions of active gene cis-regulatory elements and young TEs, respectively. How Alu elements perform their ProA function stands as a riddle at this stage, and will be addressed in the Discussion. Notably, to a degree, active and repressed chromatin states potentially confront each other at every TE in a dynamic manner, which may be captured in specific settings (73). Therefore, we propose that most if not all TEs may be involved in shaping genome organization through a combination of ProA and ProB capacity.

Compositional genomics reveals general trends for the local densities of ProA and ProB elements, and a marked heterogeneity between functionally similar territories

By way of extension, CorrA.enrichA and CorrB.enrichB sets will be referred to as ProA and ProB RepSeq sets in the following. Compositional analysis of the human genome shows a trend by which the density of both TE and non-TE RepSeq ProB gradually rises from high A character subclasses to high B character subclasses, and the reverse is true for RepSeq ProA and chromatin features associated with gene activity, in particular DHSs, with the AorB subclasses displaying intermediate profiles (Fig. 5A; see Extended Data Fig. 8A for all 24 subclasses). This general pattern is remarkably congruent with the expectation outlined in Fig. 2. In addition, there is a direct relationship between the enrichment value of a ProA RepSeq subfamily in AlwaysA and in A in general, and for a ProB subfamily in AlwaysB and in B (Fig. 5B). It appears therefore that the marked functional character of Always regions with respect to compartmentalization coincides with a marked character in ProA and ProB RepSeq composition. The absolute values of ProA TE counts along these trend lines is much higher than that of ProB RepSeq, such that to facilitate comparisons we will divide ProA TE counts by 3 (referred to as ProA* count) in subsequent analyses and representations. Thus, using this arbitrary gimmick to calculate a ProA/ProB RepSeq ratio, the ProA "weight" appears, on average, four times higher than the ProB weight in AlwaysA subclasses with the highest A character compared with AlwaysB subclasses with the highest B character, and the mirror image is true for bins with high B character (Extended Data Fig. 8A). Interestingly, again using ProA* and ProB counts, the total RepSeq count appears relatively constant in all subclasses, on average (Extended Data Fig. 8A, middle panel). That both absolute and relative ProA and ProB densities appear to obey such general trends is consistent with a model in which the ProA and ProB elements are organized into two systems in which functionally interchangeable elements cooperate at a distance, all the more efficiently when they are closer, and which ultimately obey, as systems, the toggle-switch rule shown in Fig. 2. Cooperation between ProB elements may occur according to principles explained before (17), building a heterochromatin network which conceivably represents a major organizational force, while ProA elements are envisioned to cooperate in antagonizing the latter. Gene complexes mainly fall into B chromatin subclasses 5-8 and 13-16 and display a strikingly chimeric composition, harboring high densities of both ProA and ProB RepSeq (Extended Data Fig. 8A).
It is important to note that these are only general trends and that a wide range of values is observed for each feature considered within each subclass, indicating a marked heterogeneity between bins <u>within</u> a subclass and prominent overlap

between subclasses (Fig. 5; Extended Data Fig. 8C). A similar picture is observed when %GC classes are analyzed (Extended Data Fig. 8B).

Is such a heterogeneity compatible with the notion of RepSeqs as key players in genome organization, or may it even have a specific role? In the following analysis (Fig. 5C), multiple features are assessed in AorB bins that are outliers in HUVEC compared to the seven other reference cell lines (i.e. HUVEC.A(vecB) bins which score A in HUVEC but display a B trend, and vice-versa for HUVEC.B(vecA) bins). Only AorB bins which are "outstanding outliers" in HUVEC, i.e. displaying a large difference between HiC EV and AorBvec values, are considered (see Resources and Methods section). Such HUVEC.A(vecB) and HUVEC.B(vecA) regions feature especially low and high levels of Polycomb facultative chromatin-associated markers, respectively, but otherwise display the expected average levels of chromatin markers associated with gene activity for A and B bins (Fig. 5C). However, they tend to display a remarkable, chimeric composition with regards to RepSeq. Thus, the %GC and RepSeq composition of HUVEC.A(vecB) globally follow the trend of the corresponding HUVEC.B subclasses, consistent with them belonging to the B compartment in a majority of cell lines. The exception are CorrB RepSeq, with CorrB.TE density tending to approach that found in the corresponding A subclasses and CorrB non-TE density being especially high, in the range found in AlwaysB regions (Fig. 5C). The mirror image is true for HUVEC.B(vecA) regions, which largely display a typical A composition with especially high densities in CorrA RepSeq while featuring at the same time CorrB RepSeq densities only slightly lower on average than the corresponding HUVEC B subclasses (Fig. 5C). Altogether, it therefore appears that bins that are outliers in HUVEC with respect to compartmentalization, i.e. that score B in HUVEC but A in a majority of the seven other cell lines or vice-versa, also appear as outliers with respect to the composition in RepSeq. Thus, the widely dispersed RepSeq composition observed within each subclass may conceivably have a role in allowing the emergence of a variety of specific regulatory subfunctionalities.

Alternatively, such a heterogeneity may have minimal relevance to gene regulation and instead reflect fundamental physical principles at work in shaping genome organization. Thus, the RepSeq and gene density-based linear models prove powerful in reproducing the general appearance of the HiC EV curve, but the curves of the models are more jagged than the HiC EV curve, which is remarkably smooth (Fig. 4B). This observation is consistent with the notion of long-distance cooperation between elements within the ProA and ProB subsystems, well beyond the 100 kb binning length that we have used. Moreover, genome organization is known to obey fractal principles, with self-similar patterns of domains recurring across scales as inferred in particular from HiC analysis (1, 5, 75, 76). Observing a large dispersion for elementary parameters determining the overall system therefore does not come as a surprise, but is actually expected (77, 78).

In conclusion, the local dosage in ProA and ProB DNA features appears to coarsely dictate the propensity of a chromatin region to promote the installation of either the A or the B compartment, or instead to confer plasticity. This should not be regarded as a strictly deterministic principle. Instead, the considerable heterogeneity observed in ProA and ProB RepSeq densities can be seen as a sign of the fractal organization of genomes, in which ProA and ProB elements are envisionned to interplay in multiple dimensions in nuclear space, and in an eminently dynamic manner.

DNase I Hypersentitive Sites oppose B compartment to open domains

DHS density emerges as the strongest single ProA chromatin feature. DHSs correspond to DNA segments actively freed from nucleosomes by the interplay of transcription factors that bind to them and recruit chromatin remodelers, which in turn promotes nucleosome sliding or eviction to both induce and sustain DHS existence (79-84). DHSs are largely tissue-specific, and approximately two thirds of DHSs in a differentiated somatic cell coincide with active gene enhancers (85). Analyzing a large panel of tissues and cell lines recently established a compendium of 3.6 million DNA elements in the human genome actuated as DHS in at least one setting (85), amounting to approximately 385,000 "active DHSs" just for cell types found in the human brain cortex (74). The number of these elements is strikingly in the same range as the total RepSeq count of chromosome arms, as discussed above. As expected for a feature associated with gene activity, DHS density is on average systematically lower in a chromatin subclass where bins display no active promoter marks, as compared with the subclass with similar gene density and chromatin color but displaying activity (Fig. 6A), and it is generally at a background level in the B compartment except for a minority of regions displaying active marks (Fig. 6A), which are mostly found at A/B junctions. In active high gene density regions, DHS density tends to increase with size, albeit only mildly (Fig. 6B). In a seemingly paradoxical twist, there is a tendency whereby, in some A subclasses, smaller domains will display a higher density of DHSs. This tendency reaches statistical significance for inactive A bins containing few or no genes (Fig. 6B, subclasses 12 and 20). In addition, both DHS and active enhancer densities are especially high in gene-containing bins that score A in HUVEC but B in a majority of the 7 other cell lines analyzed (Fig. 5C, compare HUVEC.A(vecB) bins and HUVEC.A bins). Altogether these

observations suggest that the factors and events associated with a DHS constitute an opposing force that enables the formation of an A domain within a "hostile" environment.

Strong support for this notion is to be found during development, or by experimental manipulation, in which the shift of domains from a B/closed to an A/open configuration, and relocation away from a heterochromatin-dense compartment have been reported to coincide with the appearance of DHSs scattered across the domain, in the absence of any gene transcription (3, 86-91) (illustrated Fig. 6C). Some of these DHSs will convert to bona fide enhancers, associated with typical markers such as high levels of H3K27ac and RNA (eRNA) production, and possibly contact with a target gene promoter, but this occurs only in a second step, at the same time as the activity of the cognate gene takes off (3, 87, 89, 92-96). Interestingly, some active enhancers seem not to display elevated contact frequency with any gene, in a given cell type, albeit displaying high enrichment of active enhancer features (72). A strikingly similar sequence of events may also be observed on a shorter time scale, for instance for domains that switch between A and B compartments over a circadian cycle (97-99), or upon neuron stimulation and memory encoding (100). Thus, there is strong evidence that DHS/enhancers mediate domain opening, enabling promoter activation, but also recombination and initiation of DNA replication (101, 102), and conceivably any process that is inhibited by heterochromatin in the B compartment, in an indirect and therefore non-specific manner. This function of enhancers may be described as "generalized activation", in reference to the "generalized repression" observed in the B compartment, and which is a hallmark of heterochromatin.

DHS/enhancers cooperate over a distance in opposing installation of a B chromatin system, as may be inferred from the co-actuation of clustered DHSs observed at a single-molecule level (103), and as also proposed above for ProA and ProB RepSeq. DHS/enhancers therefore appear to behave as inducible ProA features, acting to open domains by cooperatively opposing the forces that drive the B compartment, and being presumably continuously required although not continuously bound by TFs (103). It is likely that this "anti-silencing" role is expanded when a basic DHS turns into a classical enhancer. In the Discussion, we will propose a mechanism for how ProA elements oppose establishment and maintenance of the B compartment that integrates our findings presented here and what is already known from the literature.

DHS/enhancers are of RepSeq origin, switch from a ProB to a ProA state upon induction, and can globally unfold the genome

DHS/enhancers are of RepSeq origin

A growing body of literature shows that, in all tissues where this has been investigated, a significant proportion of DHS/enhancers active in a given cell type are in fact RepSeqs (26, 104-107). Mounting evidence further suggests that all DHS/enhancers are of RepSeq origin, deriving either from TE or non-TE sequences, which may essentially have become undetectable due to sequence drift across evolutionary times (see Box1) (47, 74, 108-110).

DHS/enhancers are dual-function elements that switch from a ProB to a ProA state upon induction

In cases where the DHS/enhancer considered is a Very Young TE, typically a human- or primate-specific TE, actuation of the DHS/enhancer in the course of development or differentiation actually consists in a clear switch from an initial, H3K9me3-associated ProB chromatin state, to an active DHS/enhancer chromatin state. Interestingly, integrating switches from H3K9me3 to active DHS/enhancer state on a genome-wide scale enables a compelling prediction of cell trajectories (90). However, most DHS/enhancer elements exhibit a mundane, seemingly neutral chomatin state prior to induction, blending into the landscape (56). We surmise that a large fraction of enhancers nonetheless perform a ProB function prior to induction. There are at least three possible explanations for why this is not generally apparent in the form of bound repressor TFs and repressive marks: (i) There is a conspicuous background of H3K9me3 signal across the whole genome, visible in both bulk ChIP and single-cell analyses, which would clearly mask a small peak; (ii) Highly transient binding of TFs such as Krüppel-associated box (KRAB) zinc-finger proteins (KZFPs) may evade detection by ChIP while being sufficient to maintain heterochromatin assembly over a region via discrete events of nucleation, due to memory effects within the heterochromatin system in a differentiated cell (111) (Box2). Consistent with this notion, even prototypical ProB TE most often do not appear as DHSs in spite of featuring both H3K9me3 and KZFP ChIP signals, indicative of transient TF binding. (iii) All types of heterochromatin require histones to be deacetylated for spreading. Targeted histone deacetylation can locally promote heterochromatin binding to the chromatin fiber without the need for a dedicated heterochromatin recruitment apparatus, as observed for SIRc-based heterochromatin in the budding yeast *Saccharomyces cerevisiae* (*S. cerevisiae*) (112). The resulting heterochromatin binding is more dynamic and therefore essentially not apparent using ChIP. Such a principle may account for the ProB capacity of a number of TEs, but also of gene promoters (Fig. 4B), as promoters toggle between high

and low acetylation states, coinciding with bursts of effective transcription initiation by RNA polymerase II (Pol.II) and inactivity, respectively.

Altogether, ProA and ProB RepSeqs and DHS/enhancers must therefore be considered as a single system of elements, wherein, in a first approximation, two core sets of RepSeqs with a marked ProA and a marked ProB character, namely Alu/SVA and L1/ERV elements, respectively, perform their function constitutively, while elements known as DHS/enhancers are RepSeqs or derivatives that can exist either in a ProA or a ProB state depending on the cell context. Independent support for the ProA/ProB bipotentiality of DHS/enhancers and for their RepSeq origin comes from the discovery that KZFPs are largely responsible for interindividual epigenetic variability both in mice and humans (73, 113-115). Thus, KZFPs, which are pivotal TF effectors of RepSeq repression and therefore of the ProB function in general (Box2), were also found to be prime trans-determinants of DHS/enhancer activity (113, 114). This may now be interpreted as reflecting a negative modulation of the DHS/enhancer ProA function by the associated ProB character, seemingly tuning ProA function (Fig. 7A) (113).

A scenario for domain opening that incorporates alteration of CpG methylation

The human genome of a healthy cell displays a homogeneous roughly 80% level of cytosine methylation at CpG dinucleotides. A highly localized loss of meCpG is observed along with the emergence of a DHS/enhancer, which appears as a cluster of differentially methylated CpG (DMR). This is due to the action of ten-eleven tranlocation methylcytosine dioxygenases (TET) enzymes, as recruited by bound TFs (116). However, the two processes are not strictly coupled. While loss of meCpG may precede the onset of DHS when a pioneer TF, such as Klf4, is involved with its ability to bind chromatin in isolation and to recruit TET enzymes, it may also occur with some delay after the DHS is detected (116, 117). Strikingly, in a given tissue such as the human motor cortex, approximately ten times more DMRs than DHS/enhancers are detected, which are differentially methylated or differentially open between cell types, respectively (74). Similarly to DHS/enhancers, a large proportion of DMRs coincide with RepSeqs, belonging to all TE classes, but the proportion is even greater for DMRs, and overall DMRs are less cell-type specific than DHS/enhancers except for those where DMR and DHS/enhancer coincide (74). CpG methylation is involved in TE repression (Box2). It is a pivotal player for the repression of young L1s and ERVs, which feature relatively high levels of %GC, of %CpG, and of CpG methylation in a normal tissue (Extended Data Fig. 6; Extended Data Fig. 9), and which comprise precisely the TEs with the strongest ProB character in our analysis.

In view of the above, it seems reasonable to assume that an alteration in the meCpG level of a DNA element without a DHS being concomitantly detected represents a modulation of its ProB function, whether or not the element is clearly RepSeq-related, and whether or not it has a remarkable histone mark profile. Transition from a ProB status to a ProA status may thus be conceived as a continuum, of which demethylation/ DMR may be regarded as an intermediate stage, according to the following scenario. We envision loss of DNA methylation and eviction of a nucleosome, as driven by TFs and associated cofactors binding the element, as highly dynamic processes that are largely compensated by reverse reactions in an initial stage, with loss of methylation/DMR status becoming gradually detectable. A minor subset of these elements in a B chromatin segment would clearly emerge as DHS, upon the pivotal binding of some additional TFs, only when a threshold density of such elements is reached in a domain. This would indeed switch the domain to an A configuration and lead in in turn to the all-or-none co-actuation of these DHS by cooperative effects, as experimentally supported by a recent publication (103) (Fig. 6C). The other DMRs, which are much more abundant than DHSs, would instead remain as DMRs, possibly progressing to a clearer demethylated stage without ever reaching a detectable DHS stage. They may thus contribute to domain opening in a more subtle way than DHSs, by virtue of modulating the ProB function of elements without a clear switch to ProA. Such a model allows, in particular, to understand how continuous phenomena, such as the alteration of the concentration of specific TFs, can be responsible for domain switching from a B to an A state, which appears essentially binary. This is due to the fact that domain opening and DHSs mutually reinforce each other, concurring to the apparent self-sustainment of the A state at the domain scale (Fig. 7A). Whether ProB RepSeqs whose function is simply modulated without switching to an overt ProA DHS or enhancer state play an important role in this picture remains to be investigated.

Viruses drive global genome opening by switching or tuning selected sets of DHS/enhancers towards ProA

Viruses massively reprogram cells, which has been the subject of intense research over the last fifty years. We decided to re-examine this phenomenon for selected viruses to tentatively gain further insights into general principles of genome organization. Infection of cells by viruses has been reported to result in sets of TE inserts, most commonly ERV LTRs, switching in a sustained manner over time from a repressed chromatin state to an active enhancer or active promoter state; this is associated with loss of DNA methylation, and for a subset of them with active transcription (118, 119). This is the case

in particular for Epstein-Barr Virus (EBV) infecting resting B cells or epithelial cells (119-122), where a large number of ERV subfamilies of all ages have members that are induced or over-activated, although only a minority of inserts within a subfamily are involved (Extended Data Fig. 7B). These induced LTRs are indistinguishable from conventional host cell enhancers except that they result from the cooperation of tissue-specific TFs, expressed by the host cell, and TFs encoded by the virus ("viral transactivators"). Induced LTRs may also interact *in trans* with non-integrated EBV episomes (121, 123). Such neo-enhancers appear to switch entire domains from B to A (121, 122), similarly to cellular enhancers as described above. Infection of B cells by EBV triggers a transformation by which B cells exit quiescence and reach a state akin to that of activated and self-renewing B cells, exhibiting a globally unfolded genome in a large nucleus (122, 124-126). The TFs NF-KB and Myc, as well as the MAPK signaling pathway, have been implicated in both situations, Myc being directly upregulated by EBV transactivators (127). A plausible explanation for such a global genome unfolding is that shifting a large number of RepSeqs along the genome to a DHS/enhancer state can globally tip the equilibrium towards A according to the genome partitioning toggle-switch scheme (Fig. 7A), opening up the genome due to the loss of cooperative effects in the ProB system of elements, at all scales. It has previously been experimentally demonstrated using the mouse early development embryo as a model system (128) that switching a limited number of RepSeqs along the genome from a ProB to a ProA state, with or without associated transcriptional induction, can indeed globally unfold the genome. Strengthening the ProB state of the same RepSeqs had the opposite effect, compacting the genome (128), strongly supporting the notion that tuning the ProA / ProB functions of RepSeqs is the common way by which cells adjust genome folding at all scales to their needs.

Global opening of the host cell genome has been observed upon infection by many viruses, or after reactivation of viruses from latency. It is associated with an exit from cell quiescence, which appears necessary for viral multiplication (122). Viruses use a variety of strategies to eventually tip the toggle-switch towards A. Small DNA tumor viruses such as simian virus 40 (SV40) and polyomavirus, but also papillomaviruses, and Hepatitis B and C viruses (HBV, HCV) acutely push p53 towards degradation (129), p53 being a pivotal TF for the repression of RepSeqs (130-133). Notably, inactivation of p53 also suppresses the p53-driven stress response (129). Other viruses such as HTLV-I and HBV express "promiscuous transactivators" that empower specific TFs and chromatin modifiers typically acting at active promoters and enhancers, respectively (134-137). Yet other viruses, such as influenza A virus (IAV), inhibit transcription termination, and ensuing readthrough transcription over extended chromatin segments exerts itself the ProA antisilencing function, switching entire domains from B to A (138). In addition, many viruses antagonize heterochromatin because it acts primarily as a "restriction factor", silencing invading nucleic acids, and antagonizing at the same time the function of ProB elements. Viruses may trigger the degradation of heterochromatin components, such as Smc5/6, which is targeted by HBV, or components of PML bodies, which provide a favorable environment for heterochromatin formation (139, 140). This can, for example, be PML itself or SP100, which are targeted by HSV-1 (141), or ATRX/DAXX, targeted by adenovirus (139).

RepSeq ProB function is impaired in cancer by both specific and global mechanisms, resulting in genome opening and plasticity

The chromatin of a cancer cell is generally less compact and more homogeneous than that of a healthy, differentiated cell, and compartments are more intermingled, similar to the situation in a progenitor or embryonic stem cell (13). These observations are typically indicative of global genome unfolding, which leads us to suggest that limited RepSeq sets may switch their function from ProB to ProA, and therefore shift domains from the B to the A compartments, in cancer, similarly to what occurs during viral infection. Such RepSeqs would eventually be differentially enriched in the A compartment. In order to identify such RepSeqs, we compared RepSeq enrichment in the A and B compartments in the liver cancer cell line HepG2 and in the primary cell line HUVEC. While enrichment of L1 and Alu elements, as well as old ERVs is similar in the B compartment between the two cell lines, a large number of young and middle-aged ERVs tend to be more enriched in the A compartment in HepG2 cells, which is exactly what one would expect for drivers of compartment switching and, ultimately, global genome unfolding (Fig. 7E).

It was independently shown that a large number of ERV LTRs, together with some L1s, are transcriptionally induced in liver cancer, both in human and in mouse. This appears to result from transactivation by Myc in cooperation with liver-specific TFs and in particular with the pioneer TF FOXA3 (57, 72, 142). Myc is commonly upregulated in liver cancer, in both humans and animal models, and considered to be the main driver of hepatocarcinogenesis. Its overexpression in mouse liver potently induces liver tumors (142-147). The upregulation of ERVs is more pronounced in liver tumors of viral etiology, associated with chronic infection by either HBV or HCV, and inflammation (142). Myc-driven hepatocarcinogenesis may be further empowered by p53 haploinsufficiency, a highly prevalent genetic alteration in all cancer types (148). These findings are

consistent with a model in which Myc upregulation, due to both genetic and epigenetic alterations in the course of liver carcinogenesis (145), switches or tunes a large number of RepSeqs, and in particular ERVs, toward a more ProA function, bringing about global genome unfolding. p53 participates in the repression of a host of RepSeqs as discussed above, and therefore the loss of p53 may facilitate the ProB to ProA switch by weakening the ProB state. This model also provides a rationale for the well-described but hitherto not understood genome-unfolding ability of Myc (149, 150). Presumably different RepSeq sets and Myc partner TFs are involved depending on cellular context. Interestingly, iPS reprogramming is promoted by both Myc overexpression and p53 inactivation, and also by drugs that weaken HP1α /H3K9me3-based heterochromatin, suggesting a mechanism similar to that outlined above for cancer. Intriguingly, a number of RepSeq subfamilies are known to be responsive to the Yamanaka factors (151-153).

As mentioned above, ProB elements that switch to ProA generally lose CpG methylation. We therefore decided to assess the fraction of CpG dinucleotides methylated for each individual RepSeq copy in the genome ("meCpG fraction") in normal liver and in a liver tumor sample, using Nanopore sequencing (Fig. 7B. Strikingly, not only ERVs, but a majority of RepSeq subfamilies trending towards ProB character and appearing in the left wing of our UMAP analysis appear to lose DNA methylation in the tumor, while subfamilies in the ProA-trending right wing are largely spared or even gain methylation, in the tumor. Comparing RepSeq DNA methylation in liver tumor and adjacent non-tumorous tissue over the same 25 Mb region provides an explanation for this surprising finding (Fig. 7C). Our analysis was restricted to Alu and SVA elements, on the one hand, and ERV and L1 elements, on the other hand, as these are the main families contributing core ProA and ProB TE sets, respectively. Clearly, both Alu/SVA and L1/ERV TEs seem to globally lose meCpG in the B compartment, but not in the A compartment, in the tumor (Fig. 7C). Visual inspection of meCpG levels for 25 kb DNA segments leads to a similar conclusion (Fig. 7C). DNA methylation may actually occur at a higher level in the tumor than in the non-tumor sample in some A domains, such as in the 5 Mb AlwaysA domain on the left of the 25 Mb region (Fig. 7C), which had already been well established (154, 155).

Trend lines for the RepSeq meCpG fraction according to the HiC EV values of the 100 kb bins where an individual RepSeq copy is found show, first, that Alu/SVA TEs tend to display much higher methylation levels than L1/ERV TEs (Fig. 7D, top). Note that both types of TEs have a very different GC composition (Extended Data Fig. 6), and that CpG methylation plays a distinctive role (Box2). Second, both categories tend to lose DNA methylation to a similar extent in the B compartment while both display a similar moderate methylation loss in the A compartment in the tumor (Fig. 7D (top)). The main determinant for methylation loss of L1/ERV TEs in the tumor is clearly not the ProB character, nor the age of the subfamily, as may have been suspected, but instead the position of the insert in the genome, and more specifically the local chromatin flavour (Fig. 7D, middle; Extended Data Fig. 9 A and B). The reason why ProB TEs appear preferentially demethylated in the tumor is therefore that the B compartment globally loses methylation in cancer, and that ProBs are enriched in the B compartment to begin with. Centromeric and pericentromeric satellites were also found to become partially demethylated in the tumor sample (Extended Data Fig. 10). They are commonly transcribed in tumors and may display distinctively enhanced accessibility (156-161).

Although not the main determinant, the RepSeq age clearly impacts methylation, with younger TEs generally being richer in GC and CpG (Extended Data Fig. 6), and more methylated in normal liver (Extended Data Fig. 9 B-C). Interestingly, whereas Very Young L1s and ERVs are both more methylated than older counterparts in normal liver, and both lose methylation in B compartment in the tumor, this loss is more pronounced for L1s, presumably due to a lower GC content (Extended Data Fig. 9 B-C). Furthermore, whereas the L1/ERV curves are all approximately flat in Fig. 7D (middle), we noticed that, the Alu/SVA curves are all tilted, and all the more so from panel AlwaysA towards B up to panel AlwaysB. This shows that the youngest Alu subfamilies, which are enriched in B, are massively methylated in B, while the less young ones, which are gradually more enriched in A and are endowed with a strong ProA character (Fig. 3D), are partially demethylated when located in the B compartment. This is consistent with the somewhat counterintuitive notion that CpG methylation powers the ProA function of Alu/SVA and should therefore be thwarted in B such as not to jeopardize B compartment establishment (Box2). While the drop in methylation of A-enriched Alu/SVAs in B may seem modest, it is *a priori* sufficient to essentially abrogate the nucleosome positioning capacity of Alu (162), and thereby their ProA capacity as detailed in Box2. By contrast, the youngest Alu exhibit very high methylation levels in B, suggestive of a ProB function for CpG methylation in this context, which intriguingly may also involve nucleosome positioning, but combined in this case with a heterochromatin-nucleation capacity similar to that of L1/ERVs (43, 64, 66, 163). The latter would be lost by genetic drift, gradually revealing a ProA-only character, as mentioned before.

Collectively, these observations suggest a two-tiered dysfunction in cancer with respect to chromatin and genome organization. Firstly, the heterochromatin state of a large number of ProB RepSeqs is specifically weakened and some switch to a ProA function, locally shifting the balance between ProA and ProB power, unfolding domains, and eventually resulting in the global unfolding of the genome (Fig. 7A). This is similar to the aforementioned mode of genome reorganization upon viral infections, and may arise in this case as a consequence of combinations of abnormalities typically observed in cancer. Examples include constitutive activation of signaling pathways due to mutations, such as that in the Ras oncogene for the MAPK pathway, or autocrine mechanisms or inflammatory conditions, which promote an active chromatin state and therefore the ProA function of a large number of RepSeqs and known DHS/enhancers. In contrast, repressive mechanisms pivotal to ProB function are frequently dampened by genetic or epigenetic alteration in tumors. Highly prevalent, iconic examples include mutation or loss of: p53; KZFPs and TRIM28/KAP1 (164, 165); the ATRX/DAXX remodeler/chaperone pair responsible of the deposition and maintenance of H3.3K9me3 at RepSeqs (166, 167); the remodeler HELLS/LSH, which both enacts macroH2A deposition to reinforce RepSeqs repression and promotes their methylation by DNMTs (168-171). A classical, stepwise transformation protocol applied to human fibroblasts recapitulates rapid loss of CpG methylation at a subset of RepSeqs accompanied, with some delay, by RepSeq transcription (172). A second alteration, which seemingly superimposes on the first one, is a massive, systemic alteration at the compartment level, by which the B compartment, and by way of consequence ProB RepSeqs therein, globally lose DNA methylation, being therefore only partially methylated in tumors. By contrast, the A compartment may instead even gain DNA methylation in some domains, and display prominently active DHS/enhancers, including superenhancers, which altogether further shifts the balance between A and B compartments toward A (Fig. 7A) (173, 174).

Since CpG methylation is a key player in RepSeq repression (Box2) and therefore in ProB function, one may anticipate a collapse of the B compartment in cancer. However, this is not observed (Fig. 7C), even though compartmentalization is known to be weaker (175). The reason is that compensatory mechanisms arise to ensure a degree of RepSeq repression sufficient so as to maintain the integrity of the B compartment, while at the same time contributing to hiding the cell from the immune system by restraining RepSeq transcription (Box2). However, the resulting ProB RepSeq repression has neither the same efficiency nor the same quality as in a healthy cell, with three major consequences:
(i) First, irrespective of the cancer-specific mechanisms involved, a B chromatin segment harboring ProB RepSeq with altered functionality due to DNA methylation loss is anticipated to behave as a segment with ProB RepSeqs at a lower density. In other words, such a B segment would have a weakened B character, display more plasticity with respect to compartmentalization, as inferred before by comparing AlwaysB and AorB regions (see Figs. 1and 5, Extended Data Fig. 8). If this B segment contains genes, their expression would be less tightly controlled (176), which is consistent with the observed increased entropy in gene expression in cancer (175).
(ii) Second, since CpG methylation is specifically involved in guaranteeing a memory of repressed chromatin states both within a cell cycle and through multiple cell divisions (Box2) (177, 178), alterations thereof can only be poorly compensated by other mechanisms, in particular by the rise in H3K9me3 commonly observed in cancer (Box2). This suggests that the altered chromatin state at ProB RepSeqs, and more specifically the globally altered CpG methylation underlies the plasticity of chromatin states in cancer, and endows cancer cells with one of their most iconic hallmarks, the plasticity of cell fate (175, 179, 180).
(iii) RepSeqs are conspicuously transcribed in cancer (Box3), which enforces a marked ProA character, associated with a robust memory of the active chromatin state. This clearly plays an important role in B compartment decline while being at the same time promoted by it, thereby fostering a pathogenic feed-forward loop.

DISCUSSION

Eukaryotic genomes are universally partitioned into an euchromatin/A compartment and a heterochromatin/B compartment in the nucleus of differentiated cells, with euchromatin/A being conducive to gene transcription and heterochromatin/B largely non-permissive to transcription. The observations reported in this article, based on simple premises, suggest a model in which the core ProA and ProB repeat sequences (RepSeq) of a genome, defined as specifically promoting the A and B compartments, respectively, shape a basic A/B landscape, which as a first approximation can be considered invariant between different cells of an organism, and onto which genes settle. "Always" regions exhibit a strong propensity to be either A or B while "AorB" regions are more plastic, coinciding with a more balanced RepSeq composition in the latter (Fig. 8). Whereas subsets of RepSeq subfamilies as defined herein exhibit either pronounced ProA or pronounced ProB functions,

basically every RepSeq insert in the genome must be regarded as a composite with some degree of ProA and ProB capacities, which may vary depending on the cellular context and thus on the activities of TFs binding this particular copy.

Genes and their regulatory elements located in AorB regions behave as inducible ProA elements, switching these regions to A when they become active, and thereby forming new A domains (Fig. 8). Note that activity here does not primarily mean transcription but "active chromatin state", associated with the binding of TFs and the eviction of a nucleosome, forming DHS sites, which are evenly scattered along the domain. A major part of DHSs corresponds to "enhancers", known to recruit TFs and their cofactors to activate target gene promoters that can be identified as looping partners (181-183). In combination, the above-mentioned evidence suggests that another mechanism, by which enhancers and bound TFs stimulate gene transcription, is separate from *bona fide* transcriptional activation and consists in opposing the formation of B compartment and its associated generalized repression. Such "anti-silencing" activity of TFs and enhancers was actually postulated and experimentally demonstrated almost three decades ago (21, 86, 184-186). What has become appreciated more recently is that when a gene, embedded in a B domain, is silent and yet should be expressed, there is an intermediate stage when the gene is not yet expressed but the domain has already switched to an A configuration (Fig. 6C). This intermediate stage coincides with the emergence of enhancers in an immature form ("primed enhancers"), which essentially manifests itself as a DHS, with associated chromatin marks such as H3K4me1. Here we provide more evidence that a high density of DHSs of all types seemingly repels the B compartment, suggesting altogether a model in which an A domain will emerge on a B background precisely over a chromatin segment where the density of DHSs is sufficient to repel the B influence (Fig. 8). There is indeed evidence for cooperation between DHSs in doing so (103), a large part of which are enhancers in a steady-state situation, suggestive of a feed-forward loop whereby DHSs cooperatively enhance accessibility over a region, which in turn promotes TF binding at DHSs, further consolidating the presence of DHSs. There is a growing awareness that enhancers originate from RepSeq, and that a significant proportion of enhancers actually are *bona fide* RepSeq (59, 72, 74, 105, 107, 187, 188). Furthermore, the literature contains countless examples of situations in which RepSeqs, and ERVs in particular, switch from a conspicuously repressed or more often mundane chromatin state to a DHS or "enhancer" state, for example during viral infection or in cancer as reported herein, blurring the line between ProA and ProB identity. This is in fact true for every tissue- or developmental stage-specific enhancer, whether of recognized RepSeq origin or not, suggesting that DHS/enhancer induction may more generally consist in switching from a ProB to a ProA state, even in the absence of detectable TF binding and repressive marks prior to induction (189).

The genome must therefore be envisioned as a supersystem, of which the RepSeqs are the main constituents and also responsible for essentially all levels of organization. Schematically, there are three main subsystems, each composed of functionally redundant, cooperating elements: RepSeqs markedly ProA or markedly ProB, which broadly outline A/B partitioning, and RepSeqs with more indefinite characteristics. In human, a large proportion of the latter are ERV LTRs. These can shift between ProA and ProB functions, depending on the cellular context and on which TFs are expressed, and thereby cooperatively open or close the AorB domains. Such ERVs and their derivatives, and those of other RepSeqs, that emerged by idiosyncratic evolution and selection over evolutionary times, are better known as enhancers. In this picture, genes and their promoters appear like a special class of RepSeqs, "special guests at the genome party", which in their active, transcribed state reinforce the openness of their surroundings. In human, core ProA RepSeqs are essentially made up of Alu elements, whereas core ProB RepSeqs consist of young L1 and ERVs, as well as a panel of AT-rich microsatellites, and telomeric and pericentromeric satellites. The latter are the main constituents of constitutive heterochromatin, forming Megabase-sized chunks of DNA, which we assume plays a specific role in the B subsystem by massively assembling heterochromatin material (see below). This remains to be explored using in particular now available completely sequenced genomes (190).

This new perspective, which assigns the primacy in genome organization to RepSeqs, is a total change of paradigm compared to previous gene-centered models. Along these lines, another completely new insight relates to the function of enhancers. While enhancers are most widely believed to directly stimulate the activity of cognate gene promoters, the primary role that is now emerging for enhancers is to open domains (58, 59, 191). According to this view, their ability to regulate gene transcription directly comes second. The well-known synergy between distant enhancers in activating genes may therefore be partially accounted for by cooperative effects in opening domains, at different scales of genome organization (192). Consistent with these notions, enhanced opening of specific domains in two common chronic diseases, schizophrenia and bipolar disorder, correlates with the tissue-specific, coordinated upregulation of enhancers scattered along these domains (193). Nucleotide polymorphisms predisposing to the two diseases are enriched in a number of these enhancers (193), and similar findings have been made in other diseases (194, 195), suggesting more broadly that a variety

of disease-associated enhancers, for which the quest for a cognate gene has thus far proven unsuccessful (183, 196, 197), may specifically display domain-level activity.

What are the molecular mechanisms for ProA and ProB functions? We identify limited sets of RepSeq subfamilies in the human genome with archetypal features expected for ProA and ProB elements, essentially the Alu and SVA subfamilies for ProA, and the L1 and ERVs subfamilies for ProB. However, note that many more RepSeqs may have such a role in a less apparent way because of a more mixed character, or because they may elude identification at the subfamily level due to heterogeneity between individual inserts. The ProB subfamily archetypes are young L1 and ERVs with inserts known to be predominantly associated with high levels of H3K9me3/ HP1α heterochromatin-related markers; this points to a scenario in which repression of TEs is not only a necessity for the cell to maintain genome stability, but also a prime opportunity for genome organization. Pericentromeric satellites, which are the main constituents of constitutive heterochromatin sequences, and which notoriously nucleate HP1α -based heterochromatin assembly across species in spite of highly divergent sequences (17, 52, 198-200), score among the strongest ProB RepSeq in our analysis. There is evidence that simple repeats and low-complexity sequences scoring as ProB can also trigger local heterochromatin assembly, in particular via DNA:RNA hybrids or mechanisms fostered by non-B DNA (43, 50, 52). Finally, artificially targeting the KRAB domain of KAP1, a heterochromatin nucleator (see Box2), to sites along a domain can switch it from A to B (201), altogether making a strong case for an equivalence between ProB function and a capacity to nucleate HP1α -based heterochromatin assembly (Fig. 8, top).

The ProA function of Alu may similarly relate to the binding of TFs and associated cofactors (Box2). Alu subfamilies originate from the 7SL RNA gene that is transcribed by RNA polymerase III (Pol.III) from an internal promoter that they inherit, and therefore Alu elements bind the Pol.III-specific initiation factor TFIIIC (Box2). A ProA-type, antisilencing function has been demonstrated for Pol.III-independent TFIIIC sites, which has been conserved throughout evolution (202-205). All five TFIIIC subunits have histone acetyltransferase activity (HAT) (206), and each of these as well as ADNP, known to bind Alu elements as part of the ChAHP complex (64, 163), can interact with the protein Myc (207, 208). Both TFIIIC and Myc itself can recruit HATs, which overall suggests the assembly of a tree-like dynamic structure by which very high local concentrations of HATs can be attained at Alu inserts, especially in cells with high *MYC* expression (63, 125, 150), or upon cell stimulation resulting in Myc activation by phosphorylation (209). Pol.III is also activated upon cell stimulation, and subsets of both tRNA and Alu are rapidly induced (45, 210, 211). Histone acetylation prevents heterochromatin spreading, presumably by preventing nucleosome coalescence into clutches (212, 213). This could be one mechanism by which Alu elements oppose B installation, and which also applies to active enhancers, typically associated with HATs and in particular CBP/p300 (63, 181) (Fig. 8). As Myc not only binds Alu elements, but is in fact detected at every active enhancer and promoter in the genome, cooperation between these ProA elements in destabilizing nucleosome clutches may underly the massive genome opening observed upon *MYC* induction (125, 149, 150). Conversely, "low Myc" conditions may promote genome compaction by a reverse mechanism, such as in quiescent cells, which is associated with chromatin featuring predominantly low acetylation levels over the whole genome (125).

However, no clear histone acetylation signal is detected at Alu elements in a steady-state situation, nor at extra-TFIIIC sites in the genomes of either the fission yeast *Schizosaccharomyces pombe* (*S. pombe*) or the nematode *Caenorhabditis elegans* (202, 205). The fact that ProA function appears to be conserved in other species for SINE elements (214, 215), also derived from Pol.III transcripts, further suggests another, quite distinct hypothesis that would be based on a specific mechanism independent of TFs and their cofactors. A body of evidence points to a nucleosome positioning capacity for Alu elements, similar in fact to many Pol.III-transcribed genes, that would depend on CpG methylation at least in the case of Alu (Box2). By contrast, the assembly of chromosome segments into heterochromatin is envisioned to involve translational motions of nucleosomes, as enacted by ATP-dependent remodeling machineries working together with HMG B-box proteins, presumably in order to optimize HP1α binding and the bridging of nucleosomes (216-218). Support for this notion comes in particular from a newly described function of the FACT complex, which remodels H2A-H2B and is best known for its involvement in transcription elongation. FACT was found to be crucial for heterochromatin spreading and maintenance in *S. pombe* (219). It was also shown to be strongly enriched in an *in vitro* reconstituted SIRc-heterochromatin phase together with other remodelers, such as RSC, ISWI1, ISWI2, INO80, CHD1 and to a lower extent the SWI/SNF complex, in *S. cerevisiae* (220). The HMGB factor Nhp6A was also found to be enriched in the latter system (220), and both HMGB1 and HMGB2 were identified upon isolation of natural pericentromeric heterochromatin in mouse embryonic stem (ES) cells (221). In the same vein, the variant histone H2AZ, best known for its role in aiding the priming of transcription elongation by Pol.II at the +1 nucleosome of genes (222), appears, in a seemingly paradoxical way, to have a role in heterochromatin spreading, and

more broadly in the installation of a B logic over a domain (223-226). H2AZ-containing nucleosomes display a redistribution of histone-DNA interactions (227) that facilitate Pol.II passage and probably, more broadly, mobilization, substantiating the view that nucleosomes must be mobile during heterochromatin installation, similarly to transcription elongation. One likely scenario that emerges is therefore that Alu elements, and presumably also SVA elements, of which a part derives from Alu elements, impede the translational remodeling of nucleosomes (Fig. 8, bottom) and thereby interfere with cooperative phenomena associated with heterochromatin spreading (Fig. 8, right). Even a minor loss of CpG methylation, as observed in AlwaysB, is expected to significantly reduce the nucleosome positioning capacity of Alu elements (162), and therefore their ProA potential, which may account for the surprising presence of a large number of Alu elements in AlwaysB. Moreover, effects might be non-linear, being for instance negligible at a low density of Alu elements such as found in the AlwaysB subcompartment, and being prominent in regions that should crucially be kept open, such as the loci of immediate early genes, that display very high Alu densities. The combination of high ProA and high ProB RepSeq densities, such as that observed over gene complexes embedded in the B compartment, can be interpreted to provide an environment compatible both with transcription via an insulation effect and with strong repressive control of genes within (17) (Fig. 8).

As a DHS primarily represents an interruption in the "beads on a string" organization of the nucleosomal fiber, it is associated with a high nucleosome turn-over to keep the DNA element nucleosome-free. Therefore, the capacity of DHSs to oppose B installation may similarly relate to the fact that translation of a nucleosome through a DHS is virtually impossible (Fig. 8, bottom). Such a mechanism was proposed early on for TFs that potently recruit remodelers to account for their capacity to uncouple adjacent chromatin domains and to prevent heterochromatin spreading, as well as for tRNA genes, in *S. cerevisiae* (21, 228, 229). The capacity of multiple such DHSs scattered along a domain to switch it from B to A was further recapitulated by *in silico* modeling (230). More details as to how the proposed mechanisms may translate into varying proportions of ProA and ProB RepSeqs shaping distinct chromatin landscapes are provided in Fig. 8 caption. In short, the fundamental principle responsible for non-linear effects by which seemingly minor variations in ProA content tip a B domain into an A domain is primarily a strong cooperativity between ProB elements. This cooperativity is presumably underpinned by the condensate-like properties of heterochromatin material, but it also involves the spreading of heterochromatin along the chromatin fiber. ProA elements interfere with heterochromatin spreading through an activity that can be broadly described as "anti-silencing", as previously proposed (17, 21), thus breaking cooperativity between ProB elements.

Genomes are in flux, as recognized by Barbara McClintock decades ago (231), with TEs inserting more or less randomly, being possibly fixed in a population, and eventually drifting along evolutionary times. Enhancers emerge essentially from TE inserts. These harbor binding sites for tissue-specific TFs and may acquire others over time by genetic drift. This is consistent with the fact that enhancer sequences are largely species-specific, being conserved only between close phyla, in sharp contrast to genes (74, 187, 232). Simple repeats may also be involved (47, 108, 109). Any young TE is subjected to repression upon insertion and can seemingly behave as a ProB element. Our observations are consistent with a model in which ProB RepSeq operate as a system of redundant elements that cooperate by little, dynamic touches across scales and in particular at a domain level (17, 233, 234). The same can be said for ProA elements, except that ProA seemingly cooperate to oppose the genome driving force that emerges from the ensemble function of ProBs. Sequence drift upon TE aging weakens ProB function, and a TE will be counter-selected in the B compartment if a ProA capacity emerges and takes over, which would no longer be in accordance with the genomic environment. Such a scenario appears to account for the observed distribution of Alu repeats, with the youngest inserts being progressively lost in B, in particular on the Y chromosome, as recognized early on (67, 235). It should therefore be appreciated that the selection pressure that acts to maintain an adequate RepSeq composition in any DNA segment over the genome also proceeds by small touches. These operate on populations of interchangeable elements, which can be replaced when one or several gradually disappear in the course of evolution, to generate a habitat, which is diffusely more conducive to the dynamic installation of a chromatin system of a particular flavor. This is how order emerges out of chaos within eukaryotic genomes, populated for the most part by repeated sequences, which were until recently denied the right to a functional existence to the point of being described as junk DNA. Repeated sequences operate in a collective manner and thus are subject to selection as an ensemble according to a logic that is totally different from the individualistic logic that operates at the level of genes. Using this conceptual framework and unprecedented technical possibilities, we can now read and contemplate genomes as never before.

BOX 1: Life history of retrotransposable elements

Life history of retrotransposable elements include:

1- Production of an RNA from a source, full-length TE insert.

2- Reverse copying by a reverse transcriptase (RT) encoded in cis or in trans **and insertion into the genome**, producing a new TE insert. The new insert may include a promoter, especially when the original promoter is internal in the source TE (e.g. full-length L1, Alu), as well as an enhancer (e.g. the LTR of full-length ERVs, or solo LTRs, derived from ERVs by recombination between LTRs), which altogether are referred to hereafter as RepSeqReg. TE display varied levels of %GC with L1 being overall AT-rich while ERVs are GC-rich (Extended Data Fig. 6). However, RepSeqReg are usually GC-rich, and therefore appear for L1 as a GC-rich window in the context of an otherwise AT-rich sequence (236).
3 - Regulated expression: RepSeqReg are regulated by both transcriptional activators and repressors, as are cis-regulatory gene elements, hereafter referred to as GeneReg. In particular, ERV subfamilies are expressed in specific cell lineages due to the binding of **tissue-specific transcription factors** to the LTR enhancer. However, RepSeqReg are, more than GeneReg, subject to specific, **targeted heterochromatin-mediated repression, which suppresses potentially harmful transcription, but also recombination** between repeats, thus preserving the genome from instability. Heterochromatin is targeted at TE by dedicated repressive systems and involves both H3K9me3 deposition and CpG methylation and recognition by cognate binding partners (see Box2).
4- Drift of the sequence, including in particular progressive loss of CpG (converted to TpG due precisely to methylation of CpG)**, correlates both with a loss of transcription capacity and the weakening of repressive mechanisms.** Remarkably, some heterochromatin marks can still be observed at some TE inserts which have lost all transcription capacity, and there are instances where positive selection of such combination could be demonstrated (56).
5- TE as a source of innovation for gene regulation: At all stages of the TE insert idiosyncratic evolution, **RepSeqReg can specifically modulate the activity of neighboring gene promoters.** RepSeqReg are therefore subjected to **selection pressure** as **silencers or enhancers,** seemingly acquiring the binding of some TFs and loosing others. It became recently appreciated that gene enhancers: (i) are fast evolving sequences which actually are of or derive from RepSeqReg, and in particular from ERV LTR in human (74); (ii) intrinsically display both a silencer and an enhancer capacity and can commute between functions according to the cell context (237, 238).`

BOX2: CpG methylation in repression of L1s and ERVs, and in activation of Alu elements: mechanisms and functions.

The cytosine of CpG dinucleotides is subject to methylation by dedicated enzymes called DNA methyltransferases (DNMTs). CpGs are generally lost with age in the genome during evolution, due to the effects of pervasive methylation and the increased likelihood of conversion of mC to T, resulting in a global depletion of CpGs during evolution as compared to GpCs for instance. RepSeqs, which account for 56% of the human genome, harbor 61% of CpG dinucleotides, with the remainder mostly found in gene exons and their regulatory elements (239). When cells are treated with 5-azacytidine (AZA) or derivatives thereof, DNMTs are inhibited, which leads to a rapid, global loss of DNA methylation in actively dividing cells, and results in the upregulation of young ERVs (Extended Data Fig. 7) while transcription of the vast majority of Alu inserts is unaffected (61, 65, 240). In addition, levels of ongoing transcription correlates negatively with CpG methylation for L1s but positively for Alu elements, and this is also the case for Pol.III ChIP levels for Alu elements (61, 239). Finally, the GC and CpG composition of Alu and SVA clearly stands out from that of other TEs (Extended Data Fig. 6), altogether pointing to a distinctive role of CpG methylation between these TE classes, as previously suggested by others (26).

CpG methylation as a memory mark for the heterochromatin-mediated repression of L1s, ERVs, and pericentromeric satellites.

The expression of RepSeqs is inhibited by an interplay of interwoven, largely redundant repressive mechanisms (241), with a pivotal role of H3K9me3/ HP1α -based heterochromatin (132, 221, 242, 243). Heterochromatin can be nucleated by TFs, such as the KZFP multigene family, which recruit KAP1, a potent heterochromatin-scaffolder, or via RNA-based mechanisms such as the PIWI and the HUSH systems. The latter are especially prominent to silence transcription-competent inserts in cells with a fluid chromatin, such as during early development or during gametogenesis (56, 244-247). The methylation of CpGs by DNMTs is another important player in enforcing RepSeq repression in human. meCpG renders the heterochromatin state of RepSeq more stable both within a cell cycle and through generations, by introducing some extra feed-forward and reciprocal loops in the heterochromatin regulatory network (21, 248, 249). CpG methylation therefore appears to guarantee the memory of a repressed state, which was formally demonstrated experimentally by artificially targeting DNMTs, or TET enzymes that revert meCpG, at reporter genes (111, 177, 178, 241, 250). Conversely, mutating DNMTs is embryonically lethal in mouse, due to an inability of embryo cells to securely silence prior or alternative cell identities (251). Treating cultured cells with AZA not only affects TEs but also induces loss of meCpG in pericentromeric satellites and subtelomeric RepSeqs (252), as also observed in patients with Immunodeficiency with Centromeric instability and Facial dysmorphia (ICF)

syndrome, who carry mutations in DNMT3b or cofactors thereof, and display prominent decompaction of specific pericentromeric regions in affected tissues (253).

CpG methylation is pivotal for the repression of young ERV subfamilies in differentiated cells. Gradual loss thereof along with sequence drift, if one considers progressively older TE (Extended Data Fig. 9), appears to be compensated by an increased contribution of H3K9me3-based mechanisms, until this is eventually lost as well (65). However, a fraction of ancient TEs still display significant KAP1 ChIP signal, in an idiosyncratic manner (56, 65). Intriguingly, this sequence of events occurring on an evolutionary timescale is emulated *in vitro* over a period of weeks when cultured cells are treated with AZA, with H3K9me3 rising at young ERVs while CpG methylation decreases (65). A similar sequence of events is also observed during the very early stages of development in mammals when the genome, in general, and TEs, in particular, get massively demethylated and many TEs are expressed, all of which is required for totipotency (128, 250, 251). Soon thereafter, RepSeqs are repressed again by concomitantly upregulated KZFPs and cofactors, in particular the histone lysine methyl transferase SETDB1, which results in the deposition of high levels of H3K9me3 at TEs (59, 105). CpG methylation rises again as cells exit the pluripotency state, of which ES cells are an equivalent, and eventually reaches patterns and levels that are found in differentiated cells during gastrulation (251, 254). meCpG is lost in cancer across the B compartment (Fig. 7), and there is a general trend by which H3K9me3 levels increase in cancer (173), echoing the sequence of events observed during evolution and in early development. In cancer, the apparent compensatory rise in H3K9me3 may be actually promoted by anti-cancer treatments, and it is presumably due to both a physiological response to RepSeq transcription involving heterochromatinization pathways, and to a selection against RepSeq transcription, that would result in awakening the innate immunity pathway and subsequent elimination of the cancer cell (140, 171, 255-259). Although RepSeqs may be sufficiently repressed by alternative mechanisms and the compensatory rise of H3K9me3 such as not to, or minimally, trigger surveillance pathways and therefore allow "evasion" of the tumor cell from the immune system (255), the quality of repression is clearly not the same with and without the involvement of meCpG-based mechanisms (178).

CpG methylation may enforce nucleosome positioning over Alu.

There is clear evidence that a large proportion of Alu elements may be transcribed by Pol.II. This is most likely because they reside in close proximity to gene enhancers or promoters and are embedded in sequences transcribed into eRNA or uaRNA, respectively, which was recently shown to structurally stabilize enhancer-promoter interactions (260). Alternatively, Alu elements can evolve to become *bona fide* enhancers, binding Pol.II-related TFs (62-64, 74).

However, Alu elements are primarily transcribed by Pol.III, since they originated by reverse transcription from the non-coding RNA 7SL, and therefore inherited the internal Pol.III promoter of the 7SL RNA gene. Genes transcribed by Pol.III are most often short, about the size of a nucleosome, and they do indeed display a positioned nucleosome, like the 5S rRNA gene which has been used from early on as a workhorse nucleosome-positioning sequence (261). For Pol.III, unlike Pol.II, the presence of a nucleosome not only does not prevent transcription initiation, but in fact promotes it: (i) TFIIIC, which scaffolds Pol.III pre-initiation complex assembly, can bind nucleosome-embedded DNA; (ii) Transcription by Pol. III can operate without eviction of the nucleosome; (iii) *In vitro* reconstitution experiments show that transcription activation by TFIIIC is much more efficient if a nucleosome is present (262). Note that some highly transcribed Pol.III-dependent genes, such as tRNA genes in the yeast *S.cerevisiae* (229), do however exhibit nucleosome eviction, but remarkably, nucleosomes on either side of those genes are H2AZ-depleted and strongly positioned; this is required for efficient transcription (263), emphasizing that nucleosome positioning is important for Pol.III-mediated transcription. Alu elements originated as dimers of the 7SL cDNA, although internal promoter sequences were conserved only in the first monomer, and they strikingly exhibit one of the highest nucleosome occupancies in the human genome, with one nucleosome positioned on each half of the dimer (162). The two monomers are separated by an A-rich linker, and Alu elements end with a polyA stretch, both of which are known to strongly exclude nucleosomes, and are therefore predicted to contribute to the positioning of the two nucleosomes over Alu elements (264). In addition, there is evidence that mutations predicted to further promote nucleosome positioning may be positively selected over Alu elements (264-266). Considering that Alu elements are among the most highly methylated sequences in the human genome (Fig. 7; Extended Data Fig. 9), for reasons that have persisted as a riddle for almost fifty years, we favor the following model: Cytosine methylation at CpG dinucleotides, as observed at Alu, alters the DNA structure in a manner that promotes nucleosome positioning and occupancy. This is supported by a wealth of arguments: (i) the CpG distribution profile of Alu elements displays a remarkable pattern, shared by both monomers, seemingly obeying a grammar that remains to be deciphered, with the most frequent spacing between two CpGs being 8 nt, but also with spacings of 22 or 44 nt (61). This is reminiscent of the artificial nucleosome positioning sequence called Widom601, which was identified by *in vitro* selection. It contains short G/C blocks, which repeat every 10 bp, interspersed with short A/T blocks, facing outward and toward the histone octamer, respectively (216); (ii) methyl-cytosine can be considered as a fifth DNA base with distinctive

properties, with meCpG displaying heightened base stacking, duplex stability, local curvature, and torsional rigidity of the DNA as compared with unmethylated CpG (248). In addition, the two bulky, hydrophobic methyl groups on adjacent Cs in a CpG dinucleotide are predicted to lead to a slight widening of the major groove and a corresponding narrowing of the minor groove (162); (iii) CpG methylation of natural CpG-rich sequences has been reported to favor nucleosome stability *in vitro*, with a tendency for the minor groove of meCpG to be rotationally oriented toward the histone surface, which is opposite to nucleosome Widom601, and to show correlation with nucleosome occupancy *in vivo* (162, 267)
Collectively, these lines of evidence lend strong support to the idea that the *raison d'être* of cytosine methylation at Alu elements is to promote nucleosome positioning and occupancy.

BOX3: RepSeq transcription is the Achilles' heel of cancer cells

A whole body of literature shows that RepSeq transcription is a commonality of cancer. Combinatorial signatures of a limited set of RepSeqs can be defined much in the same way as this can be done with panels of genes for mutations, or altered expression or chromatin states, which allow the clustering of tumor samples with a high predictive value (159, 160, 268-271). RepSeq repression is secured by redundant pathways in a healthy cell (Box2) such that RepSeq transcription can occur only when several of them are debilitated or dampened, in particular CpG methylation, due to either targeted or global deregulation as discussed herein. In addition, the constitutive cancer-related activation of signaling- or hormone-responsive pathways translate into TFs binding and activating RepSeqs (130). However, a more detailed analysis shows that, in a given tumor, RepSeqs that exhibit transcriptional activity represent only a fraction of all RepSeq inserts that switch to a promoter- or enhancer-associated chromatin state (272), similar to what we describe herein during development or viral infection. RepSeq derepression, whether or not associated with transcription, can now be understood in terms of a switch from a ProB to ProA function, or tuning a ProA status towards more ProA, with transcription enforcing a marked ProA character. RepSeq transcription at a high level is also detected in healthy stem and progenitor cells, both in early development and in tissues (273, 274). Interestingly, while genome unfolding is prominent both in aggressive forms of cancer and healthy stem cells, TE RNA levels may be surprisingly low in the cancer cells. This appears to be accounted for by post-transcriptional suppression of TEs by a variety of mechanisms (13, 273). More broadly, the innate immunity pathway, highly conserved in the animal kingdom and best known for its defense role against microbes, is the first line immune surveillance system against cancer by detecting TE-derived dsRNA. The latter activity is physiologically suppressed by a number of mechanisms in healthy stem and progenitor cells, to accommodate prominent levels of RepSeq transcription associated with genome unfolding (275-278). The mechanisms involved in stem cells are subverted by cancer cells as an immune evasion strategy. On top of this, the innate immunity system itself is commonly crippled in cancer, via both genetic and epigenetic alterations of gene expression (279), apparently without ever being completely ablated. Acutely treating cancer cells with 5-azacytidine, a DNA demethylating agent, induces a sudden and massive burst of RepSeq transcription (Extended Data Fig. 7A). It was reported in 2015 that this
arouses what remains functional of the innate immunity system, turning the tumor into an immunogenic body: the cancer cell is unmasked, at least temporarily, before having a possibility to adapt (65, 130, 240, 268). This is in fact the rationale behind anti-cancer therapeutic strategies combining immunotherapy with a range of epidrugs that undermine RepSeq repression (280-282), which may seem paradoxical at first sight as the later treatments may be independently known as oncogenic. Weakened repression of ProB repeat sequences therefore emerges as the Achilles' heel of cancer cells, being both a necessity for cancer cell plasticity, and at the same time generating 'sneaks' that signal abnormal genome unfolding to the immune system (283).

RESOURCES & METHODS

Resources

Data pertaining to chromatin-related features (histone marks, chromatin sensitivity to DNase I) were gathered from validated ChIP-seq and DNase-seq experiments in the HUVEC reference epigenome series of the ENCODE project (ENCSR194DQD), namely for H3K4me3 (ENCSR578QSO), H3K27ac (ENCSR000ALB), H3K27me3 (ENCSR000DVO) and DHS (ENCSR000EOQ). GRO-seq cumulated counts over a 100 kb bin were used as a proxy for RNA polymerase activity, and therefore transcription intensity. GRO-seq data and H3K9me3 data for HUVEC were downloaded from the gene expression omnibus (GEO) under the dataset GSE94872. The samples selected for GRO-seq were the 4 normoxia tracks (GSM2486801 to GSM2486804). A list of regulatory elements that operationally behave as Polycomb-responsive silencers in HUVEC was established from the literature (Table S2 from Ref (284)) and named "Silencer". Note that most of these elements score as enhancers in other cellular contexts (284). RepeatMasker annotations for hg19, hg19 RefSeq genic feature locations, hg19 GC content, hg19 chromosome size, and ChromHMM tracks for HUVEC were downloaded from the

UCSC Genome Browser.

Hi-C data were retrieved from the Juicebox database via Juicer Tools (https://hicfiles.s3.amazonaws.com/hiseq). Juicer Tools was used to compute HiC EV from Hi-C data at 25, 50, and 100 kb resolutions, and EV values were used to define A and B compartments as described (1). HiC EV was thus computed for eight human cell lines:
- IMR90 are primary, expanded fibroblasts derived from embryo lung; HUVEC are primary, expanded umbilical vein endothelial cells; NHEK are expanded normal human epidermal keratinocytes;
- HMEC were obtained by SV40 T-antigen-mediated immortalization of endothelial cells; GM12878 is a lymphoblastoid cell line obtained by EBV-mediated immortalization of normal B cells;
- KBM7, K562, and Hela are cancer-derived cell lines established from chronic myeloid leukemia (KBM7, K562), and cervical carcinoma (HeLa) samples, respectively. Hela cells harbour HPV-18 sequences and in particular express the HPV-E6 protein.

AorB vector

The HiC EV for the 8 human cell lines was used to generate an aggregate parameter, denoted "AorB.vec" (for "AorB vector"), which conveys the inherent propensity of a genomic region to display A (EV>0) or B (EV<0) chromatin identity, as follows. The value 4 was attributed to bins that display (EV>0) in the 8 cell lines, 3 if (EV>0) in 7 cell lines, and so on up to -4 if (EV>0) in 0 cell lines which equates (EV<0) in the 8 cell lines. In particular, we could thereby delineate 3 main classes of bins: AlwaysA (EV>0 in 8/8 classes), AlwaysB (EV<0 in 8/8 classes), AorB (both EV>0 and EV<0 values among the 8 cell lines for this bin). In addition, AorB bins display either EV>0 (AorB:A) or EV<0 (AorB:B) in a given cell line. In the HUVEC cell line, AlwaysA, AlwaysB, and AorB fraction represent respectively 22%, 23%, and 55% of hg19 genome, which is largely devoid of pericentromeric and subtelomeric regions. Our analysis therefore applies only to chromosome arms.

RepSeq insert count and copy count in the human genome

RepSeq were identified in the hg19 genome as segments of homology with the consensus sequences of subfamilies ("NAMEs") found in the RepBase repository. Individual inserts may be counted multiple times if they are split into more than one segment, mostly due to short INDELs or to the insertion of other TEs, in particular Alu elements. We therefore obtained a proxy for the count of inserts by joining any two segments that are assigned to the same NAME and within 350 bp of each other. Note that 350 bp is slightly above the size of Alu sequences. Counting ERV occurrence requires a specific approach in that an ERV insert may consist either of a full-length provirus made of two LTRs flanking an internal region ("INT"), or of a solo LTR derived by recombination between two identical LTRs in the course of evolution. Indeed, LTR and INT sequences appear as distinct NAMEs in RepBase. ERV insert count was obtained by substracting INT insert counts from LTR insert counts. "copy count", which is the metrics classically used in bioinformatic analyses refers to the number of segments. We refer to the above proxy for insert counts as "insert count".

Correlation with HiC EV

We computed the Spearman correlation between the HiC eigen vector (HiC EV) value in the 8 human cell lines and repeat sequence density (copy counts, 25 kb bins) along the hg19 genome (Table 1, Table 2). The 1395 subfamilies of RepSeqs found in the RepBase repository (NAMEs) were used. Only the correlation with the HiC EV in the HUVEC cell line is used in this study. Correlation values and order in the classification were very similar when the EigenVector derived from another cell line was used (Extended Data Fig. 7; Table 2). Of note, Spearman correlation with the HiC EV is negatively impacted when a RepSeq subfamily is absent from a large number of bins, which occurs more often for subfamilies having a low number of copies in the genome.

Assessing feature density or summed signal in 100 kb bins

Apart from RepSeq, features used in multivariate linear models (see below) are peaks (DHS or ChIP peaks, for instance), intervals (ChromHMM states; Genes), or signals (GRO-seq). Assessing features in 100 kb bins was done with the BEDtools software suite.

Gene density was established as follows. Genes of the same symbolic name on the same chromosome and strand were merged together. Overlapping genes on the same chromosome and strand, which share a common prefix of at least one letter were also merged together. Note that in this work, merging is defined as summarizing several overlapping intervals by a single interval with a starting position as the minimum start of its components, and an end position as the maximum end of its components. Assessing gene density allowed to distribute bins into "no gene" (1-3 genes /100 kb), "Regular Gene Density" (1-3 genes /100 kb), and "High Gene Density" (>3 genes /100 kb) subclasses.

DHS count: We used the broadPeak file associated with GSM736575, in which the threshold is set at a lower value than with narrowPeak. The idea was to be more sensitive and to detect small peaks that would be missed by narrowPeak, and accordingly the former detects twice more peaks than the latter ($25,5 \times 10^4$ versus 12×10^4). However, such sensitivity is also associated with a higher background of false peaks, estimated at about 2-5 peaks/100 kb, more pronounced in the B compartment than in the A compartment.
GRO-seq signals of the four experimental replicates for HUVEC cultivated in normoxic conditions (see above) were concatenated, and a sum was calculated, which conveys global transcription intensity over a bin. Outliers displaying abnormally strong signals were brought down to a threshold defined according to the Tukey's Fences method with k=5 upper bound.
ChromHMM segments: The HUVEC ChromHMM track was split into individual files for each state.
ChromHMM state 4 and state 5 intervals, corresponding to strong enhancers found within or outside of a gene, were gathered into a single list ("Strong.Enhancers"). ChromHMM state 6 and state 7 intervals, corresponding to weak and poised enhancers, respectively, were also grouped together in a single list ("Weak.Enhancers"). Most intervals detected as "Strong promoter state" (state 1) are bracketted by adjacent "weak promoter state" intervals (state 2), which in fact relate to the same *cis*-regulatory element. In order to avoid overestimating the number of active promoters, we decided to merge state 2 intervals within 1000 bp of a state 1 interval with the latter. The name "Strong.Promoter" was used for the resulting list of merged state 1 and state 2 intervals, and "Weak.promoter" for the list of state 2 intervals depleted of intervals included in the "Strong.Promoter" list. A bin was considered as "Active" when it contains at least one "Strong.Promoter" or one "Weak.Promoter" interval. Note that ChromHMM state 1 and 2 do not always overlap with gene promoters.

Multivariate linear model of HiC EV profile

Multivariate linear models were generated using either density or mean value for a number of features over 100 kb bins in order to predict the HUVEC HiC EV. Both DNA sequence features and epigenomic features as detailed above were used in all possible combinations. DNA features comprised: copy count for different RepSeq subsets, and gene density. Epigenomic features comprised DHS peaks, Strong.Promoter, Weak.promoter, Strong.Enhancers, Weak.Enhancers, silencer intervals, H3K4me3, H3K27ac, and H3K27me3 peaks, and GRO-seq summed signal. Models were established using values of features as assessed using the hg19.noflank version of the genome, with the idea that trends might be clearer using only the core of chromatin domains, and discarding transition bins. Models were then compared using the adjusted square of the sample Pearson correlation coefficient to estimate the fraction of the variance in the EV that is explained by the parameter vector, whilst adjusting for the number of parameters in the model. Those models with the highest adjusted R-squared value were defined as the most performing.

GO analysis

A GO analysis was performed for gene-containing bin subclasses 1-16 as defined in our study using data obtained with HUVEC cells. The mapping of GO terms on human genes was obtained from the R package "org.Hs.eg.db" version 3.12. Overrepresentation of GO terms in bin subclasses as defined with HUVEC data was assessed using the classical approach as described (285). P-values were calculated with the one-sided Fischer's exact test and adjusted for multiple comparison with the BH method. GO terms referred to herein as "enriched GO terms" (or eGO) are those for which enrichment reaches statistical significance ($P < 0.05$) at the level of the subclass considered, amounting to 1938 GO terms in total for the 16 subclasses. In order to determine whether genes embedded in individual bin subclasses display any specific pattern with regard to functional specialization, we created a list of 34 terms more cross-sectional than GO terms, dubbed "GF-GO.slim", such that it is possible to assign a unique GF-GO.slim to each GO term by following a procedure detailed below. Although there is some overlap, GF-GO.slim are distinct from classical GO.slim in that a GO term is commonly associated with several GO.slim. First, the full list of GO terms identified in our analysis as enriched in at least one bin subclass was searched for a number of strings, and a GF-GO.slim was attributed in case the string was identified, as detailed below. When the string "signal" was identified, the GF-GO.slim "signaling" was attributed. When either of the strings "stress", "stimul", "response" were identified, the GF-GO.slim "stress, stimulus" was attributed, except for GO terms, which are immunity-related, in which case the GF-GO.slim "immunity" was attributed. When either of the strings "nerv", "neuron", "axon", "synapse", or "dendrit" were identified, the GF-GO.slim "nervous system" was attributed, except if the GO.terms included the words development, differentiation, or commitment, in which case the GF-GO.slim " development, differentiation" was attributed. When either of the strings "skin", "keratin", or "corn" were identified, the GF-GO.slim "skin" was attributed, except if the GO.terms included the words development, differentiation, or commitment, in which case the GF-GO.slim " development, differentiation" was attributed. When the string "musc" was identified, the GF-GO.slim "muscle" was attributed, except if the GO.terms included the words development, differentiation, or commitment, in which case the GF-GO.slim " development, differentiation" was

attributed. When the string " plasma membrane" was identified, the GF-GO.slim " plasma membrane" was attributed, except if the GO.terms included the words transport, regulation, adhesion, in which case the GF-GO.slim intracellular transport, localization, cytoskeleton", "regulation, system process", or "cell motility, adhesion, polarity, localization", respectively, were attributed. When the string "regulation", or exactly the wording "system process" were identified, the GF-GO.slim "regulation" or "system process", respectively, were attributed, irrespective of the process targeted by the regulation in question. When the string "checkpoint", "homeosta", "viral", or "chromosome" were identified, the GF-GO.slim "checkpoint", "homeostasis", "virus related", or "chromatin, transcription, chromosome", respectively, were attributed. When none of these possibilities were met, GF-GO.slim were manually attributed in a unambiguous manner, i.e. such that priority matters did not have to be considered. Notably, the GF-GO.slim "development, differentiation" includes terms related to morphogenesis. GF.GO slim were grouped into four clades: I (01-04): Crosscutting principles; II (05-16): Typical housekeeping functions, such as metabolism and intracellular transport, also including cell cycle, apoptosis, autophagy; III (17-21): Housekeeping functions associated with a specific cellular substructure, possibly associated with a cell differentiated phenotype; IV (22-34): Cell communication with its environment (23,24), chromatin and transcription (25), development and differentiation, and differentiated cell functions, in particular related to Immunity (34). Clade IV thus encompasses what may be called "evolutionarily specialized function", as opposed to housekeeping functions, which are represented par excellence in clade II and to some extent also in clade III.

Identification of "outlier bins": HUVEC.A(vecB), HUVEC.B (vecA)

There are 1601 bins, which display a HiC EV sign in HUVEC different from the sign of the AorB vector, i.e. which adopt a chromatin configuration in HUVEC, which does not follow the general trend of the majority of tissues ("outlier bins"). In order to study the characteristics of these regions and to observe marked features, we focussed only a subset of these bins ("outstanding outliers") by applying the following selection criteria: (i) We discarded outsider bins that lie within (-100/+100 kb of an A/B junction, as well as isolated bin outsiders, i.e. we retained only those bins that form an outsider cluster containing at least 2 bins in a row; (ii) we retained only those outlier bins for which not only the HiC EV sign and AorBvec sign are different, but for which the difference in value is significant. To do this, we reduced AorBvec values to -0.05/+0.05) (AorB.vec*), and retained only those outlier bins for which the difference between HUVEC HiC EV and AorBvec* is above 0.03 (for HUVEC.A(vecB) bins) and below -0.03 (for HUVEC.B(vecA) bins).

Analysis of CpG methylation

Oxford Nanopore Technologies (ONT) PromethION data for liver samples were sourced from a previous publication (236). CpG Methylation and canonical bases were called from .fast5 files using ONT guppy version 6.2.1 using the model "dna_r9.4.1_450bps_modbases_5mc_cg_sup_prom.cfg". Methylation statistics were generated using methylartist "segmeth" (286) for RepeatMasker annotations downloaded from the UCSC Genome Browser assembly hg38 (287). RepeatMasker classes retained for analysis were: LTR/ERV1, LTR/ERV1?, LTR/ERVK, LTR/ERVL, LTR/ERVL?, LTR/ERVL-MaLR, LINE/L1, SINE/Alu, and Retroposon/SVA.

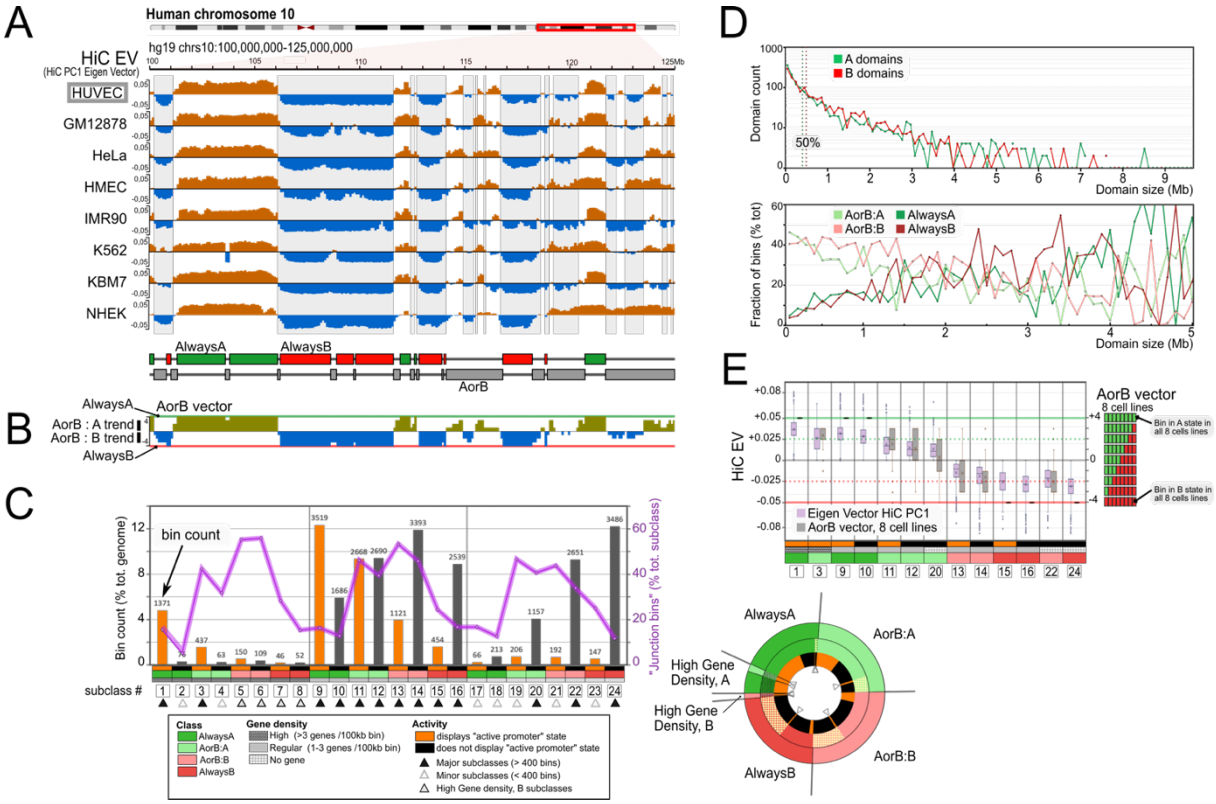


Fig. 1: High conservation of the A/B compartment profile points to the existence of pivotal DNA determinants.

(A) The Eigen Vector of the HiC PC1 is aligned for 8 ENCODE human reference cell lines, over a 25 Mb region of human chromosome 10. HUVEC and NHEK are primary, expanded cells; IMR90 is derived from fetal lung; HMEC and GM12878 are immortalized by SV40 T antigen and EBV, respectively; the others are cancer cell lines. Bottom: Regions that are A (respectively B) in all 8 cell lines are indicated as "AlwaysA" green boxes ("AlwaysB" red boxes, respectively).

(B,E) The "AorB vector" function sums the compartment state of the 8 cell lines as shown in A. Its value is +4 for AlwaysA regions and -4 for AlwaysB regions, and has intermediates values otherwise. E, Boxplot representation of AorB vector distribution shown for the 13 major bin subclasses alongside with HUVEC HiC EV.

(C) Partitioning of the human genome into 24 subclasses of 100 kb bins according to compartment, gene density, and "activity". The latter corresponds to the presence of at least one segment of "active promoter state" (ChromHMM state1 or 2 (25)). For each subclass, the absolute value of bin counts is shown at the top of histogram bars and as % of the total bin count on the Y-axis on the left; The pink line indicates the fraction of bins in a subclass lying within 200 kb of an A/B transition ("junction bins"; Y-axis on the right). Scheme on the right: Pie representation of the 24 subclasses.

(D) Domain size as a function of number of domains with a certain size, shown independently for A domains and B domains (top); domain size as a function of fraction of total bins (bottom). The total population of bins belonging to domains of a certain domain size is considered. Shown is the fraction of such bins belonging to each of the 4 bin classes (AorB:A, AorB:B, AlwaysA, AlwaysB)

A domain is defined as a group of contiguous bins between two A/B transitions. Dotted line, median values.

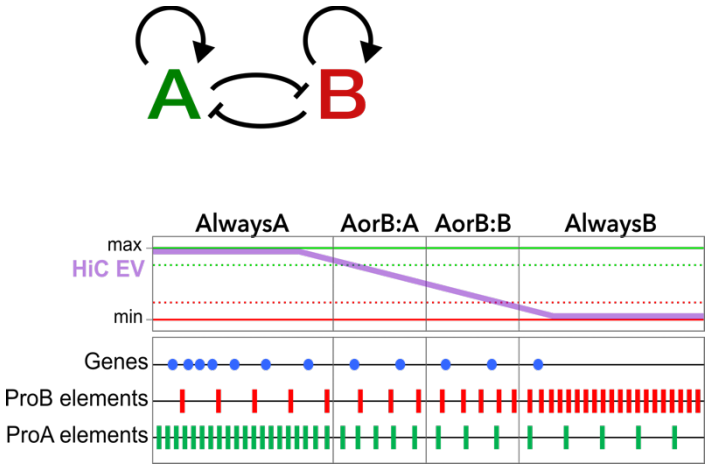


Fig. 2: A/B compartment toggle-switch points to ProA and ProB determinants as both enriched in the cognate compartment and correlating with HiC EV value.

A/B compartmentalization is driven by a toggle-switch principle (upper part) which allows prediction of the distribution of ProA and ProB elements, defined as cis-determinants of A/B partitioning promoting A and B compartments, respectively.

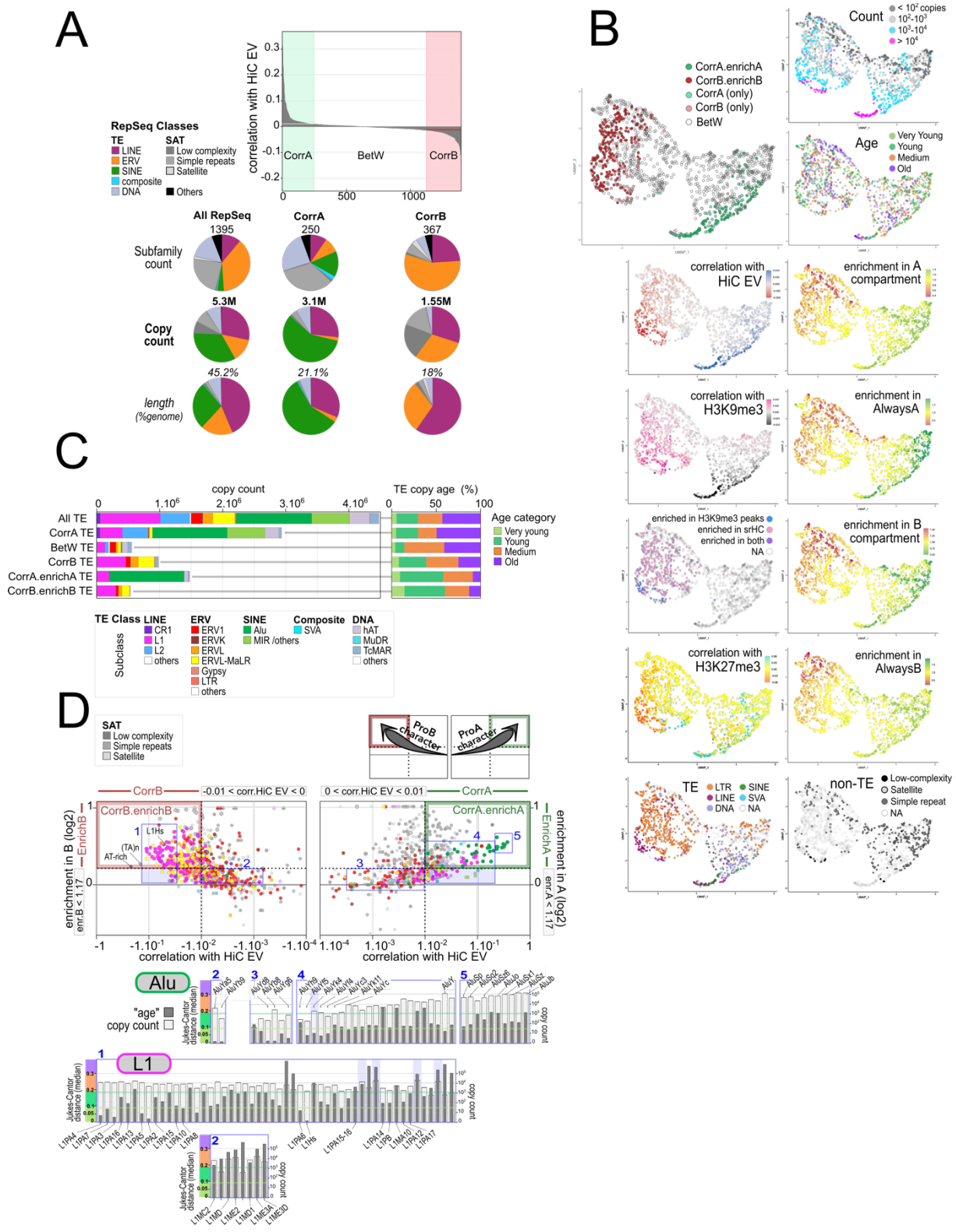

A
correlation with HiC EV
CorrA
BetW
CorrB
RepSeq Classes
TE
LINE
ERV
SINE
composite
DNA
SAT
Low complexity
Simple repeats
Satellite
Others
All RepSeq
CorrA
CorrB
Subfamily count
Copy count
length (%genome)
B
CorrA.enrichA
CorrB.enrichB
CorrA (only)
CorrB (only)
BetW
Count
Age
Very Young
Young
Medium
Old
correlation with HiC EV
enrichment in A compartment
correlation with H3K9me3
enrichment in AlwaysA
enriched in H3K9me3 peaks
enriched in srHC
enriched in both
NA
enrichment in B compartment
correlation with H3K27me3
enrichment in AlwaysB
TE
LTR
LINE
DNA
SINE
SVA
NA
non-TE
Low-complexity
Satellite
Simple repeat
NA
C
copy count
TE copy age (%)
All TE
CorrA TE
BetW TE
CorrB TE
CorrA.enrichA TE
CorrB.enrichB TE
Age category
Very young
Young
Medium
Old
TE Class
Subclass
LINE
CR1
L1
L2
others
ERV
ERV1
ERVK
ERVL
ERVL-MaLR
Gypsy
LTR
others
SINE
Alu
MIR /others
Composite
SVA
DNA
hAT
MuDR
TcMAR
others
D
SAT
Low complexity
Simple repeats
Satellite
ProB character
ProA character
CorrB
-0.01 < corr.HiC EV < 0
0 < corr.HiC EV < 0.01
CorrA
CorrB.enrichB
CorrA.enrichA
L1Hs
(TA)n
AT-rich
EnrichB
EnrichA
enrichment in B (log2)
enrichment in A (log2)
enr.B < 1.17
enr.A < 1.17
correlation with HiC EV
Alu
"age"
copy count
Jukes-Cantor distance (median)
L1
L1PA4
L1PA7
L1PA3
L1PA16
L1PA13
L1PA5
L1PA2
L1PA15
L1PA10
L1PA8
L1PA6
L1Hs
L1PA15-16
L1PA14
L1PB
L1MA10
L1PA12
L1PA17
L1MC2
L1MD
L1ME2
L1MD1
L1ME3A
L1ME3D

Fig. 3: Identifiying ProA and ProB RepSeq in the human genome.

(A) Top: All RepSeq subfamilies from RepBase ordered by Spearman correlation with the value of the HiC EV obtained with the HUVEC cell line, with bins of 25kb. Graph showing distribution of CorrA and CorrB RepSeq subsets, with correlation values above 0.01 and below 0.01, respectively. Pie charts at the bottom: Composition of the full complement of RepSeq in the human genome ("All RepSeq", hg19 version), and of CorrA and CorrB sets, according to RepSeq classes. The category "Others" includes snRNA, srpRNA, tRNA, rRNA, scRNA. Copy count in millions (M), length covered in % of total genome length (hg19). Note that "copy" refers to a segment of continuous homology with a subfamily consensus sequence. TE inserts frequently appear as a suite of more than one such segments (see Material and method section). See Extended Data Fig. 4 for a version of the figure including insert count.

(B) UMAP (Uniform Manifold Approximation and Projection) analysis of all RepSeq subfamilies from RepBase, based on the following parameters: age and count, enrichment (in A, in B, in AlwaysA, in AlwaysB); correlation (with HiC EV; with H3K9me3 ChIP-seq signal; with H3K27me3 ChIP-seq signal); 100 kb binning; data from HUVEC. Other panels highlight Repseq classes, and subsets of subfamilies as follows: subsets defined below for panel D using thresholds for correlation with HiC.EV and/or enrichment in A or B (larger panel at upper left); subsets enriched in biochemically defined fractions of the human genome (39), sonication-resistant heterochromatin (srHC), or H3K9me3 ChIP peaks in srHC-free regions.

(C,D) composition and characteristics of RepSeq sets.

(D, upper part) Each RepSeq subfamily is displayed as a point, with its color indicating TE class or nonTE ("SAT") class, on a Euclidian plane according to correlation with HiC EV and enrichment in compartments A (right) or B (left). Delimitations of CorrA and CorrB spaces, and EnrichA and EnrichB spaces are indicated on the respective axes. CorrA.enrichA and CorrB.enrichB spaces are highlighted with green and red boxes, respectively. Scheme upper right, ProA and ProB trends (see text). There is a strong overlap between CorrB.enrichB and L1 family, and between CorrA.enrichA and Alu family.

(C) TE copy count in RepSeq sets, displayed by subfamily (on the left in absolute counts) or age category (on the right as % of total count in the set considered). All TE, full complement of TE in the human genome. BetW RepSeq, correlation with HiC EV between -0.01 and 0.01. The median Jukes-Cantor distance (JCD) calculated as in Ref. (26) is used as a proxy for the age of a subfamily. Very young, JCD < 0.09; Young, 0.09 <JCD < 0.18; Medium 0.19 <JCD < 0.3; Old, JCD > 0.3. 0.09 was chosen to demarcate Young and Very Young TEs, because it is an intermediate value between the 0.08 median JCD of the L1PA7 subfamily, the "oldest" among the young L1 to display significant KAP1 binding at a subfamily level, as defined in ES cells (288), and that of the older L1PA8A (median JCD 0.10) which does not.

(D, lower part) Alu and L1 RepSeq subfamilies falling in the blue boxes noted 1 to 5 in the upper part are displayed as for their copy count and age (JCD median), ordered by increasing correlation with HiC EV, from left to right. The youngest L1 subfamilies are among the ones scoring best as CorrB.enrichB, whereas the youngest Alu subfamilies are not (Blue box2) or little (Blue box3) enriched in A, and the ones scoring best as CorrA.enrichA (Blue box5) are among the "less young" Alu (Alu elements and AluJ clades). Notably however, all Alu subfamilies are Young or Very young.

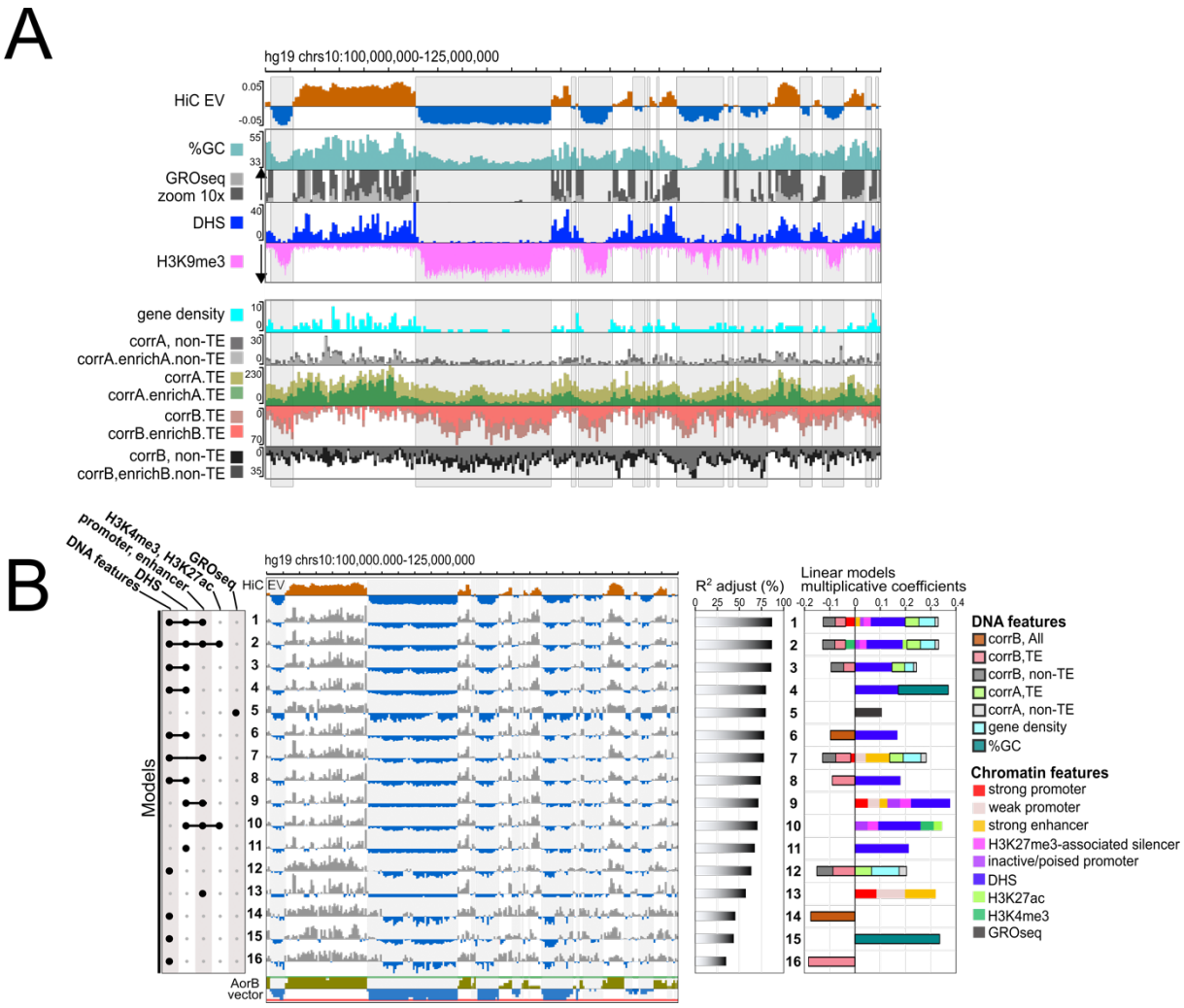


Fig. 4: CorrA and CorrB RepSeq local densities predict HiC EV.

(A) Browser view along the same 25 Mb region of chromosome 10 as in Fig. 1, for features as indicated. GROseq, signal summed over 100 kb; H3K9me3, ChIP signal; %GC, averaged over 100 kb; Gene and RepSeq sets, DHS, density over 100 kb.

(B) Linear models developed using combinations of DNA features and Chromatin features (left) recapitulate to some extent HUVEC HiC EV (middle) with a performance indicated by adjusted R^2 (middle right), expressed as % of the performance obtained by combining all features. Multiplicative coefficients in the linear equation (right) indicate the relative importance of the features used in the prediction (indicated by colors as in the legend, far right), as well the directionality, positive (resp. negative) coefficients being suggestive of a ProA (resp. ProB) role. Chromatin features (calculated for HUVEC cell line): ChromHMM segment count, for chromatin states associated with active promoters or enhancers, as well as for inactive/poised, H3K27me3-associated promoters; silencer count, for H3K27me3-associated DHS correlating with gene downregulation (284); H3K27ac and H3K4me3, peak density over 100 kb (see Resource and Methods section). Legend is otherwise the same as Fig. 4A.

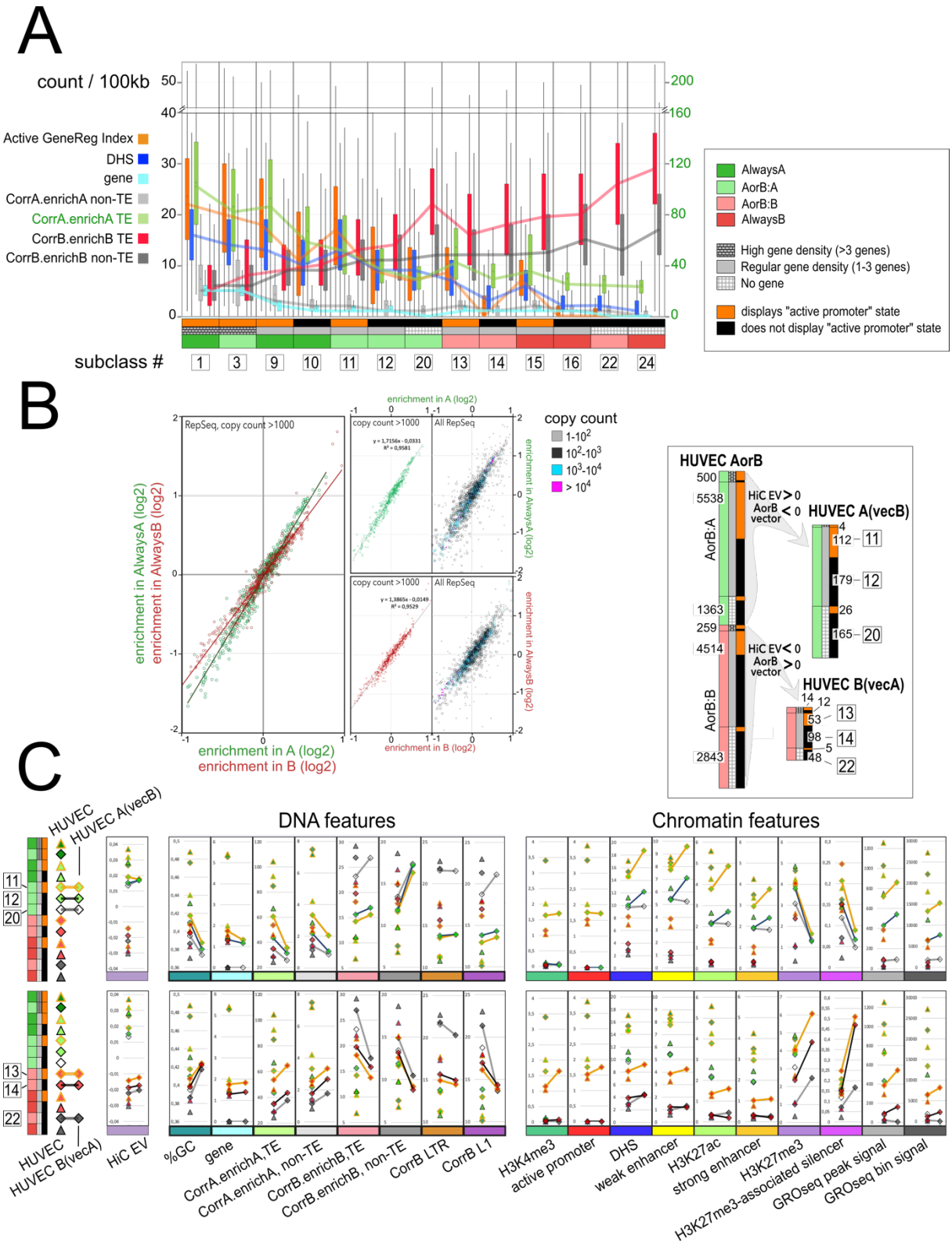


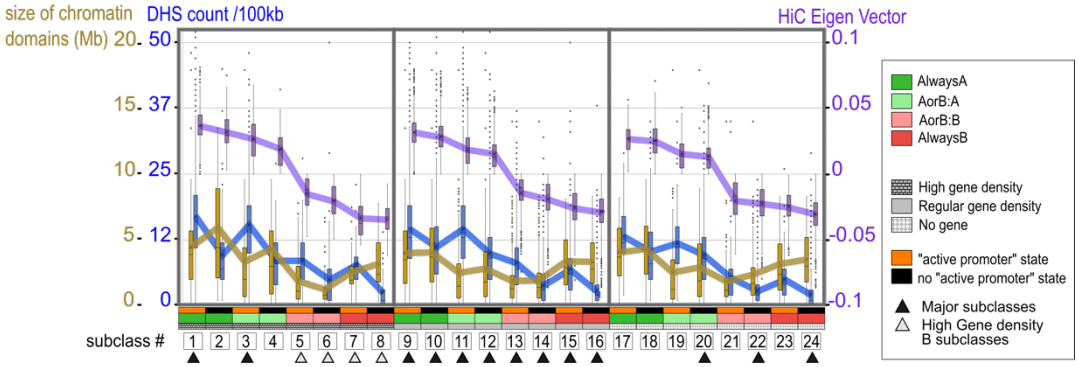
Fig. 5: Compositional genomics reveals general trends for the local densities of ProA and ProB elements, and an outlier composition for outlier bins.

(A) Boxplot representation of the density of DNA and chromatin features in the 13 major bin subclasses. Active GeneReg Index is a sum calculated from ChromHMM segment counts as follows: 0.5*WeakPromoter + 1.5*StrongPromoter + 0.5*WeakEnhancer + 1.5*StrongEnhancer. Lines connect successive median values. Note the scale is different for CorrA.enrichA TE density, as indicated in green on the right of the graph. For other details, see also legend to Fig.4.

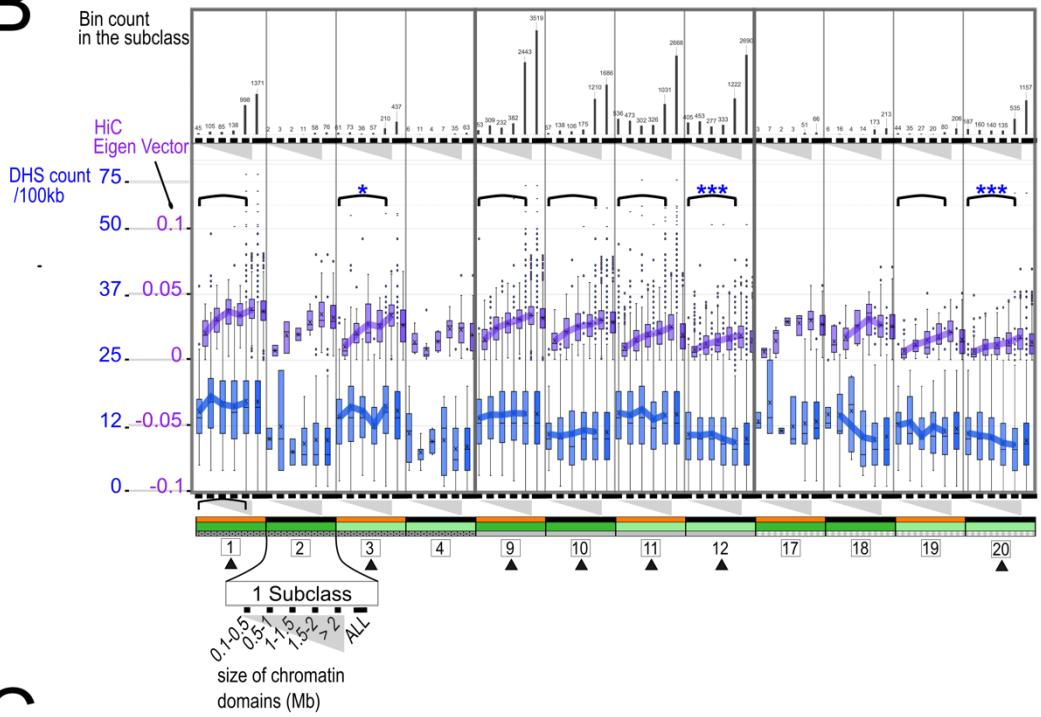
(B) Direct mathematical relationship between enrichment in AlwaysA and A, and AlwaysB and B, for each RepSeq subfamily.

Enrichment calculation is not reliable for subfamiliy counts below 1000 (right panels), which are discarded for the linear regression shown on the left with an overlay of the data points for copy counts >1000 from the central panels.
(C) Outlier AorB domains in HUVEC display a chimeric RepSeq composition. Outlier domains in HUVEC are AorB domains for which HiC EV has a sign opposite to AorBvec, meaning that the compartment "color" is different in HUVEC from the majority of the other seven reference cell lines shown in Fig. 1. Only "outstanding outlier" bins with the following characteristics were retained: (i) a marked difference between HiC EV and AorBvec (see Resources and Methods section); (ii) not found at an A/B junction (within 100 kb); (iii) part of domains with more than 2 successive bins showing such characteristics. Upper part, partitioning of AorB domains, Left, and of HUVEC A(vecB) and HUVEC B(vecA) outliers, Right. Only subclasses containing more than 40 bins are considered in the analysis. Lower part, mean values of HiC EV, DNA features, and chromatin features as indicated, calculated for individual bin subclasses and represented with the same symbols shown on the very left. Values are shown for all 13 major bin subclasses in HUVEC, and for the selected outlier subclasses, 3 HUVEC A(vecB) subclasses and 3 HUVEC B(vecA) subclasses.
The same symbols are used in both cases, connected with a line. For other details, see also legend to Fig. 4B.

A


size of chromatin domains (Mb)
DHS count /100kb
HiC Eigen Vector
subclass #
AlwaysA
AorB:A
AorB:B
AlwaysB
High gene density
Regular gene density
No gene
"active promoter" state
no "active promoter" state
Major subclasses
High Gene density B subclasses

B


Bin count in the subclass
HiC Eigen Vector
DHS count /100kb
1 Subclass
0.1-0.5
0.5-1
1-1.5
1.5-2
> 2
ALL
size of chromatin domains (Mb)

C

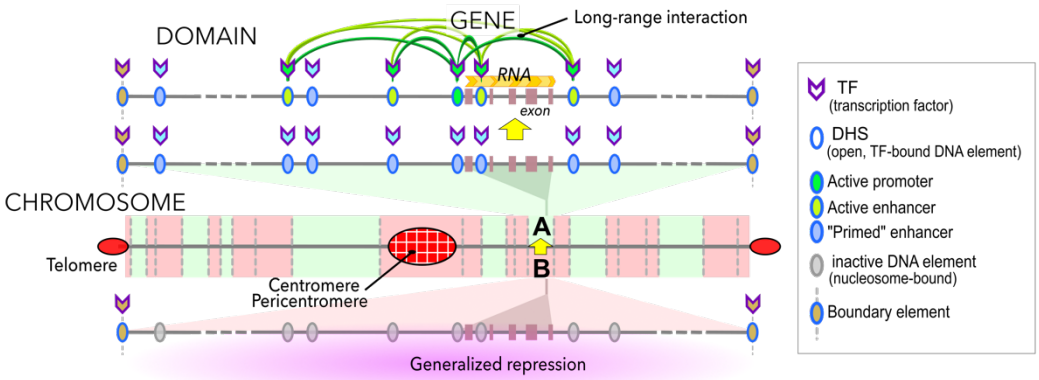

DOMAIN
GENE
Long-range interaction
RNA
exon
CHROMOSOME
Telomere
Centromere
Pericentromere
A
B
Generalized repression
TF (transcription factor)
DHS (open, TF-bound DNA element)
Active promoter
Active enhancer
"Primed" enhancer
inactive DNA element (nucleosome-bound)
Boundary element

Fig. 6: Evidence that DHS open domains.

(A) Boxplot showing size of chromatin domains (brown), DHS counts (blue), and HiC EV (purple) in the 24 bin subclasses. Note the different scales of the Y axes. Lines connect mean values.

(B) Boxplot as in A where each subclass is further subdivided in groups of bins according to the size of the domain in which the bin resides (lower part, zoomed out for subclass 2). Upper part: corresponding bin counts. Lower part: P-value calculated for comparison between domains with sizes comprised between 0.1 and 0.5 Mb with domains above 2Mb; statistical significance for this comparison is reached for subclasses 12 and 20 ($p < 0.05$, three asterisks) with bins that belong to smaller domains containing more DHSs on average than bins in larger domains, and for subclass 3 ($p<0.1$, one asterisk) in the other direction, i.e., larger domains displaying more DHSs. There are too few bins in subclasses 2,4,17 and 18 to assess the statistical significance (no parenthesis).

(C) DHS occurrence strictly correlates with domain opening. Schematical representation of a prototype gene, embedded in a prototype domain displaying either A or B state, itself embedded in a prototype chromosome, each displaying general trends.

Upper part: In a steady state situation, the gene is transcribed and the domain is part of the A compartment. Promoter activity depends on enhancers which are contained within the chromatin domain (although this is not always the case). The domain may be delimitated by clear boundaries, in which case these boundaries also depend on TFs. Although some TFs are specific for a given type of cis-element, there is overall a large overlap between TFs, which can perform distinct functions depending on which cofactors they associate with.

Lower part: Earlier in development, the domain is part of the B compartment. The DNA displays low accessibility to trans-acting factors, in particular to nucleases, and the gene is silent and its regulatory elements are occupied by nucleosomes, with chromatin marks largely indistinguishable from nondescript chromatin in the neighborhood ("generalized repression", also referred to as "closed configuration").

Middle: When development proceeds and expression of the gene is required for differentiation in a particular cell lineage, for instance, there is an intermediate stage preceding transcription when the domain adopts all characteristic features of the A compartment ("open configuration"). In particular, the domain unfolds in nuclear space and contacts within the domain increase at the expense of contacts with B domains in the vicinity, and generalized sensitivity to nucleases increases. This stage strongly correlates with the loss of nucleosomes at enhancers, which essentially appear as DHSs. This is enacted by TFs recruiting nucleosome remodeling machineries, with a key role for so-called pioneer TFs. Enhancer DNA generally shifts from methylated to unmethylated in the open state, as enacted and maintained by TET enzymes, with an intermediate, transient stage where 5hmC is detected. Loss of DNA methylation is assumed to aid in the DNA binding of TFs, but strikingly it appears not to be an absolute prerequisite for DNA binding of a large number of TFs, even factors known to be methylation-sensitive. It may occur either before, or concomitantly, or with some delay relative to the appearance of a DHS, the first option being characteristic of sites bound by certain pioneer factors such as Klf4 (99, 117, 251, 289, 290).Only in a distinct, final stage, coinciding with the onset of gene transcription, chromatin marks typical of active enhancers or active promoters appear at the corresponding DHS, while not at other DHS, which are dubbed "primed enhancers".

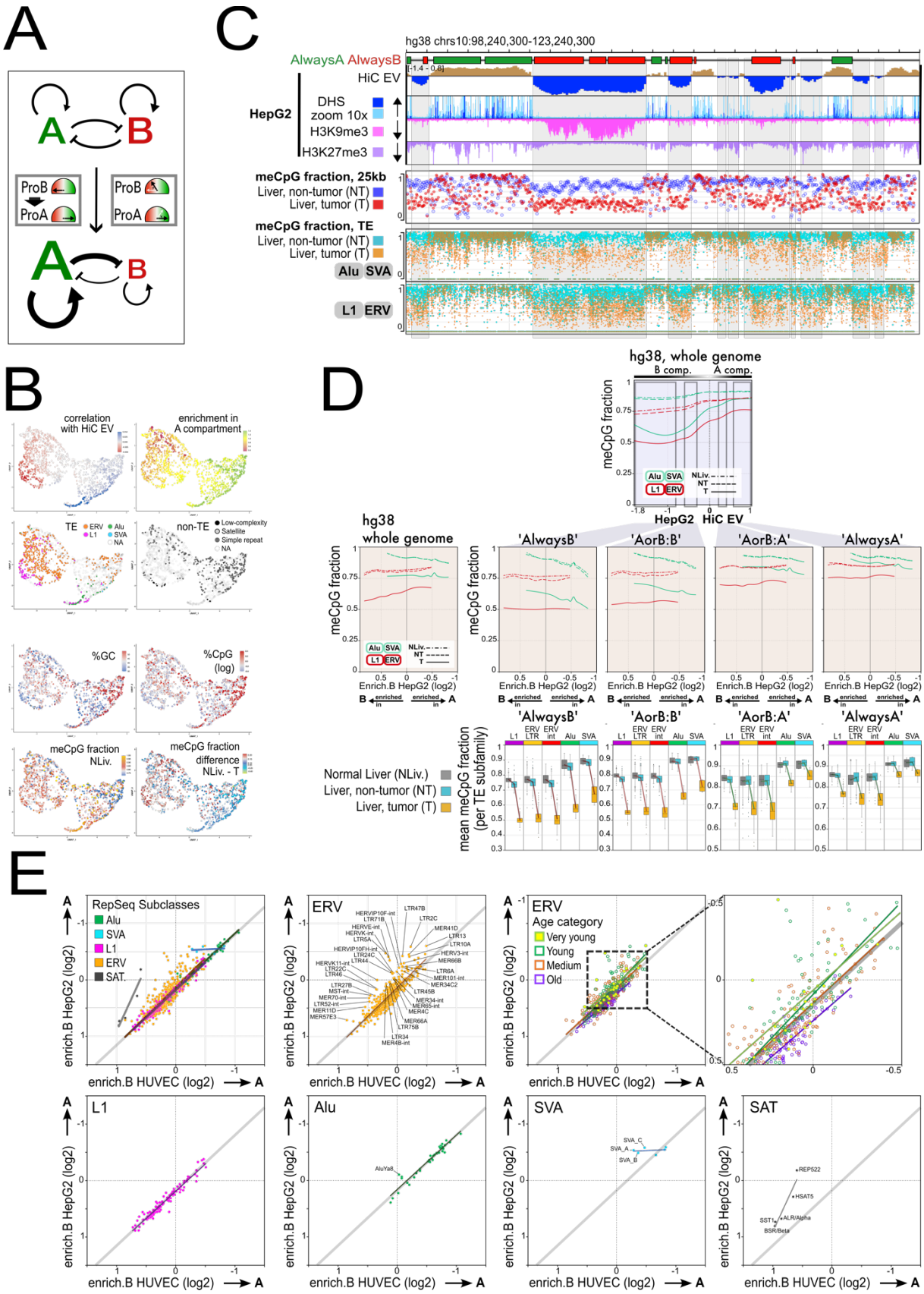

A
B
C
D
E
ProB
ProA
hg38 chrs10:98,240,300-123,240,300
AlwaysA AlwaysB
HiC EV
HepG2
DHS
zoom 10x
H3K9me3
H3K27me3
meCpG fraction, 25kb
Liver, non-tumor (NT)
Liver, tumor (T)
meCpG fraction, TE
Alu SVA
L1 ERV
correlation with HiC EV
enrichment in A compartment
TE
non-TE
%GC
%CpG (log)
meCpG fraction NLiv.
meCpG fraction difference NLiv. - T
hg38, whole genome
B comp.
A comp.
meCpG fraction
HepG2 HiC EV
hg38 whole genome
'AlwaysB'
'AorB:B'
'AorB:A'
'AlwaysA'
Enrich.B HepG2 (log2)
Normal Liver (NLiv.)
Liver, non-tumor (NT)
Liver, tumor (T)
mean meCpG fraction (per TE subfamily)
RepSeq Subclasses
Alu
SVA
L1
ERV
SAT
Age category
Very young
Young
Medium
Old
enrich.B HepG2 (log2)
enrich.B HUVEC (log2)

Fig. 7: The function of ProB RepSeqs is impaired in cancer, which drives genome opening.

(A) Modulating or switching RepSeqs from a ProB state to a ProA state, of the type found at active DHS/enhancers, tips the genome partitioning toggle-switch toward A, opening the genome at all scales. RepSeq inserts should be regarded as composites endowed each with both a ProA and a ProB potential,
as represented by a dial with green and red sides, respectively. Switching a number of RepSeqs along a chromatin segment from a ProB state, typically associated with H3K9me3 marks, to a ProA state, of the type found at active DHS/enhancers, or even just ramping up the ProA activity of a number of RepSeqs and/or decreasing the ProB capacity of others, tip the toggle-switch toward A, unfolding the segment. This principle applies at every scale of genome organization, ultimately resulting in whole genome opening.

B) Altered DNA methylation of RepSeq in cancer. UMAP analysis of all RepSeq subfamilies from RepBase as in Fig. 3B, based on the following parameters: enrichment (in A, in B, in AlwaysA, in Always B); correlation (with HiC EV; with H3K9me3 ChIP-seq signal; with H3K27me3 ChIP-seq signal); 100 kb binning; data from HUVEC; age and count categories. Four upper panels: RepSeq families are highlighted. Coloring according to enrichment and to correlation with ChIP signal for panels "enrichment in A compartment" and "correlation with HiC EV", which are also shown in Fig. 3B and replicated here as a reference. Four lower panels: Coloring according to mean %GC or mean % of CpG dinucleotides (CpG, log10 of %) calculated for each subfamily (upper panels), and to mean meCpG fraction values in a normal liver sample or meCpG loss in a liver tumor (lower panels). Mean meCpG fraction (per RepSeq subfamily): ratio of CpG dinucleotides with cytosine methylation to the total number of CpG dinucleotides over a RepSeq copy, as assessed for individual RepSeq copies by nanopore sequencing for a normal liver sample and averaged for each subfamily; a meCpG fraction difference was calculated by subtracting for each subfamily the mean meCpG fraction value obtained for a liver tumor (T) sample from the normal liver (NL) meCpG fraction (lower panel, right). Negative values correspond to increases in DNA methylation in the tumor.

(C) Most of the loss of CpG methylation in cancer occurs in the B compartment. Browser view of chromatin features and TE along the same 25 Mb region of chromosome 10 as in Fig. 1. The "AlwaysA AlwaysB" line is the same as in Fig. 1A, and highlights segments, which are always found embedded in either the A or the B compartment, respectively, as experimentally determined with eight cell lines. Four upper epigenomics tracks, chromatin features as indicated, assessed in the cell line HepG2, which was derived from a hepatocellular carcinoma. meCpG fraction, 25kb: ratio of CpG dinucleotides with cytosine methylation to the total number of CpG dinucleotides over a 25 kb chromosome segment, as assessed by nanopore sequencing for a liver tumor sample (T), and for a sample of adjacent, non-tumorous tissue (NT). meCpG fraction, TE: meCpG fraction similarly assessed for individual RepSeq copies by nanopore sequencing. Points indicate indifferently Alu or SVA copies in the upper track, and L1 or ERV copies in the lower track.

(D) TE methylation and its loss in cancer is mainly determined by the chromatin context. Upper panel: Each TE copy in the hg38 genome is represented as a point on a Euclidian plane according to the value of the HiC EV in HepG2 cells for the bin where it lies, and to its meCpG fraction in T and NT samples as above, and in a normal liver sample (NLiv.). Only regression curves are shown, with Alu and SVA copies, on the one hand, and L1 and ERV copies, on the other hand, analyzed together. Middle row: Each TE copy in the hg38 genome is represented as a point on a Euclidian plane according to the mean enrichment in the HepG2 B compartment for the TE subfamily it belongs to, and to its meCpG fraction in T, NT, and NLiv samples as above. Only regression curves are shown, with Alu and SVA copies, on the one hand, and L1 and ERV copies, on the other hand, analyzed together. This analysis was conducted for the full genome (left panel), or for four portions thereof shown each in one panel as indicated, assumed to belong to the chromatin classes AlwaysB ('AlwaysB', HiC EV < -0.8), AorB:B ('AorB:B', -0.6 < HiC EV < -0.3), AorB:A ('AorB:A', 0.2 < HiC EV < 0.4), or AlwaysA ('AlwaysA', > 0.55) on the basis of either extreme or intermediate HiC EV values . Note that enrichment in B is indicated in reverse order, such that subfamilies enriched in the A compartment fall on the right of the central 0 axis. Extended Data Fig. 9A shows the same curves overlaid with points indicating the mean meCpG fraction per TE subfamily.
Bottom row: boxplot representations of the mean meCpG fraction per TE subfamily in T, NT, and NLiv. samples as above, for L1, ERV, Alu and SVA subfamilies, in the same four genome portions as indicated. ERV sequences corresponding to LTR ("LTR") and to internal sequences ("int") were analyzed separately. Only subfamilies with more than 30 copies per genome portion considered were taken into account. See Extended Data Fig. 9C for a similar analysis according to age category.

(E) A number of young ERVs are more enriched in A compartment in cancer, suggestive of a shift in the balance between ProB and ProA functions toward ProA. Each RepSeq subfamily belonging to subclasses as indicated is represented as a point on a Euclidian plane according to the mean enrichment in the B compartment of HUVEC and HepG2 cells. Color code is the same for subclasses and trend lines. A thick grey line indicates the approximate position of the trend line for Alu, SVA and L1 altogether; it has a slope of 1 but is slightly offset from the diagonal. Only subfamilies with more than

200 copies are shown. All five RepSeq subclasses are shown together in the upper left panel, and individually in the other panels. Two upper right panels: the median age of ERV subfamilies is indicated by colors, as in Fig. 3, with the central field displayed at a higher magnification; the trend lines according to age are shown with the same color code. Only the names of those subfamilies are indicated for which the difference with respect to the grey trend line is higher than 0.2 in absolute log2 value, going in the A direction, and thus seemingly losing some ProB character and gaining more ProA character in HepG2 by comparison with HUVEC.

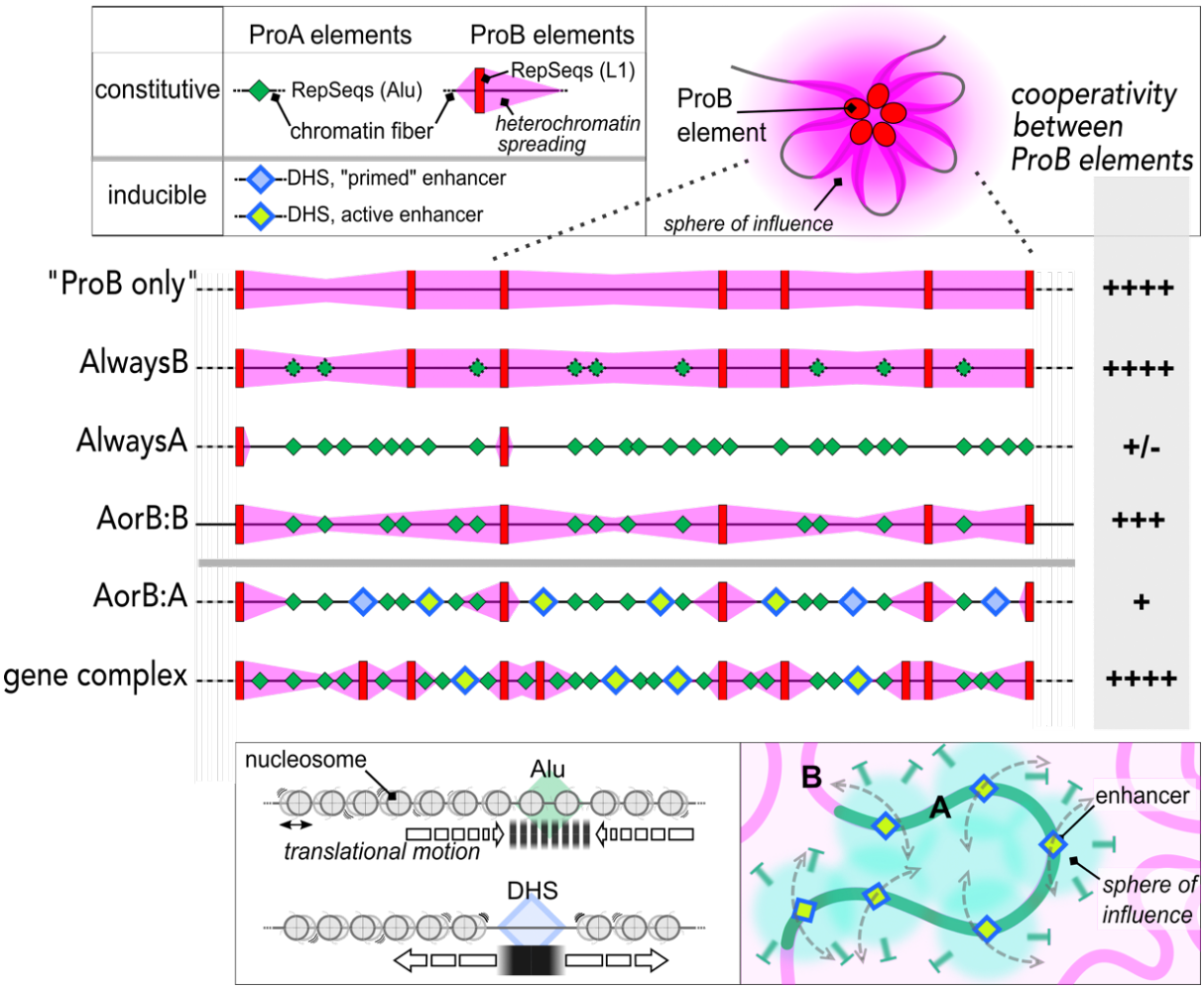


Fig. 8: General principles for how densities in ProA and ProB elements determine A/B partitioning.

Schematic drawing illustrating in an extremely simplified manner two main principles of genome organization by ProA and ProB elements: (i) ProB elements, typically RepSeqs such as young L1s, cooperate to establish a B logic over a domain; (ii) ProA, both RepSeqs, typically Alu elements, and inducible ProA, such as DHSs and enhancers, break the cooperativity between ProB elements to install an A logic (note that active genes, not shown in the scheme, can also do that). They do so indirectly, by altering the chromatin fiber in a way that antagonizes phenomena that are crucial for heterochromatin function, and therefore for the formation and maintenance of the B compartment, which otherwise occurs as a default state over a domain due to the presence of ProB elements. It has already been proposed and validated by modeling that such a limited toolkit of ProA and ProB DNA elements and set of rules form the framework of genome organization and is responsible for the toggle-switch scheme illustrated in Fig. 2 (21, 230).

Top: Upper part, left: ProB elements are nucleation points for heterochromatin, from which it seemingly "spreads" along the chromatin fiber. Pink triangles indicate the extent of heterochromatin spreading away from a ProB element, i.e. segments of the chromatin fiber where: (i) high levels of heterochromatin marks such as H3K9me3 can be detected by ChIP; (ii) transcription is markedly inhibited. Both features are also hallmarks of the B compartment, which therefore appears as regions of the genome dominated by heterochromatin spreading. Spreading depends both on factors diffusing away from a ProB element in 3D and on stochastic events where individual loci along the fiber interact with the ProB element (291), both mechanisms being all the more efficient in the proximity to a ProB element. Spreading is also thought to depend on the translational motions of nucleosomes within a fiber. Upper part, right: ProB elements in a domain stochastically interact in 3D, which is stabilized by heterochromatin-associated coalescence forces. This in turn reinforces their ability to assemble and anchor heterochromatin, and forms the basis of functional cooperativity between ProBs (adapted from (17)). The probability of presence of heterochromatin factors ("heterochromatin ambiance") is heightened in a nuclear subvolume around such dynamic structures ("sphere of influence"), and in turn around each ProB element in a manner that is strongly dependent on the frequency of long-range interactions with other ProB elements in the neighborhood.

The chromatin fiber itself is an integral player in the ProB system, through the phenomenon of nucleosome coalescence, which is powered by structural factors of heterochromatin and is at the heart of heterochromatin spreading. Nucleosome modifications such as histone acetylation can inhibit nucleosome coalescence, while nucleosome partners, such as histone H1, can promote it, and therefore confer to the chromatin fiber the capacity to ultimately interfere with, or promote, the long-range cooperation between ProB elements along a domain. In addition, the chromatin fiber contributes heterochromatin ambiance with factors that dissociate from the fiber. When heterochromatin spreading is unimpeded, such as within a hypothetical B locus harboring only ProB elements ("ProB only"), heterochromatin spreads along the entire length of the chromatin segment embedded within the sphere of high influence, such that spreading appears strictly continuous, forming a strong B compartment locus.

Bottom: Hypothetical molecular mechanisms by which three types of ProA elements are envisioned to alter the chromatin fiber, thereby interfering with heterochromatin spreading. Both Alu TEs and DHS are thought to hinder the translational motion of nucleosomes, strongly in the case of DHS due to an interruption in the nucleosomal fiber, whereas two strongly positioned nucleosomes over an individual Alu element conceivably represent a milder impediment. CpG methylation is believed to promote nucleosome positioning (162, 267), and thereby the ProA function of Alu elements. Notably, active DNA demethylation as generally detected at DHSs shortly after the detection of DHSs themselves may therefore have as its main purpose to facilitate nucleosome displacement. Active enhancers display a sphere of influence enriched in transcription cofactors many of which actually oppose the B logics. This is true in particular for at least two classes of chromatin modifiers which can make a chromatin fiber highly resistant to heterochromatin spreading: histone acetyl-transferases (HATs), responsible for histone acetylation, and Poly (ADP-ribose) Polymerase 1 (PARP1) (181, 191). The latter parylates chromatin components and in particular histone H1, resulting in its inactivation (292). Whereas "primed" enhancers distributed along a domain are sufficient to switch a B domain to A, "active enhancer"-associated HATs are envisioned to secure the A character by cooperatively repelling B influence over a domain, through diffusively opposing histone deacetylases (HDACs). The latter are enriched in heterochromatin ambiance and predominate in the human nucleus, where they promote nucleosome coalescence and therefore heterochromatin sprading (212, 293). This mechanism clearly also applies to some extent to enhancers in their "primed" state, which is associated with a range of similar histone modifications as "active enhancers" except that histone acetylation levels are much lower. In addition, the natural substrate for HP1α binding is an H1-containing nucleosomal fiber (35), and therefore inactivation of H1 by PARP1 is predicted to independently hinder heterochromatin spreading.

Middle: In AlwaysB regions, a high density of ProB RepSeqs is found together with a low density of ProA. Alu elements are envisioned to mildly mitigate spreading, leaving cooperativity between ProBs essentially unaltered, hence generating a strong B character. Notably, Alu elements have their ProA function at least partially inactivated in AlwaysB due to a loss of CpG methylation (green diamonds with dotted contour). In AlwaysA regions, proportions are reversed, and, in addition, there are active genes (not shown in the scheme). Cooperativity between ProBs is essentially absent, because of their low density, and because heterochromatin spreading is crippled. In AorB regions, intermediate densities of both ProA and ProB RepSeqs allow for strong cooperativity between ProBs while spreading is mitigated to a degree. It may be further inhibited by the emergence of inducible ProAs, switching the domain from B to A. Notably, ProB themselves, in particular ERVs, may switch to ProA, and it is likely that sequences presently considered merely as enhancers actually switch broadly from a ProB to a ProA state (not shown in the scheme for the sake of simplicity). Large regions of gene complexes are embedded in the B compartment and characteristically display a high density of both ProA and ProB RepSeqs, allowing both a marked cooperation between ProBs and the emergence of restricted, transcription-conducive domains.

REFERENCES

1. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326:289-93.
2. Harris HL, Gu H, Olshansky M, Wang A, Farabella I, Eliaz Y, Kalluchi A, Krishna A, Jacobs M, Cauer G, Pham M, Rao SSP, Dudchenko O, Omer A, Mohajeri K, Kim S, Nichols MH, Davis ES, Gkountaroulis D, Udupa D, Aiden AP, Corces VG, Phanstiel DH, Noble WS, Nir G, Di Pierro M, Seo JS, Talkowski ME, Aiden EL, Rowley MJ. 2023. Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. Nat Commun 14:3303.
3. Brown JM, Roberts NA, Graham B, Waithe D, Lagerholm C, Telenius JM, De Ornellas S, Oudelaar AM, Scott C, Szczerbal I, Babbs C, Kassouf MT, Hughes JR, Higgs DR, Buckle VJ. 2018. A tissue-specific self-interacting chromatin domain forms independently of enhancer-promoter interactions. Nat Commun 9:3849.
4. Haws SA, Simandi Z, Barnett RJ, Phillips-Cremins JE. 2022. 3D genome, on repeat: Higher-order folding principles of the heterochromatinized repetitive genome. Cell 185:2690-2707.
5. Ye Y, Zhang S, Gao L, Zhu Y, Zhang J. 2023. Deciphering Hierarchical Chromatin Domains and Preference of Genomic Position Forming Boundaries in Single Mouse Embryonic Stem Cells. Adv Sci (Weinh) doi:10.1002/advs.202205162:e2205162.
6. Barth R, Fourel G, Shaban HA. 2020. Dynamics as a cause for the nanoscale organization of the genome. Nucleus 11:83-98.
7. Keizer VIP, Grosse-Holz S, Woringer M, Zambon L, Aizel K, Bongaerts M, Delille F, Kolar-Znika L, Scolari VF, Hoffmann S, Banigan EJ, Mirny LA, Dahan M, Fachinetti D, Coulon A. 2022. Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics. Science 377:489-495.
8. Gabriele M, Brandao HB, Grosse-Holz S, Jha A, Dailey GM, Cattoglio C, Hsieh TS, Mirny L, Zechner C, Hansen AS. 2022. Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. Science 376:496-501.
9. Mach P, Kos PI, Zhan Y, Cramard J, Gaudin S, Tunnermann J, Marchi E, Eglinger J, Zuin J, Kryzhanovska M, Smallwood S, Gelman L, Roth G, Nora EP, Tiana G, Giorgetti L. 2022. Cohesin and CTCF control the dynamics of chromosome folding. Nat Genet 54:1907-1918.
10. Davidson IF, Barth R, Zaczek M, van der Torre J, Tang W, Nagasaka K, Janissen R, Kerssemakers J, Wutz G, Dekker C, Peters JM. 2023. CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. Nature 616:822-827.
11. Mirny LA, Imakaev M, Abdennur N. 2019. Two major mechanisms of chromosome organization. Curr Opin Cell Biol 58:142-152.
12. Belaghzal H, Borrman T, Stephens AD, Lafontaine DL, Venev SV, Weng Z, Marko JF, Dekker J. 2021. Liquid chromatin Hi-C characterizes compartment-dependent chromatin interaction dynamics. Nat Genet 53:367-378.
13. Ugarte F, Sousae R, Cinquin B, Martin EW, Krietsch J, Sanchez G, Inman M, Tsang H, Warr M, Passegue E, Larabell CA, Forsberg EC. 2015. Progressive Chromatin Condensation and H3K9 Methylation Regulate the Differentiation of Embryonic and Hematopoietic Stem Cells. Stem Cell Reports 5:728-740.
14. Magnitov MD, Garaev AK, Tyakht AV, Ulianov SV, Razin SV. 2022. Pentad: a tool for distance-dependent analysis of Hi-C interactions within and between chromatin compartments. BMC Bioinformatics 23:116.
15. Rodriguez-Terrones D, Torres-Padilla ME. 2018. Nimble and Ready to Mingle: Transposon Outbursts of Early Development. Trends Genet 34:806-820.
16. Guthmann M, Qian C, Gialdini I, Nakatani T, Ettinger A, Schauer T, Kukhtevich I, Schneider R, Lamb DC, Burton A, Torres-Padilla ME. 2023. A change in biophysical properties accompanies heterochromatin formation in mouse embryos. Genes Dev doi:10.1101/gad.350353.122.
17. Fourel G, Lebrun E, Gilson E. 2002. Protosilencers as building blocks for heterochromatin. Bioessays 24:828-35.
18. Sanulli S, Gross JD, Narlikar GJ. 2019. Biophysical Properties of HP1-Mediated Heterochromatin. Cold Spring Harb Symp Quant Biol 84:217-225.
19. Haddad N, Jost D, Vaillant C. 2017. Perspectives: using polymer modeling to understand the formation and function of nuclear compartments. Chromosome Res 25:35-50.
20. Falk M, Feodorova Y, Naumova N, Imakaev M, Lajoie BR, Leonhardt H, Joffe B, Dekker J, Fudenberg G, Solovei I, Mirny LA. 2019. Heterochromatin drives compartmentalization of inverted and conventional nuclei. Nature 570:395-399.
21. Fourel G, Magdinier F, Gilson E. 2004. Insulator dynamics and the setting of chromatin domains. Bioessays 26:523-32.
22. Rada-Iglesias A, Grosveld FG, Papantonis A. 2018. Forces driving the three-dimensional folding of eukaryotic genomes. Mol Syst Biol 14:e8214.
23. Bailey LT, Northall SJ, Schalch T. 2021. Breakers and amplifiers in chromatin circuitry: acetylation and ubiquitination control the heterochromatin machinery. Curr Opin Struct Biol 71:156-163.
24. Bell O, Burton A, Dean C, Gasser SM, Torres-Padilla ME. 2023. Heterochromatin definition and function. Nat Rev Mol Cell Biol doi:10.1038/s41580-023-00599-7.
25. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473:43-9.
26. Pehrsson EC, Choudhary MNK, Sundaram V, Wang T. 2019. The epigenomic landscape of transposable elements across normal human development and anatomy. Nat Commun 10:5640.
27. Carron L, Concia L, Grob S, Barneche F, Carbone A, Mozziconacci J. 2023. A layer cake model for plant and metazoan chromatin. bioRxiv doi:10.1101/2023.02.24.529851:2023.02.24.529851.
28. Shopland LS, Lynch CR, Peterson KA, Thornton K, Kepper N, Hase J, Stein S, Vincent S, Molloy KR, Kreth G, Cremer C, Bult CJ, O'Brien TP. 2006. Folding and organization of a contiguous chromosome region according to the gene distribution pattern in primary genomic sequence. J Cell Biol 174:27-38.
29. van de Werken HJG, Haan JC, Feodorova Y, Bijos D, Weuts A, Theunis K, Holwerda SJB, Meuleman W, Pagie L, Thanisch K, Kumar P, Leonhardt H, Marynen P, van Steensel B, Voet T, de Laat W, Solovei I, Joffe B. 2017. Small chromosomal regions position themselves autonomously according to their chromatin class. Genome Res 27:922-933.
30. Jabbari K, Bernardi G. 2017. An Isochore Framework Underlies Chromatin Architecture. PLoS One 12:e0168023.

31. Beaufay F, Amemiya HM, Guan J, Basalla J, Meinen BA, Chen Z, Mitra R, Bardwell JCA, Biteen JS, Vecchiarelli AG, Freddolino PL, Jakob U. 2021. Polyphosphate drives bacterial heterochromatin formation. Sci Adv 7:eabk0233.
32. Larson AG, Elnatan D, Keenen MM, Trnka MJ, Johnston JB, Burlingame AL, Agard DA, Redding S, Narlikar GJ. 2017. Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. Nature 547:236-240.
33. Chapard C, Meneu L, Serizay J, Routhier E, Ruault M, Bignaud A, Gourgues G, Lartigue C, Piazza A, Taddei A, Beckouët F, Mozziconacci J, Koszul R. 2023. Exogenous chromosomes reveal how sequence composition drives chromatin assembly, activity, folding and compartmentalization. bioRxiv doi:10.1101/2022.12.21.520625:2022.12.21.520625.
34. Lensch S, Herschl MH, Ludwig CH, Sinha J, Hinks MM, Mukund A, Fujimori T, Bintu L. 2022. Dynamic spreading of chromatin-mediated gene silencing and reactivation between neighboring genes in single cells. Elife 11.
35. Bryan LC, Weilandt DR, Bachmann AL, Kilic S, Lechner CC, Odermatt PD, Fantner GE, Georgeon S, Hantschel O, Hatzimanikatis V, Fierz B. 2017. Single-molecule kinetic analysis of HP1-chromatin binding reveals a dynamic network of histone modification and DNA interactions. Nucleic Acids Res 45:10504-10517.
36. Biswas S, Chen Z, Karslake JD, Farhat A, Ames A, Raiymbek G, Freddolino PL, Biteen JS, Ragunathan K. 2022. HP1 oligomerization compensates for low-affinity H3K9me recognition and provides a tunable mechanism for heterochromatin-specific localization. Sci Adv 8:eabk0793.
37. Quante T, Bird A. 2016. Do short, frequent DNA sequence motifs mould the epigenome? Nat Rev Mol Cell Biol 17:257-62.
38. Kuwayama N, Kujirai T, Kishi Y, Hirano R, Echigoya K, Fang L, Watanabe S, Nakao M, Suzuki Y, Ishiguro KI, Kurumizaka H, Gotoh Y. 2023. HMGA2 directly mediates chromatin condensation in association with neuronal fate regulation. Nat Commun 14:6420.
39. Becker JS, McCarthy RL, Sidoli S, Donahue G, Kaeding KE, He Z, Lin S, Garcia BA, Zaret KS. 2017. Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. Mol Cell 68:1023-1037 e15.
40. Kouzine F, Wojtowicz D, Baranello L, Yamane A, Nelson S, Resch W, Kieffer-Kwon KR, Benham CJ, Casellas R, Przytycka TM, Levens D. 2017. Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome. Cell Syst 4:344-356 e7.
41. Herbert A. 2020. Simple Repeats as Building Blocks for Genetic Computers. Trends Genet 36:739-750.
42. Senturk Cetin N, Kuo CC, Ribarska T, Li R, Costa IG, Grummt I. 2019. Isolation and genome-wide characterization of cellular DNA:RNA triplex structures. Nucleic Acids Res 47:2306-2321.
43. Zhang X, Jiang Q, Li J, Zhang S, Cao Y, Xia X, Cai D, Tan J, Chen J, Han JJ. 2022. KCNQ1OT1 promotes genome-wide transposon repression by guiding RNA-DNA triplexes and HP1 binding. Nat Cell Biol doi:10.1038/s41556-022-01008-5.
44. Grapotte M, Saraswat M, Bessiere C, Menichelli C, Ramilowski JA, Severin J, Hayashizaki Y, Itoh M, Tagami M, Murata M, Kojima-Ishiyama M, Noma S, Noguchi S, Kasukawa T, Hasegawa A, Suzuki H, Nishiyori-Sueki H, Frith MC, consortium F, Chatelain C, Carninci P, de Hoon MJL, Wasserman WW, Brehelin L, Lecellier CH. 2021. Discovery of widespread transcription initiation at microsatellites predictable by sequence-based deep neural network. Nat Commun 12:3297.
45. Maag JL, Panja D, Sporild I, Patil S, Kaczorowski DC, Bramham CR, Dinger ME, Wibrand K. 2015. Dynamic expression of long noncoding RNAs and repeat elements in synaptic plasticity. Front Neurosci 9:351.
46. Crossley MP, Song C, Bocek MJ, Choi JH, Kousorous J, Sathirachinda A, Lin C, Brickner JR, Bai G, Lans H, Vermeulen W, Abu-Remaileh M, Cimprich KA. 2022. R-loop-derived cytoplasmic RNA-DNA hybrids activate an immune response. Nature doi:10.1038/s41586-022-05545-9.
47. Horton CA, Alexandari AM, Hayes MGB, Marklund E, Schaepe JM, Aditham AK, Shah N, Suzuki PH, Shrikumar A, Afek A, Greenleaf WJ, Gordan R, Zeitlinger J, Kundaje A, Fordyce PM. 2023. Short tandem repeats bind transcription factors to tune eukaryotic gene expression. Science 381:eadd1250.
48. Herbert A. 2019. Z-DNA and Z-RNA in human disease. Commun Biol 2:7.
49. Jiao H, Wachsmuth L, Kumari S, Schwarzer R, Lin J, Eren RO, Fisher A, Lane R, Young GR, Kassiotis G, Kaiser WJ, Pasparakis M. 2020. Z-nucleic-acid sensing triggers ZBP1-dependent necroptosis and inflammation. Nature 580:391-395.
50. Gambelli A, Ferrando A, Boncristiani C, Schoeftner S. 2023. Regulation and function of R-loops at repetitive elements. Biochimie doi:10.1016/j.biochi.2023.08.013.
51. Liu Y, Liu Q, Su H, Liu K, Xiao X, Li W, Sun Q, Birchler JA, Han F. 2021. Genome-wide mapping reveals R-loops associated with centromeric repeats in maize. Genome Res 31:1409-1418.
52. Zhou J, Lei X, Shafiq S, Zhang W, Li Q, Li K, Zhu J, Dong Z, He XJ, Sun Q. 2023. DDM1-mediated R-loop resolution and H2A.Z exclusion facilitates heterochromatin formation in Arabidopsis. Sci Adv 9:eadg2699.
53. Niehrs C, Luke B. 2020. Regulatory R-loops as facilitators of gene expression and genome stability. Nat Rev Mol Cell Biol 21:167-178.
54. Moore BL, Aitken S, Semple CA. 2015. Integrative modeling reveals the principles of multi-scale chromatin boundary formation in human nuclear organization. Genome Biol 16:110.
55. Cao Q, Zhang Z, Fu AX, Wu Q, Lee T-L, Lo E, Cheng ASL, Cheng C, Leung D, Yip KY. 2020. A unified framework for integrative study of heterogeneous gene regulatory mechanisms. Nature Machine Intelligence 2:447-456.
56. Imbeault M, Helleboid PY, Trono D. 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature 543:550-554.
57. Sun X, Wang X, Tang Z, Grivainis M, Kahler D, Yun C, Mita P, Fenyo D, Boeke JD. 2018. Transcription factor profiling reveals molecular choreography and key regulators of human retrotransposon expression. Proc Natl Acad Sci U S A 115:E5526-E5535.
58. Pontis J, Planet E, Offner S, Turelli P, Duc J, Coudray A, Theunissen TW, Jaenisch R, Trono D. 2019. Hominoid-Specific Transposable Elements and KZFPs Facilitate Human Embryonic Genome Activation and Control Transcription in Naive Human ESCs. Cell Stem Cell 24:724-735 e5.
59. Pontis J, Pulver C, Playfoot CJ, Planet E, Grun D, Offner S, Duc J, Manfrin A, Lutolf MP, Trono D. 2022. Primate-specific transposable elements shape transcriptional networks during human development. Nat Commun 13:7178.
60. Ernst J, Melnikov A, Zhang X, Wang L, Rogov P, Mikkelsen TS, Kellis M. 2016. Genome-scale high-resolution mapping of activating and repressive nucleotides in regulatory regions. Nat Biotechnol 34:1180-1190.
61. Varshney D, Vavrova-Anderson J, Oler AJ, Cowling VH, Cairns BR, White RJ. 2015. SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. Nat Commun 6:6569.

62. Bouttier M, Laperriere D, Memari B, Mangiapane J, Fiore A, Mitchell E, Verway M, Behr MA, Sladek R, Barreiro LB, Mader S, White JH. 2016. Alu repeats as transcriptional regulatory platforms in macrophage responses to M. tuberculosis infection. Nucleic Acids Res 44:10571-10587.
63. Policarpi C, Crepaldi L, Brookes E, Nitarska J, French SM, Coatti A, Riccio A. 2017. Enhancer SINEs Link Pol III to Pol II Transcription in Neurons. Cell Rep 21:2879-2894.
64. Ferrari R, de Llobet Cucalon LI, Di Vona C, Le Dilly F, Vidal E, Lioutas A, Oliete JQ, Jochem L, Cutts E, Dieci G, Vannini A, Teichmann M, de la Luna S, Beato M. 2020. TFIIIC Binding to Alu Elements Controls Gene Expression via Chromatin Looping and Histone Acetylation. Mol Cell 77:475-487 e11.
65. Ohtani H, Liu M, Zhou W, Liang G, Jones PA. 2018. Switching roles for DNA and histone methylation depend on evolutionary ages of human endogenous retroviruses. Genome Res 28:1147-1157.
66. de Tribolet-Hardy J, Thorball CW, Forey R, Planet E, Duc J, Coudray A, Khubieh B, Offner S, Pulver C, Fellay J, Imbeault M, Turelli P, Trono D. 2023. Genetic features and genomic targets of human KRAB-zinc finger proteins. Genome Res doi:10.1101/gr.277722.123.
67. Medstrand P, van de Lagemaat LN, Mager DL. 2002. Retroelement distributions in the human genome: variations associated with age and proximity to genes. Genome Res 12:1483-95.
68. Fuentes DR, Swigut T, Wysocka J. 2018. Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. Elife 7.
69. Pal D, Patel M, Boulet F, Sundarraj J, Grant OA, Branco MR, Basu S, Santos S, Zabet NR, Scaffidi P, Pradeepa MM. 2022. H4K16ac activates the transcription of transposable elements and contributes to their cis-regulatory function. bioRxiv doi:10.1101/2022.04.29.488986:2022.04.29.488986.
70. Fueyo R, Judd J, Feschotte C, Wysocka J. 2022. Roles of transposable elements in the regulation of mammalian transcription. Nat Rev Mol Cell Biol 23:481-497.
71. Carter TA, Singh M, Dumbovic G, Chobirko JD, Rinn JL, Feschotte C. 2022. Mosaic cis-regulatory evolution drives transcriptional partitioning of HERVH endogenous retrovirus in the human embryo. Elife 11.
72. Karttunen K, Patel D, Xia J, Fei L, Palin K, Aaltonen L, Sahu B. 2023. Transposable elements as tissue-specific enhancers in cancers of endodermal lineage. Nat Commun 14:5313.
73. Costello KR, Leung A, Trac C, Lee M, Basam M, Pospisilik JA, Schones DE. 2021. Sequence features of retrotransposons allow for epigenetic variability. Elife 10.
74. Zemke NR, Armand EJ, Wang W, Lee S, Zhou J, Li YE, Liu H, Tian W, Nery JR, Castanon RG, Bartlett A, Osteen JK, Li D, Zhuo X, Xu V, Miller M, Krienen FM, Zhang Q, Taskin N, Ting J, Feng G, McCarroll SA, Callaway EM, Wang T, Behrens MM, Lein ES, Ecker JR, Ren B. 2023. Comparative single cell epigenomic analysis of gene regulatory programs in the rodent and primate neocortex. bioRxiv doi:10.1101/2023.04.08.536119.
75. Bancaud A, Huet S, Daigle N, Mozziconacci J, Beaudouin J, Ellenberg J. 2009. Molecular crowding affects diffusion and binding of nuclear proteins in heterochromatin and reveals the fractal organization of chromatin. EMBO J 28:3785-98.
76. Polovnikov KE, Slavov B, Belan S, Imakaev M, Brandão HB, Mirny LA. 2023. Crumpled polymer with loops recapitulates key features of chromosome organization. bioRxiv doi:10.1101/2022.02.01.478588:2022.02.01.478588.
77. Mandelbrot BB, Van Ness JW. 1968. Fractional Brownian Motions, Fractional Noises and Applications. SIAM Review 10:10.1137/1010093.
78. Strogatz SH. 2018. Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. CRC Press.
79. Mito Y, Henikoff JG, Henikoff S. 2007. Histone replacement marks the boundaries of cis-regulatory domains. Science 315:1408-11.
80. Vierstra J, Lazar J, Sandstrom R, Halow J, Lee K, Bates D, Diegel M, Dunn D, Neri F, Haugen E, Rynes E, Reynolds A, Nelson J, Johnson A, Frerker M, Buckley M, Kaul R, Meuleman W, Stamatoyannopoulos JA. 2020. Global reference mapping of human transcription factor footprints. Nature 583:729-736.
81. Lazar JE, Stehling-Sun S, Nandakumar V, Wang H, Chee DR, Howard NP, Acosta R, Dunn D, Diegel M, Neri F, Castillo A, Ibarrientos S, Lee K, Lescano N, Van Biber B, Nelson J, Halow J, Sandstrom R, Bates D, Urnov FD, Stamatoyannopoulos JA, Funnell APW. 2020. Global Regulatory DNA Potentiation by SMARCA4 Propagates to Selective Gene Expression Programs via Domain-Level Remodeling. Cell Rep 31:107676.
82. Comoglio F, Simonatto M, Polletti S, Liu X, Smale ST, Barozzi I, Natoli G. 2019. Dissection of acute stimulus-inducible nucleosome remodeling in mammalian cells. Genes Dev 33:1159-1174.
83. Paun O, Tan YX, Patel H, Strohbuecker S, Ghanate A, Cobolli-Gigli C, Llorian Sopena M, Gerontogianni L, Goldstone R, Ang SL, Guillemot F, Dias C. 2023. Pioneer factor ASCL1 cooperates with the mSWI/SNF complex at distal regulatory elements to regulate human neural differentiation. Genes Dev 37:218-242.
84. Michael AK, Stoos L, Crosby P, Eggers N, Nie XY, Makasheva K, Minnich M, Healy KL, Weiss J, Kempf G, Cavadini S, Kater L, Seebacher J, Vecchia L, Chakraborty D, Isbel L, Grand RS, Andersch F, Fribourgh JL, Schubeler D, Zuber J, Liu AC, Becker PB, Fierz B, Partch CL, Menet JS, Thoma NH. 2023. Cooperation between bHLH transcription factors and histones for DNA access. Nature doi:10.1038/s41586-023-06282-3.
85. Meuleman W, Muratov A, Rynes E, Halow J, Lee K, Bates D, Diegel M, Dunn D, Neri F, Teodosiadis A, Reynolds A, Haugen E, Nelson J, Johnson A, Frerker M, Buckley M, Sandstrom R, Vierstra J, Kaul R, Stamatoyannopoulos J. 2020. Index and biological spectrum of human DNase I hypersensitive sites. Nature 584:244-251.
86. Francastel C, Walters MC, Groudine M, Martin DI. 1999. A functional enhancer suppresses silencing of a transgene and prevents its localization close to centrometric heterochromatin. Cell 99:259-69.
87. Schubeler D, Francastel C, Cimbora DM, Reik A, Martin DI, Groudine M. 2000. Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. Genes Dev 14:940-50.
88. Therizols P, Illingworth RS, Courilleau C, Boyle S, Wood AJ, Bickmore WA. 2014. Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. Science 346:1238-42.
89. Karagianni P, Moulos P, Schmidt D, Odom DT, Talianidis I. 2020. Bookmarking by Non-pioneer Transcription Factors during Liver Development Establishes Competence for Future Gene Activation. Cell Rep 30:1319-1328 e6.

90. Tedesco M, Giannese F, Lazarevic D, Giansanti V, Rosano D, Monzani S, Catalano I, Grassi E, Zanella ER, Botrugno OA, Morelli L, Panina Bordignon P, Caravagna G, Bertotti A, Martino G, Aldrighetti L, Pasqualato S, Trusolino L, Cittaro D, Tonon G. 2022. Chromatin Velocity reveals epigenetic dynamics by single-cell profiling of heterochromatin and euchromatin. Nat Biotechnol 40:235-244.
91. Shah PP, Keough KC, Gjoni K, Santini GT, Abdill RJ, Wickramasinghe NM, Dundes CE, Karnay A, Chen A, Salomon REA, Walsh PJ, Nguyen SC, Whalen S, Joyce EF, Loh KM, Dubois N, Pollard KS, Jain R. 2023. An atlas of lamina-associated chromatin across twelve human cell types reveals an intermediate chromatin subtype. Genome Biol 24:16.
92. Madsen-Osterbye J, Abdelhalim M, Baudement MO, Collas P. 2022. Local euchromatin enrichment in lamina-associated domains anticipates their repositioning in the adipogenic lineage. Genome Biol 23:91.
93. Madrigal P, Deng S, Feng Y, Militi S, Goh KJ, Nibhani R, Grandy R, Osnato A, Ortmann D, Brown S, Pauklin S. 2023. Epigenetic and transcriptional regulations prime cell fate before division during human pluripotent stem cell differentiation. Nat Commun 14:405.
94. Schröder CM, Zissel L, Mersiowsky S-L, Tekman M, Probst S, Schüle KM, Preissl S, Schilling O, Timmers HTM, Arnold SJ. 2023. An EOMES induced epigenetic deflection initiates lineage commitment at mammalian gastrulation. bioRxiv doi:10.1101/2023.03.15.532746:2023.03.15.532746.
95. Hirabayashi S, Bhagat S, Matsuki Y, Takegami Y, Uehata T, Kanemaru A, Itoh M, Shirakawa K, Takaori-Kondo A, Takeuchi O, Carninci P, Katayama S, Hayashizaki Y, Kere J, Kawaji H, Murakawa Y. 2019. NET-CAGE characterizes the dynamics and topology of human transcribed cis-regulatory elements. Nat Genet 51:1369-1379.
96. Barshad G, Lewis JJ, Chivu AG, Abuhashem A, Krietenstein N, Rice EJ, Ma Y, Wang Z, Rando OJ, Hadjantonakis AK, Danko CG. 2023. RNA polymerase II dynamics shape enhancer-promoter interactions. Nat Genet 55:1370-1380.
97. Menet JS, Pescatore S, Rosbash M. 2014. CLOCK:BMAL1 is a pioneer-like transcription factor. Genes Dev 28:8-13.
98. Furlan-Magaril M, Ando-Kuri M, Arzate-Mejia RG, Morf J, Cairns J, Roman-Figueroa A, Tenorio-Hernandez L, Poot-Hernandez AC, Andrews S, Varnai C, Virk B, Wingett SW, Fraser P. 2021. The global and promoter-centric 3D genome organization temporally resolved during a circadian cycle. Genome Biol 22:162.
99. Misra N, Damara M, Ye T, Chambon P. 2023. The circadian demethylation of a unique intronic deoxymethylCpG-rich island boosts the transcription of its cognate circadian clock output gene. Proc Natl Acad Sci U S A 120:e2214062120.
100. Marco A, Meharena HS, Dileep V, Raju RM, Davila-Velderrain J, Zhang AL, Adaikkan C, Young JZ, Gao F, Kellis M, Tsai LH. 2020. Mapping the epigenomic and transcriptomic interplay during memory formation and recall in the hippocampal engram ensemble. Nat Neurosci doi:10.1038/s41593-020-00717-0.
101. Barajas-Mora EM, Lee L, Lu H, Valderrama JA, Bjanes E, Nizet V, Feeney AJ, Hu M, Murre C. 2023. Enhancer-instructed epigenetic landscape and chromatin compartmentalization dictate a primary antibody repertoire protective against specific bacterial pathogens. Nat Immunol doi:10.1038/s41590-022-01402-z.
102. Sima J, Chakraborty A, Dileep V, Michalski M, Klein KN, Holcomb NP, Turner JL, Paulsen MT, Rivera-Mulia JC, Trevilla-Garcia C, Bartlett DA, Zhao PA, Washburn BK, Nora EP, Kraft K, Mundlos S, Bruneau BG, Ljungman M, Fraser P, Ay F, Gilbert DM. 2019. Identifying cis Elements for Spatiotemporal Control of Mammalian DNA Replication. Cell 176:816-830 e18.
103. Stergachis AB, Debo BM, Haugen E, Churchman LS, Stamatoyannopoulos JA. 2020. Single-molecule regulatory architectures captured by chromatin fiber sequencing. Science 368:1449-1454.
104. Chuong EB, Rumi MA, Soares MJ, Baker JC. 2013. Endogenous retroviruses function as species-specific enhancer elements in the placenta. Nat Genet 45:325-9.
105. Xu R, Li S, Wu Q, Li C, Jiang M, Guo L, Chen M, Yang L, Dong X, Wang H, Wang C, Liu X, Ou X, Gao S. 2022. Stage-specific H3K9me3 occupancy ensures retrotransposon silencing in human pre-implantation embryos. Cell Stem Cell 29:1051-1066 e8.
106. Ito J, Seita Y, Kojima S, Parrish NF, Sasaki K, Sato K. 2022. A hominoid-specific endogenous retrovirus may have rewired the gene regulatory network shared between primordial germ cells and naive pluripotent cells. PLoS Genet 18:e1009846.
107. Roller M, Stamper E, Villar D, Izuogu O, Martin F, Redmond AM, Ramachanderan R, Harewood L, Odom DT, Flicek P. 2021. LINE retrotransposons characterize mammalian tissue-specific and evolutionarily dynamic regulatory regions. Genome Biol 22:62.
108. Balestrieri C, Alfarano G, Milan M, Tosi V, Prosperini E, Nicoli P, Palamidessi A, Scita G, Diaferia GR, Natoli G. 2018. Co-optation of Tandem DNA Repeats for the Maintenance of Mesenchymal Identity. Cell 173:1150-1164 e14.
109. Mangan RJ, Alsina FC, Mosti F, Sotelo-Fonseca JE, Snellings DA, Au EH, Carvalho J, Sathyan L, Johnson GD, Reddy TE, Silver DL, Lowe CB. 2022. Adaptive sequence divergence forged new neurodevelopmental enhancers in humans. Cell 185:4587-4603 e23.
110. Andrews G, Fan K, Pratt HE, Phalke N, Zoonomia Consortium section s, Karlsson EK, Lindblad-Toh K, Gazal S, Moore JE, Weng Z. 2023. Mammalian evolution of human cis-regulatory elements and transcription factor binding sites. Science 380:eabn7930.
111. Tatarakis A, Saini H, Moazed D. 2023. Requirements for establishment and epigenetic stability of mammalian heterochromatin. bioRxiv doi:10.1101/2023.02.27.530221:2023.02.27.530221.
112. Radman-Livaja M, Verzijlbergen KF, Weiner A, van Welsem T, Friedman N, Rando OJ, van Leeuwen F. 2011. Patterns and mechanisms of ancestral histone protein inheritance in budding yeast. PLoS Biol 9:e1001075.
113. Baker CL, Walker M, Arat S, Ananda G, Petkova P, Powers NR, Tian H, Spruce C, Ji B, Rausch D, Choi K, Petkov PM, Carter GW, Paigen K. 2019. Tissue-Specific Trans Regulation of the Mouse Epigenome. Genetics 211:831-845.
114. Hawe JS, Wilson R, Schmid KT, Zhou L, Lakshmanan LN, Lehne BC, Kuhnel B, Scott WR, Wielscher M, Yew YW, Baumbach C, Lee DP, Marouli E, Bernard M, Pfeiffer L, Matias-Garcia PR, Autio MI, Bourgeois S, Herder C, Karhunen V, Meitinger T, Prokisch H, Rathmann W, Roden M, Sebert S, Shin J, Strauch K, Zhang W, Tan WLW, Hauck SM, Merl-Pham J, Grallert H, Barbosa EGV, Mu TC, Illig T, Peters A, Paus T, Pausova Z, Deloukas P, Foo RSY, Jarvelin MR, Kooner JS, Loh M, Heinig M, Gieger C, Waldenberger M, Chambers JC. 2022. Genetic variation influencing DNA methylation provides insights into molecular mechanisms regulating genomic function. Nat Genet 54:18-29.
115. Bertozzi TM, Elmer JL, Macfarlan TS, Ferguson-Smith AC. 2020. KRAB zinc finger protein diversification drives mammalian interindividual methylation variability. Proc Natl Acad Sci U S A 117:31290-31300.

116. Sardina JL, Collombet S, Tian TV, Gomez A, Di Stefano B, Berenguer C, Brumbaugh J, Stadhouders R, Segura-Morales C, Gut M, Gut IG, Heath S, Aranda S, Di Croce L, Hochedlinger K, Thieffry D, Graf T. 2018. Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate. Cell Stem Cell 23:727-741 e9.
117. Stadhouders R, Vidal E, Serra F, Di Stefano B, Le Dily F, Quilez J, Gomez A, Collombet S, Berenguer C, Cuartero Y, Hecht J, Filion GJ, Beato M, Marti-Renom MA, Graf T. 2018. Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat Genet 50:238-249.
118. Castro FL, Brustolini OJB, Geddes VEV, Souza J, Alves-Leon SV, Aguiar RS, Vasconcelos ATR. 2022. Modulation of HERV Expression by Four Different Encephalitic Arboviruses during Infection of Human Primary Astrocytes. Viruses 14.
119. Leung A, Trac C, Kato H, Costello KR, Chen Z, Natarajan R, Schones DE. 2018. LTRs activated by Epstein-Barr virus-induced transformation of B cells alter the transcriptome. Genome Res 28:1791-1798.
120. Zhao B, Zou J, Wang H, Johannsen E, Peng CW, Quackenbush J, Mar JC, Morton CC, Freedman ML, Blacklow SC, Aster JC, Bernstein BE, Kieff E. 2011. Epstein-Barr virus exploits intrinsic B-lymphocyte transcription programs to achieve immortal cell growth. Proc Natl Acad Sci U S A 108:14902-7.
121. Okabe A, Huang KK, Matsusaka K, Fukuyo M, Xing M, Ong X, Hoshii T, Usui G, Seki M, Mano Y, Rahmutulla B, Kanda T, Suzuki T, Rha SY, Ushiku T, Fukayama M, Tan P, Kaneda A. 2020. Cross-species chromatin interactions drive transcriptional rewiring in Epstein-Barr virus-positive gastric adenocarcinoma. Nat Genet 52:919-930.
122. Wang C, Liu X, Liang J, Narita Y, Ding W, Li D, Zhang L, Wang H, Leong MML, Hou I, Gerdt C, Jiang C, Zhong Q, Tang Z, Forney C, Kottyan L, Weirauch MT, Gewurz BE, Zeng MS, Jiang S, Teng M, Zhao B. 2023. A DNA tumor virus globally reprograms host 3D genome architecture to achieve immortal growth. Nat Commun 14:1598.
123. Wang L, Laing J, Yan B, Zhou H, Ke L, Wang C, Narita Y, Zhang Z, Olson MR, Afzali B, Zhao B, Kazemian M. 2020. Epstein-Barr Virus Episome Physically Interacts with Active Regions of the Host Genome in Lymphoblastoid Cells. J Virol 94.
124. Stanfield BA, Luftig MA. 2017. Recent advances in understanding Epstein-Barr virus. F1000Res 6:386.
125. Kieffer-Kwon KR, Nimura K, Rao SSP, Xu J, Jung S, Pekowska A, Dose M, Stevens E, Mathe E, Dong P, Huang SC, Ricci MA, Baranello L, Zheng Y, Tomassoni Ardori F, Resch W, Stavreva D, Nelson S, McAndrew M, Casellas A, Finn E, Gregory C, St Hilaire BG, Johnson SM, Dubois W, Cosma MP, Batchelor E, Levens D, Phair RD, Misteli T, Tessarollo L, Hager G, Lakadamyali M, Liu Z, Floer M, Shroff H, Aiden EL, Casellas R. 2017. Myc Regulates Chromatin Decompaction and Nuclear Architecture during B Cell Activation. Mol Cell 67:566-578 e10.
126. Xu M, Lin J, Yang S, Yao J, Chen M, Feng J, Zhang L, Zhou L, Zhang J, Qin Q. 2023. Epstein-Barr virus-encoded miR-BART11-3p modulates the DUSP6-MAPK axis to promote gastric cancer cell proliferation and metastasis. J Virol doi:10.1128/jvi.00881-23:e0088123.
127. Wood CD, Veenstra H, Khasnis S, Gunnell A, Webb HM, Shannon-Lowe C, Andrews S, Osborne CS, West MJ. 2016. MYC activation and BCL2L11 silencing by a tumour virus through the large-scale reconfiguration of enhancer-promoter hubs. Elife 5.
128. Jachowicz JW, Bing X, Pontabry J, Boskovic A, Rando OJ, Torres-Padilla ME. 2017. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. Nat Genet 49:1502-1510.
129. Aloni-Grinstein R, Charni-Natan M, Solomon H, Rotter V. 2018. p53 and the Viral Connection: Back into the Future (double dagger). Cancers (Basel) 10.
130. Leonova KI, Brodsky L, Lipchick B, Pal M, Novototskaya L, Chenchik AA, Sen GC, Komarova EA, Gudkov AV. 2013. p53 cooperates with DNA methylation and a suicidal interferon response to maintain epigenetic silencing of repeats and noncoding RNAs. Proc Natl Acad Sci U S A 110:E89-98.
131. Wylie A, Jones AE, D'Brot A, Lu WJ, Kurtz P, Moran JV, Rakheja D, Chen KS, Hammer RE, Comerford SA, Amatruda JF, Abrams JM. 2016. p53 genes function to restrain mobile elements. Genes Dev 30:64-77.
132. Tiwari B, Jones AE, Caillet CJ, Das S, Royer SK, Abrams JM. 2020. p53 directly represses human LINE1 transposons. Genes Dev doi:10.1101/gad.343186.120.
133. Sun S, Hong J, You E, Tsanov KM, Chacon-Barahona J, Gioacchino AD, Hoyos D, Li H, Jiang H, Ly H, Marhon S, Murali R, Chanda P, Karacay A, Vabret N, De Carvalho DD, LaCava J, Lowe SW, Ting DT, Iacobuzio-Donahue CA, Solovyov A, Greenbaum BD. 2023. Cancer cells co-evolve with retrotransposons to mitigate viral mimicry. bioRxiv doi:10.1101/2023.05.19.541456.
134. Giam CZ, Semmes OJ. 2016. HTLV-1 Infection and Adult T-Cell Leukemia/Lymphoma-A Tale of Two Proteins: Tax and HBZ. Viruses 8.
135. Alarcon V, Hernandez S, Rubio L, Alvarez F, Flores Y, Varas-Godoy M, De Ferrari GV, Kann M, Villanueva RA, Loyola A. 2016. The enzymes LSD1 and Set1A cooperate with the viral protein HBx to establish an active hepatitis B viral chromatin state. Sci Rep 6:25901.
136. Dias JD, Sarica N, Cournac A, Koszul R, Neuveut C. 2022. Crosstalk between Hepatitis B Virus and the 3D Genome Structure. Viruses 14.
137. Liu W, Yao Q, Su X, Deng Y, Yang M, Peng B, Zhao F, Du C, Zhang X, Zhu J, Wang D, Li W, Li H. 2023. Molecular insights into Spindlin1-HBx interplay and its impact on HBV transcription from cccDNA minichromosome. Nat Commun 14:4663.
138. Heinz S, Texari L, Hayes MGB, Urbanowski M, Chang MW, Givarkes N, Rialdi A, White KM, Albrecht RA, Pache L, Marazzi I, Garcia-Sastre A, Shaw ML, Benner C. 2018. Transcription Elongation Can Affect Genome 3D Structure. Cell 174:1522-1536 e22.
139. Ryabchenko B, Sroller V, Hornikova L, Lovtsov A, Forstova J, Huerfano S. 2023. The interactions between PML nuclear bodies and small and medium size DNA viruses. Virol J 20:82.
140. Corpet A, Kleijwegt C, Roubille S, Juillard F, Jacquet K, Texier P, Lomonte P. 2020. PML nuclear bodies and chromatin dynamics: catch me if you can! Nucleic Acids Res 48:11890-11912.
141. Jan Fada B, Guha U, Zheng Y, Reward E, Kaadi E, Dourra A, Gu H. 2023. A Novel Recognition by the E3 Ubiquitin Ligase of HSV-1 ICP0 Enhances the Degradation of PML Isoform I to Prevent ND10 Reformation in Late Infection. Viruses 15.
142. Hashimoto K, Suzuki AM, Dos Santos A, Desterke C, Collino A, Ghisletti S, Braun E, Bonetti A, Fort A, Qin XY, Radaelli E, Kaczkowski B, Forrest AR, Kojima S, Samuel D, Natoli G, Buendia MA, Faivre J, Carninci P. 2015. CAGE profiling of ncRNAs in hepatocellular carcinoma reveals widespread activation of retroviral LTR promoters in virus-induced tumors. Genome Res 25:1812-24.

143. Kaposi-Novak P, Libbrecht L, Woo HG, Lee YH, Sears NC, Coulouarn C, Conner EA, Factor VM, Roskams T, Thorgeirsson SS. 2009. Central role of c-Myc during malignant conversion in human hepatocarcinogenesis. Cancer Res 69:2775-82.
144. D'Artista L, Moschopoulou AA, Barozzi I, Craig AJ, Seehawer M, Herrmann L, Minnich M, Kang TW, Rist E, Henning M, Klotz S, Heinzmann F, Harbig J, Sipos B, Longerich T, Eilers M, Dauch D, Zuber J, Wang XW, Zender L. 2023. MYC determines lineage commitment in KRAS-driven primary liver cancer development. J Hepatol 79:141-149.
145. Fourel G. 1994. Genetic and epigenetic alteration of gene expression in the course of hepatocarcinogenesis, p 297-343. *In* Tronche F, Yaniv M (ed), Liver gene expression. R.G. Landes Company, Austin.
146. Buendia MA, Neuveut C. 2015. Hepatocellular carcinoma. Cold Spring Harb Perspect Med 5:a021444.
147. Etiemble J, Degott C, Renard CA, Fourel G, Shamoon B, Vitvitski-Trepo L, Hsu TY, Tiollais P, Babinet C, Buendia MA. 1994. Liver-specific expression and high oncogenic efficiency of a c-myc transgene activated by woodchuck hepatitis virus insertion. Oncogene 9:727-37.
148. Renard CA, Fourel G, Bralet MP, Degott C, De La Coste A, Perret C, Tiollais P, Buendia MA. 2000. Hepatocellular carcinoma in WHV/N-myc2 transgenic mice: oncogenic mutations of beta-catenin and synergistic effect of p53 null alleles. Oncogene 19:2678-86.
149. Cotterman R, Jin VX, Krig SR, Lemen JM, Wey A, Farnham PJ, Knoepfler PS. 2008. N-Myc regulates a widespread euchromatic program in the human genome partially independent of its role as a classical transcription factor. Cancer Res 68:9654-62.
150. Baluapuri A, Wolf E, Eilers M. 2020. Target gene-independent functions of MYC oncoproteins. Nat Rev Mol Cell Biol 21:255-267.
151. Hong H, Takahashi K, Ichisaka T, Aoi T, Kanagawa O, Nakagawa M, Okita K, Yamanaka S. 2009. Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. Nature 460:1132-5.
152. Nicetto D, Zaret KS. 2019. Role of H3K9me3 heterochromatin in cell identity establishment and maintenance. Curr Opin Genet Dev 55:1-10.
153. Ito J, Sugimoto R, Nakaoka H, Yamada S, Kimura T, Hayano T, Inoue I. 2017. Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. PLoS Genet 13:e1006883.
154. Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, Noushmehr H, Lange CP, van Dijk CM, Tollenaar RA, Van Den Berg D, Laird PW. 2011. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. Nat Genet 44:40-6.
155. Feinberg AP. 2018. The Key Role of Epigenetics in Human Disease Prevention and Mitigation. N Engl J Med 378:1323-1334.
156. Narayan A, Ji W, Zhang XY, Marrogi A, Graff JR, Baylin SB, Ehrlich M. 1998. Hypomethylation of pericentromeric DNA in breast adenocarcinomas. Int J Cancer 77:833-8.
157. Feinberg AP, Tycko B. 2004. The history of cancer epigenetics. Nat Rev Cancer 4:143-53.
158. Pappalardo XG, Barra V. 2021. Losing DNA methylation at repetitive elements and breaking bad. Epigenetics Chromatin 14:25.
159. Eymery A, Horard B, El Atifi-Borel M, Fourel G, Berger F, Vitte AL, Van den Broeck A, Brambilla E, Fournier A, Callanan M, Gazzeri S, Khochbin S, Rousseaux S, Gilson E, Vourc'h C. 2009. A transcriptomic analysis of human centromeric and pericentric sequences in normal and tumor cells. Nucleic Acids Res 37:6340-54.
160. Solovyov A, Vabret N, Arora KS, Snyder A, Funt SA, Bajorin DF, Rosenberg JE, Bhardwaj N, Ting DT, Greenbaum BD. 2018. Global Cancer Transcriptome Quantifies Repeat Element Polarization between Immunotherapy Responsive and T Cell Suppressive Classes. Cell Rep 23:512-521.
161. Perelli L, Zhang L, Mangiameli S, Russell AJC, Giannese F, Peng F, Carbone F, Le C, Khan H, Citron F, Soeung M, Lam TNA, Lundgren S, Zhu C, Catania D, Feng N, Gurreri E, Sgambato A, Tortora G, Draetta GF, Tonon G, Futreal A, Giuliani V, Carugo A, Viale A, Heffernan TP, Wang L, Cittaro D, Chen F, Genovese G. 2023. Evolutionary fingerprints of EMT in pancreatic cancers. bioRxiv doi:10.1101/2023.09.18.558231:2023.09.18.558231.
162. Collings CK, Anderson JN. 2017. Links between DNA methylation and nucleosome occupancy in the human genome. Epigenetics Chromatin 10:18.
163. Kaaij LJT, Mohn F, van der Weide RH, de Wit E, Buhler M. 2019. The ChAHP Complex Counteracts Chromatin Looping at CTCF Sites that Emerged from SINE Expansions in Mouse. Cell 178:1437-1451 e14.
164. Ito J, Kimura I, Soper A, Coudray A, Koyanagi Y, Nakaoka H, Inoue I, Turelli P, Trono D, Sato K. 2020. Endogenous retroviruses drive KRAB zinc-finger protein family expression for tumor suppression. Sci Adv 6.
165. Wegert J, Fischer AK, Palhazi B, Treger TD, Hilgers C, Ziegler B, Jung H, Juttner E, Waha A, Fuchs J, Warmann SW, Fruhwald MC, Hubertus J, Pritchard-Jones K, Graf N, Behjati S, Furtwangler R, Gessler M, Vokuhl C. 2023. TRIM28 inactivation in epithelial nephroblastoma is frequent and often associated with predisposing TRIM28 germline variants. J Pathol doi:10.1002/path.6206.
166. Dyer MA, Qadeer ZA, Valle-Garcia D, Bernstein E. 2017. ATRX and DAXX: Mechanisms and Mutations. Cold Spring Harb Perspect Med 7.
167. Carraro M, Hendriks IA, Hammond CM, Solis-Mezarino V, Volker-Albert M, Elsborg JD, Weisser MB, Spanos C, Montoya G, Rappsilber J, Imhof A, Nielsen ML, Groth A. 2023. DAXX adds a de novo H3.3K9me3 deposition pathway to the histone chaperone network. Mol Cell 83:1075-1092 e9.
168. Ni K, Ren J, Xu X, He Y, Finney R, Braun SMG, Hathaway NA, Crabtree GR, Muegge K. 2020. LSH mediates gene repression through macroH2A deposition. Nat Commun 11:5647.
169. Ni K, Muegge K. 2021. LSH catalyzes ATP-driven exchange of histone variants macroH2A1 and macroH2A2. Nucleic Acids Res 49:8024-8036.
170. Lee SC, Adams DW, Ipsaro JJ, Cahn J, Lynn J, Kim HS, Berube B, Major V, Calarco JP, LeBlanc C, Bhattacharjee S, Ramu U, Grimanelli D, Jacob Y, Voigt P, Joshua-Tor L, Martienssen RA. 2023. Chromatin remodeling of histone H3 variants by DDM1 underlies epigenetic inheritance of DNA methylation. Cell 186:4100-4116 e15.
171. Moore PC, Henderson KW, Classon M. 2023. The epigenome and the many facets of cancer drug tolerance. Adv Cancer Res 158:1-39.

172. Kanholm T, Rentia U, Hadley M, Karlow JA, Cox OL, Diab N, Bendall ML, Dawson T, McDonald JI, Xie W, Crandall KA, Burns KH, Baylin SB, Easwaran H, Chiappinelli KB. 2023. Oncogenic Transformation Drives DNA Methylation Loss and Transcriptional Activation of Transposable Element Loci. Cancer Res doi:10.1158/0008-5472.CAN-22-3485.
173. Gopi LK, Kidder BL. 2021. Integrative pan cancer analysis reveals epigenomic variation in cancer type and cell specific chromatin domains. Nat Commun 12:1419.
174. Kent D, Marchetti L, Mikulasova A, Russell LJ, Rico D. 2023. Broad H3K4me3 domains: Maintaining cellular identity and their implication in super-enhancer hijacking. Bioessays 45:e2200239.
175. Feinberg AP, Levchenko A. 2023. Epigenetics as a mediator of plasticity in cancer. Science 379:eaaw3835.
176. Nicetto D, Donahue G, Jain T, Peng T, Sidoli S, Sheng L, Montavon T, Becker JS, Grindheim JM, Blahnik K, Garcia BA, Tan K, Bonasio R, Jenuwein T, Zaret KS. 2019. H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. Science 363:294-297.
177. Hathaway NA, Bell O, Hodges C, Miller EL, Neel DS, Crabtree GR. 2012. Dynamics and memory of heterochromatin in living cells. Cell 149:1447-60.
178. Bintu L, Yong J, Antebi YE, McCue K, Kazuki Y, Uno N, Oshimura M, Elowitz MB. 2016. Dynamics of epigenetic regulation at the single-cell level. Science 351:720-4.
179. Jenkinson G, Pujadas E, Goutsias J, Feinberg AP. 2017. Potential energy landscapes identify the information-theoretic nature of the epigenome. Nat Genet 49:719-729.
180. Househam J, Heide T, Cresswell GD, Spiteri I, Kimberley C, Zapata L, Lynn C, James C, Mossner M, Fernandez-Mateos J, Vinceti A, Baker AM, Gabbutt C, Berner A, Schmidt M, Chen B, Lakatos E, Gunasri V, Nichol D, Costa H, Mitchinson M, Ramazzotti D, Werner B, Iorio F, Jansen M, Caravagna G, Barnes CP, Shibata D, Bridgewater J, Rodriguez-Justo M, Magnani L, Sottoriva A, Graham TA. 2022. Phenotypic plasticity and genetic control in colorectal cancer evolution. Nature 611:744-753.
181. Karr JP, Ferrie JJ, Tjian R, Darzacq X. 2022. The transcription factor activity gradient (TAG) model: contemplating a contact-independent mechanism for enhancer-promoter communication. Genes Dev 36:7-16.
182. Cullen KE, Kladde MP, Seyfred MA. 1993. Interaction between transcription regulatory regions of prolactin chromatin. Science 261:203-6.
183. Zaugg JB, Sahlen P, Andersson R, Alberich-Jorda M, de Laat W, Deplancke B, Ferrer J, Mandrup S, Natoli G, Plewczynski D, Rada-Iglesias A, Spicuglia S. 2022. Current challenges in understanding the role of enhancers in disease. Nat Struct Mol Biol doi:10.1038/s41594-022-00896-3.
184. Fourel G, Couturier J, Wei Y, Apiou F, Tiollais P, Buendia MA. 1994. Evidence for long-range oncogene activation by hepadnavirus insertion. EMBO J 13:2526-34.
185. Walters MC, Magis W, Fiering S, Eidemiller J, Scalzo D, Groudine M, Martin DI. 1996. Transcriptional enhancers act in cis to suppress position-effect variegation. Genes Dev 10:185-95.
186. Fourel G, Boscheron C, Revardel E, Lebrun E, Hu YF, Simmen KC, Muller K, Li R, Mermod N, Gilson E. 2001. An activation-independent role of transcription factors in insulator function. EMBO Rep 2:124-32.
187. Judd J, Sanderson H, Feschotte C. 2021. Evolution of mouse circadian enhancers from transposable elements. Genome Biol 22:193.
188. Neumayr C, Haberle V, Serebreni L, Karner K, Hendy O, Boija A, Henninger JE, Li CH, Stejskal K, Lin G, Bergauer K, Pagani M, Rath M, Mechtler K, Arnold CD, Stark A. 2022. Differential cofactor dependencies define distinct types of human enhancers. Nature doi:10.1038/s41586-022-04779-x.
189. Zeller P, Yeung J, Vinas Gaza H, de Barbanson BA, Bhardwaj V, Florescu M, van der Linden R, van Oudenaarden A. 2023. Single-cell sortChIC identifies hierarchical chromatin dynamics during hematopoiesis. Nat Genet 55:333-345.
190. Nurk S, Koren S, Rhie A, Rautiainen M, Bzikadze AV, Mikheenko A, Vollger MR, Altemose N, Uralsky L, Gershman A, Aganezov S, Hoyt SJ, Diekhans M, Logsdon GA, Alonge M, Antonarakis SE, Borchers M, Bouffard GG, Brooks SY, Caldas GV, Chen NC, Cheng H, Chin CS, Chow W, de Lima LG, Dishuck PC, Durbin R, Dvorkina T, Fiddes IT, Formenti G, Fulton RS, Fungtammasan A, Garrison E, Grady PGS, Graves-Lindsay TA, Hall IM, Hansen NF, Hartley GA, Haukness M, Howe K, Hunkapiller MW, Jain C, Jain M, Jarvis ED, Kerpedjiev P, Kirsche M, Kolmogorov M, Korlach J, Kremitzki M, Li H, et al. 2022. The complete sequence of a human genome. Science 376:44-53.
191. Benabdallah NS, Williamson I, Illingworth RS, Kane L, Boyle S, Sengupta D, Grimes GR, Therizols P, Bickmore WA. 2019. Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. Mol Cell 76:473-484 e7.
192. Lin X, Liu Y, Liu S, Zhu X, Wu L, Zhu Y, Zhao D, Xu X, Chemparathy A, Wang H, Cao Y, Nakamura M, Noordermeer JN, La Russa M, Wong WH, Zhao K, Qi LS. 2022. Nested epistasis enhancer networks for robust genome regulation. Science 377:1077-1085.
193. Girdhar K, Hoffman GE, Bendl J, Rahman S, Dong P, Liao W, Hauberg ME, Sloofman L, Brown L, Devillers O, Kassim BS, Wiseman JR, Park R, Zharovsky E, Jacobov R, Flatow E, Kozlenkov A, Gilgenast T, Johnson JS, Couto L, Peters MA, Phillips-Cremins JE, Hahn CG, Gur RE, Tamminga CA, Lewis DA, Haroutunian V, Psych EC, Dracheva S, Lipska BK, Marenco S, Kundakovic M, Fullard JF, Jiang Y, Roussos P, Akbarian S. 2022. Chromatin domain alterations linked to 3D genome organization in a large cohort of schizophrenia and bipolar disorder brains. Nat Neurosci 25:474-483.
194. Factor DC, Barbeau AM, Allan KC, Hu LR, Madhavan M, Hoang AT, Hazel KEA, Hall PA, Nisraiyya S, Najm FJ, Miller TE, Nevin ZS, Karl RT, Lima BR, Song Y, Sibert AG, Dhillon GK, Volsko C, Bartels CF, Adams DJ, Dutta R, Gallagher MD, Phu W, Kozlenkov A, Dracheva S, Scacheri PC, Tesar PJ, Corradin O. 2020. Cell Type-Specific Intralocus Interactions Reveal Oligodendrocyte Mechanisms in MS. Cell 181:382-395 e21.
195. Chatterjee S, Kapoor A, Akiyama JA, Auer DR, Lee D, Gabriel S, Berrios C, Pennacchio LA, Chakravarti A. 2016. Enhancer Variants Synergistically Drive Dysfunction of a Gene Regulatory Network In Hirschsprung Disease. Cell 167:355-368 e10.
196. Morris JA, Caragine C, Daniloski Z, Domingo J, Barry T, Lu L, Davis K, Ziosi M, Glinos DA, Hao S, Mimitou EP, Smibert P, Roeder K, Katsevich E, Lappalainen T, Sanjana NE. 2023. Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens. Science doi:10.1126/science.adh7699:eadh7699.
197. Connally NJ, Nazeen S, Lee D, Shi H, Stamatoyannopoulos J, Chun S, Cotsapas C, Cassa CA, Sunyaev SR. 2022. The missing link between genetic association and regulatory function. Elife 11.

198. Bulut-Karslioglu A, Perrera V, Scaranaro M, de la Rosa-Velazquez IA, van de Nobelen S, Shukeir N, Popow J, Gerle B, Opravil S, Pagani M, Meidhof S, Brabletz T, Manke T, Lachner M, Jenuwein T. 2012. A transcription factor-based mechanism for mouse heterochromatin formation. Nat Struct Mol Biol 19:1023-30.
199. Velazquez Camacho O, Galan C, Swist-Rosowska K, Ching R, Gamalinda M, Karabiber F, De La Rosa-Velazquez I, Engist B, Koschorz B, Shukeir N, Onishi-Seebacher M, van de Nobelen S, Jenuwein T. 2017. Major satellite repeat RNA stabilize heterochromatin retention of Suv39h enzymes by RNA-nucleosome association and RNA:DNA hybrid formation. Elife 6.
200. Thakur J, Packiaraj J, Henikoff S. 2021. Sequence, Chromatin and Evolution of Satellite DNA. Int J Mol Sci 22.
201. Feng Y, Wang Y, Wang X, He X, Yang C, Naseri A, Pederson T, Zheng J, Zhang S, Xiao X, Xie W, Ma H. 2020. Simultaneous epigenetic perturbation and genome imaging reveal distinct roles of H3K9me3 in chromatin architecture and transcription. Genome Biol 21:296.
202. Noma K, Cam HP, Maraia RJ, Grewal SI. 2006. A role for TFIIIC transcription factor complex in genome organization. Cell 125:859-72.
203. Hiraga S, Botsios S, Donze D, Donaldson AD. 2012. TFIIIC localizes budding yeast ETC sites to the nuclear periphery. Mol Biol Cell 23:2741-54.
204. Yokoshi M, Segawa K, Fukaya T. 2020. Visualizing the Role of Boundary Elements in Enhancer-Promoter Communication. Mol Cell 78:224-235 e5.
205. Stutzman AV, Liang AS, Beilinson V, Ikegami K. 2020. Transcription-independent TFIIIC-bound sites cluster near heterochromatin boundaries within lamina-associated domains in C. elegans. Epigenetics Chromatin 13:1.
206. Kundu TK, Wang Z, Roeder RG. 1999. Human TFIIIC relieves chromatin-mediated repression of RNA polymerase III transcription and contains an intrinsic histone acetyltransferase activity. Mol Cell Biol 19:1605-15.
207. Siddaway R, Milos S, Coyaud E, Yun HY, Morcos SM, Pajovic S, Campos EI, Raught B, Hawkins C. 2022. The in vivo Interaction Landscape of Histones H3.1 and H3.3. Mol Cell Proteomics 21:100411.
208. Oughtred R, Rust J, Chang C, Breitkreutz BJ, Stark C, Willems A, Boucher L, Leung G, Kolas N, Zhang F, Dolma S, Coulombe-Huntington J, Chatr-Aryamontri A, Dolinski K, Tyers M. 2021. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein Sci 30:187-200.
209. Olsen JV, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P, Mann M. 2006. Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127:635-48.
210. Reverendo M, Arguello RJ, Polte C, Valecka J, Camosseto V, Auphan-Anezin N, Ignatova Z, Gatti E, Pierre P. 2019. Polymerase III transcription is necessary for T cell priming by dendritic cells. Proc Natl Acad Sci U S A 116:22721-22729.
211. Linker SB, Randolph-Moore L, Kottilil K, Qiu F, Jaeger BN, Barron J, Gage FH. 2020. Identification of bona fide B2 SINE retrotransposon transcription through single-nucleus RNA-seq of the mouse hippocampus. Genome Res 30:1643-1654.
212. Gibson BA, Doolittle LK, Schneider MWG, Jensen LE, Gamarra N, Henry L, Gerlich DW, Redding S, Rosen MK. 2019. Organization of Chromatin by Intrinsic and Regulated Phase Separation. Cell 179:470-484 e21.
213. Oki M, Valenzuela L, Chiba T, Ito T, Kamakaka RT. 2004. Barrier proteins remodel and modify chromatin to restrict silenced domains. Mol Cell Biol 24:1956-67.
214. Solovei I, Thanisch K, Feodorova Y. 2016. How to rule the nucleus: divide et impera. Curr Opin Cell Biol 40:47-59.
215. Lu JY, Shao W, Chang L, Yin Y, Li T, Zhang H, Hong Y, Percharde M, Guo L, Wu Z, Liu L, Liu W, Yan P, Ramalho-Santos M, Sun Y, Shen X. 2020. Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. Cell Rep 30:3296-3311 e5.
216. Travers AA, Vaillant C, Arneodo A, Muskhelishvili G. 2012. DNA structure, nucleosome placement and chromatin remodelling: a perspective. Biochem Soc Trans 40:335-40.
217. Abdulhay NJ, McNally CP, Hsieh LJ, Kasinathan S, Keith A, Estes LS, Karimzadeh M, Underwood JG, Goodarzi H, Narlikar GJ, Ramani V. 2020. Massively multiplex single-molecule oligonucleosome footprinting. Elife 9.
218. Kilic S, Felekyan S, Doroshenko O, Boichenko I, Dimura M, Vardanyan H, Bryan LC, Arya G, Seidel CAM, Fierz B. 2018. Single-molecule FRET reveals multiscale chromatin dynamics modulated by HP1alpha. Nat Commun 9:235.
219. Takahata S, Chida S, Ohnuma A, Ando M, Asanuma T, Murakami Y. 2021. Two secured FACT recruitment mechanisms are essential for heterochromatin maintenance. Cell Rep 36:109540.
220. Xue Y, Pradhan SK, Sun F, Chronis C, Tran N, Su T, Van C, Vashisht A, Wohlschlegel J, Peterson CL, Timmers HTM, Kurdistani SK, Carey MF. 2017. Mot1, Ino80C, and NC2 Function Coordinately to Regulate Pervasive Transcription in Yeast and Mammals. Mol Cell 67:594-607 e4.
221. Saksouk N, Barth TK, Ziegler-Birling C, Olova N, Nowak A, Rey E, Mateos-Langerak J, Urbach S, Reik W, Torres-Padilla ME, Imhof A, Dejardin J, Simboeck E. 2014. Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. Mol Cell 56:580-94.
222. Ranjan A, Nguyen VQ, Liu S, Wisniewski J, Kim JM, Tang X, Mizuguchi G, Elalaoui E, Nickels TJ, Jou V, English BP, Zheng Q, Luk E, Lavis LD, Lionnet T, Wu C. 2020. Live-cell single particle imaging reveals the role of RNA polymerase II in histone H2A.Z eviction. Elife 9.
223. Zhang C, Tian Y, Song S, Zhang L, Dang Y, He Q. 2022. H3K56 deacetylation and H2A.Z deposition are required for aberrant heterochromatin spreading. Nucleic Acids Res 50:3852-3866.
224. Imre L, Nánási P, Bosire R, Csóti Á, Enyedi KN, Mező G, Kusakabe M, Ausio J, Harata M, Szabó G. 2021. Fundamental Role Of The H2A.Z C-Terminal Tail In The Formation Of Constitutive Heterochromatin. bioRxiv doi:10.1101/2021.02.22.432230:2021.02.22.432230.
225. Tartour K, Andriani F, Folco EG, Letkova D, Schneider R, Saidi I, Sato T, Welz PS, Benitah SA, Allier C, Padmanabhan K. 2022. Mammalian PERIOD2 regulates H2A.Z incorporation in chromatin to orchestrate circadian negative feedback. Nat Struct Mol Biol 29:549-562.
226. Hardy S, Robert F. 2010. Random deposition of histone variants: A cellular mistake or a novel regulatory mechanism? Epigenetics 5:368-72.
227. Chen Z, Gabizon R, Brown AI, Lee A, Song A, Diaz-Celis C, Kaplan CD, Koslover EF, Yao T, Bustamante C. 2019. High-resolution and high-accuracy topographic and transcriptional maps of the nucleosome barrier. Elife 8.

228. Fourel G, Miyake T, Defossez PA, Li R, Gilson E. 2002. General regulatory factors (GRFs) as genome partitioners. J Biol Chem 277:41736-43.
229. Dion MF, Kaplan T, Kim M, Buratowski S, Friedman N, Rando OJ. 2007. Dynamics of replication-independent histone turnover in budding yeast. Science 315:1405-8.
230. Dodd IB, Sneppen K. 2011. Barriers and silencers: a theoretical toolkit for control and containment of nucleosome-based epigenetic states. J Mol Biol 414:624-37.
231. McClintock B. 1984. The significance of responses of the genome to challenge. Science 226:792-801.
232. Kaplow IM, Lawler AJ, Schaffer DE, Srinivasan C, Sestili HH, Wirthlin ME, Phan BN, Prasad K, Brown AR, Zhang X, Foley K, Genereux DP, Zoonomia C, Karlsson EK, Lindblad-Toh K, Meyer WK, Pfenning AR. 2023. Relating enhancer genetic variation across mammals to complex phenotypes using machine learning. Science 380:eabm7993.
233. Griffin GK, Wu J, Iracheta-Vellve A, Patti JC, Hsu J, Davis T, Dele-Oni D, Du PP, Halawi AG, Ishizuka JJ, Kim SY, Klaeger S, Knudsen NH, Miller BC, Nguyen TH, Olander KE, Papanastasiou M, Rachimi S, Robitschek EJ, Schneider EM, Yeary MD, Zimmer MD, Jaffe JD, Carr SA, Doench JG, Haining WN, Yates KB, Manguso RT, Bernstein BE. 2021. Epigenetic silencing by SETDB1 suppresses tumour intrinsic immunogenicity. Nature 595:309-314.
234. Soares ML, Edwards CA, Dearden FL, Ferron SR, Curran S, Corish JA, Rancourt RC, Allen SE, Charalambous M, Ferguson-Smith MA, Rens W, Adams DJ, Ferguson-Smith AC. 2018. Targeted deletion of a 170-kb cluster of LINE-1 repeats and implications for regional control. Genome Res 28:345-56.
235. Kvikstad EM, Makova KD. 2010. The (r)evolution of SINE versus LINE distributions in primate genomes: sex chromosomes are important. Genome Res 20:600-13.
236. Ewing AD, Smits N, Sanchez-Luque FJ, Faivre J, Brennan PM, Richardson SR, Cheetham SW, Faulkner GJ. 2020. Nanopore Sequencing Enables Comprehensive Transposable Element Epigenomic Profiling. Mol Cell 80:915-928 e5.
237. Zhu Y, van Essen D, Saccani S. 2012. Cell-type-specific control of enhancer activity by H3K9 trimethylation. Mol Cell 46:408-23.
238. Yelagandula R, Stecher K, Novatchkova M, Michetti L, Michlits G, Wang J, Hofbauer P, Vainorius G, Pribitzer C, Isbel L, Mendjan S, Schubeler D, Elling U, Brennecke J, Bell O. 2023. ZFP462 safeguards neural lineage specification by targeting G9A/GLP-mediated heterochromatin to silence enhancers. Nat Cell Biol doi:10.1038/s41556-022-01051-2.
239. Hoyt SJ, Storer JM, Hartley GA, Grady PGS, Gershman A, de Lima LG, Limouse C, Halabian R, Wojenski L, Rodriguez M, Altemose N, Rhie A, Core LJ, Gerton JL, Makalowski W, Olson D, Rosen J, Smit AFA, Straight AF, Vollger MR, Wheeler TJ, Schatz MC, Eichler EE, Phillippy AM, Timp W, Miga KH, O'Neill RJ. 2022. From telomere to telomere: The transcriptional and epigenetic state of human repeat elements. Science 376:eabk3112.
240. Mehdipour P, Marhon SA, Ettayebi I, Chakravarthy A, Hosseini A, Wang Y, de Castro FA, Loo Yau H, Ishak C, Abelson S, O'Brien CA, De Carvalho DD. 2020. Epigenetic therapy induces transcription of inverted SINEs and ADAR1 dependency. Nature 588:169-173.
241. Deniz O, Frost JM, Branco MR. 2019. Regulation of transposable elements by DNA modifications. Nat Rev Genet 20:417-431.
242. Nishibuchi G, Dejardin J. 2017. The molecular basis of the organization of repetitive DNA-containing constitutive heterochromatin in mammals. Chromosome Res 25:77-87.
243. McCarthy RL, Kaeding KE, Keller SH, Zhong Y, Xu L, Hsieh A, Hou Y, Donahue G, Becker JS, Alberto O, Lim B, Zaret KS. 2021. Diverse heterochromatin-associated proteins repress distinct classes of genes and repetitive elements. Nat Cell Biol 23:905-914.
244. Wang L, Gao Y, Zheng X, Liu C, Dong S, Li R, Zhang G, Wei Y, Qu H, Li Y, Allis CD, Li G, Li H, Li P. 2019. Histone Modifications Regulate Chromatin Compartmentalization by Contributing to a Phase Separation Mechanism. Mol Cell 76:646-659 e6.
245. Liu N, Lee CH, Swigut T, Grow E, Gu B, Bassik MC, Wysocka J. 2018. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. Nature 553:228-232.
246. Tan K, Kim ME, Song HW, Skarbrevik D, Babajanian E, Bedrosian TA, Gage FH, Wilkinson MF. 2021. The Rhox gene cluster suppresses germline LINE1 transposition. Proc Natl Acad Sci U S A 118.
247. Shaban HA, Barth R, Recoules L, Bystricky K. 2020. Hi-D: nanoscale mapping of nuclear dynamics in single living cells. Genome Biol 21:95.
248. Rausch C, Hastert FD, Cardoso MC. 2019. DNA Modification Readers and Writers and Their Interplay. J Mol Biol doi:10.1016/j.jmb.2019.12.018.
249. Li Y, Chen X, Lu C. 2021. The interplay between DNA and histone methylation: molecular mechanisms and disease implications. EMBO Rep 22:e51803.
250. Buckberry S, Liu X, Poppe D, Tan JP, Sun G, Chen J, Nguyen TV, de Mendoza A, Pflueger J, Frazer T, Vargas-Landin DB, Paynter JM, Smits N, Liu N, Ouyang JF, Rossello FJ, Chy HS, Rackham OJL, Laslett AL, Breen J, Faulkner GJ, Nefzger CM, Polo JM, Lister R. 2023. Transient naive reprogramming corrects hiPS cells functionally and epigenetically. Nature doi:10.1038/s41586-023-06424-7.
251. Clark SJ, Argelaguet R, Lohoff T, Krueger F, Drage D, Gottgens B, Marioni JC, Nichols J, Reik W. 2022. Single-cell multi-omics profiling links dynamic DNA methylation to cell fate decisions during mouse early organogenesis. Genome Biol 23:202.
252. Horard B, Eymery A, Fourel G, Vassetzky N, Puechberty J, Roizes G, Lebrigand K, Barbry P, Laugraud A, Gautier C, Simon EB, Devaux F, Magdinier F, Vourc'h C, Gilson E. 2009. Global analysis of DNA methylation and transcription of human repetitive sequences. Epigenetics 4:339-50.
253. Francastel C, Magdinier F. 2019. DNA methylation in satellite repeats disorders. Essays Biochem 63:757-771.
254. Argelaguet R, Clark SJ, Mohammed H, Stapel LC, Krueger C, Kapourani CA, Imaz-Rosshandler I, Lohoff T, Xiang Y, Hanna CW, Smallwood S, Ibarra-Soria X, Buettner F, Sanguinetti G, Xie W, Krueger F, Gottgens B, Rugg-Gunn PJ, Kelsey G, Dean W, Nichols J, Stegle O, Marioni JC, Reik W. 2019. Multi-omics profiling of mouse gastrulation at single-cell resolution. Nature 576:487-491.
255. Cuellar TL, Herzner AM, Zhang X, Goyal Y, Watanabe C, Friedman BA, Janakiraman V, Durinck S, Stinson J, Arnott D, Cheung TK, Chaudhuri S, Modrusan Z, Doerr JM, Classon M, Haley B. 2017. Silencing of retrotransposons by SETDB1 inhibits the interferon response in acute myeloid leukemia. J Cell Biol 216:3535-3549.

256. Guler GD, Tindell CA, Pitti R, Wilson C, Nichols K, KaiWai Cheung T, Kim HJ, Wongchenko M, Yan Y, Haley B, Cuellar T, Webster J, Alag N, Hegde G, Jackson E, Nance TL, Giresi PG, Chen KB, Liu J, Jhunjhunwala S, Settleman J, Stephan JP, Arnott D, Classon M. 2017. Repression of Stress-Induced LINE-1 Expression Protects Cancer Cell Subpopulations from Lethal Drug Exposure. Cancer Cell 32:221-237 e13.
257. Baratchian M, Tiwari R, Khalighi S, Chakravarthy A, Yuan W, Berk M, Li J, Guerinot A, de Bono J, Makarov V, Chan TA, Silverman RH, Stark GR, Varadan V, De Carvalho DD, Chakraborty AA, Sharifi N. 2022. H3K9 methylation drives resistance to androgen receptor-antagonist therapy in prostate cancer. Proc Natl Acad Sci U S A 119:e2114324119.
258. Canat A, Veillet A, Batrin R, Dubourg C, Illingworth R, Fabre E, Therizols P. 2023. DAXX safeguards heterochromatin formation in embryonic stem cells. bioRxiv doi:10.1101/2021.04.28.441827:2021.04.28.441827.
259. Lindholm HT, Chen R, De Carvalho DD. 2023. Endogenous retroelements as alarms for disruptions to cellular homeostasis. Trends Cancer 9:55-68.
260. Liang L, Cao C, Ji L, Cai Z, Wang D, Ye R, Chen J, Yu X, Zhou J, Bai Z, Wang R, Yang X, Zhu P, Xue Y. 2023. Complementary Alu sequences mediate enhancer-promoter selectivity. Nature 619:868-875.
261. Virstedt J, Berge T, Henderson RM, Waring MJ, Travers AA. 2004. The influence of DNA stiffness upon nucleosome formation. J Struct Biol 148:66-85.
262. Vinayachandran V, Shivaswamy S, Shukla A, Kendyala N, Bhargava P. 2023. Terminator-dependent facilitated recycling of RNA polymerase III couples transcriptional activation and chromatin remodeling in vivo. bioRxiv doi:10.1101/2023.02.28.530554:2023.02.28.530554.
263. Tramantano M, Sun L, Au C, Labuz D, Liu Z, Chou M, Shen C, Luk E. 2016. Constitutive turnover of histone H2A.Z at yeast promoters requires the preinitiation complex. Elife 5.
264. Barbier J, Vaillant C, Volff JN, Brunet FG, Audit B. 2021. Coupling between Sequence-Mediated Nucleosome Organization and Genome Evolution. Genes (Basel) 12.
265. Drillon G, Audit B, Argoul F, Arneodo A. 2016. Evidence of selection for an accessible nucleosomal array in human. BMC Genomics 17:526.
266. Brunet FG, Audit B, Drillon G, Argoul F, Volff JN, Arneodo A. 2018. Evidence for DNA Sequence Encoding of an Accessible Nucleosomal Array across Vertebrates. Biophys J 114:2308-2316.
267. Collings CK, Waddell PJ, Anderson JN. 2013. Effects of DNA methylation on nucleosome stability. Nucleic Acids Res 41:2918-31.
268. Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, Shen SY, Han H, Liang G, Jones PA, Pugh TJ, O'Brien C, De Carvalho DD. 2015. DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. Cell 162:961-73.
269. Zapatka M, Borozan I, Brewer DS, Iskar M, Grundhoff A, Alawi M, Desai N, Sultmann H, Moch H, Pathogens P, Cooper CS, Eils R, Ferretti V, Lichter P, Consortium P. 2020. The landscape of viral associations in human cancers. Nat Genet 52:320-330.
270. Colombo AR, Triche T, Jr., Ramsingh G. 2018. Transposable Element Expression in Acute Myeloid Leukemia Transcriptome and Prognosis. Sci Rep 8:16449.
271. Porter RL, Sun S, Flores MN, Berzolla E, You E, Phillips IE, Kc N, Desai N, Tai EC, Szabolcs A, Lang ER, Pankaj A, Raabe MJ, Thapar V, Xu KH, Nieman LT, Rabe DC, Kolin DL, Stover EH, Pepin D, Stott SL, Deshpande V, Liu JF, Solovyov A, Matulonis UA, Greenbaum BD, Ting DT. 2022. Satellite repeat RNA expression in epithelial ovarian cancer associates with a tumor-immunosuppressive phenotype. J Clin Invest 132.
272. Deniz O, Ahmed M, Todd CD, Rio-Machin A, Dawson MA, Branco MR. 2020. Endogenous retroviruses are a source of enhancers with oncogenic potential in acute myeloid leukaemia. Nat Commun 11:3506.
273. Colombo AR, Zubair A, Thiagarajan D, Nuzhdin S, Triche TJ, Ramsingh G. 2017. Suppression of Transposable Elements in Leukemic Stem Cells. Sci Rep 7:7029.
274. Torres-Padilla ME. 2020. On transposons and totipotency. Philos Trans R Soc Lond B Biol Sci 375:20190339.
275. Lamers MM, van den Hoogen BG, Haagmans BL. 2019. ADAR1: "Editor-in-Chief" of Cytoplasmic Innate Immunity. Front Immunol 10:1763.
276. Guo YL. 2019. The underdeveloped innate immunity in embryonic stem cells: The molecular basis and biological perspectives from early embryogenesis. Am J Reprod Immunol 81:e13089.
277. Poirier EZ, Buck MD, Chakravarty P, Carvalho J, Frederico B, Cardoso A, Healy L, Ulferts R, Beale R, Reis e Sousa C. 2021. An isoform of Dicer protects mammalian stem cells against multiple RNA viruses. Science 373:231-236.
278. Chung H, Calis JJA, Wu X, Sun T, Yu Y, Sarbanes SL, Dao Thi VL, Shilvock AR, Hoffmann HH, Rosenberg BR, Rice CM. 2018. Human ADAR1 Prevents Endogenous RNA from Triggering Translational Shutdown. Cell 172:811-824 e14.
279. Bonnin M, Fares N, Testoni B, Estornes Y, Weber K, Vanbervliet B, Lefrancois L, Garcia A, Kfoury A, Pez F, Coste I, Saintigny P, Viari A, Lang K, Guey B, Hervieu V, Bancel B, Bartoch B, Durantel D, Renno T, Merle P, Lebecque S. 2019. Toll-like receptor 3 downregulation is an escape mechanism from apoptosis during hepatocarcinogenesis. J Hepatol 71:763-772.
280. Chen R, Ishak CA, De Carvalho DD. 2021. Endogenous Retroelements and the Viral Mimicry Response in Cancer Therapy and Cellular Homeostasis. Cancer Discov 11:2707-2725.
281. Loo Yau H, Ettayebi I, De Carvalho DD. 2019. The Cancer Epigenome: Exploiting Its Vulnerabilities for Immunotherapy. Trends Cell Biol 29:31-43.
282. Sheng W, LaFleur MW, Nguyen TH, Chen S, Chakravarthy A, Conway JR, Li Y, Chen H, Yang H, Hsu PH, Van Allen EM, Freeman GJ, De Carvalho DD, He HH, Sharpe AH, Shi Y. 2018. LSD1 Ablation Stimulates Anti-tumor Immunity and Enables Checkpoint Blockade. Cell 174:549-563 e19.
283. Lindholm HT, Chen R, De Carvalho DD. 2022. Endogenous retroelements as alarms for disruptions to cellular homeostasis. Trends Cancer doi:10.1016/j.trecan.2022.09.001.
284. Huang D, Petrykowska HM, Miller BF, Elnitski L, Ovcharenko I. 2019. Identification of human silencers by correlating cross-tissue epigenetic profiles and gene expression. Genome Res 29:657-667.
285. Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM, Sherlock G. 2004. GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. Bioinformatics 20:3710-5.

286. Cheetham SW, Kindlova M, Ewing AD. 2022. Methylartist: tools for visualizing modified bases from nanopore sequence data. Bioinformatics 38:3109-3112.
287. Lee BT, Barber GP, Benet-Pages A, Casper J, Clawson H, Diekhans M, Fischer C, Gonzalez JN, Hinrichs AS, Lee CM, Muthuraman P, Nassar LR, Nguy B, Pereira T, Perez G, Raney BJ, Rosenbloom KR, Schmelter D, Speir ML, Wick BD, Zweig AS, Haussler D, Kuhn RM, Haeussler M, Kent WJ. 2022. The UCSC Genome Browser database: 2022 update. Nucleic Acids Res 50:D1115-D1122.
288. Castro-Diaz N, Ecco G, Coluccio A, Kapopoulou A, Yazdanpanah B, Friedli M, Duc J, Jang SM, Turelli P, Trono D. 2014. Evolutionally dynamic L1 regulation in embryonic stem cells. Genes Dev 28:1397-409.
289. Kreibich E, Kleinendorst R, Barzaghi G, Kaspar S, Krebs AR. 2023. Single-molecule footprinting identifies context-dependent regulation of enhancers by DNA methylation. Mol Cell 83:787-802 e9.
290. Loyfer N, Magenheim J, Peretz A, Cann G, Bredno J, Klochendler A, Fox-Fisher I, Shabi-Porat S, Hecht M, Pelet T, Moss J, Drawshy Z, Amini H, Moradi P, Nagaraju S, Bauman D, Shveiky D, Porat S, Dior U, Rivkin G, Or O, Hirshoren N, Carmon E, Pikarsky A, Khalaileh A, Zamir G, Grinbaum R, Abu Gazala M, Mizrahi I, Shussman N, Korach A, Wald O, Izhar U, Erez E, Yutkin V, Samet Y, Rotnemer Golinkin D, Spalding KL, Druid H, Arner P, Shapiro AMJ, Grompe M, Aravanis A, Venn O, Jamshidi A, Shemer R, Dor Y, Glaser B, Kaplan T. 2023. A DNA methylation atlas of normal human cell types. Nature 613:355-364.
291. Abdulla AZ, Vaillant C, Jost D. 2022. Painters in chromatin: a unified quantitative framework to systematically characterize epigenome regulation and memory. Nucleic Acids Res 50:9083-9104.
292. Azad GK, Ito K, Sailaja BS, Biran A, Nissim-Rafinia M, Yamada Y, Brown DT, Takizawa T, Meshorer E. 2018. PARP1-dependent eviction of the linker histone H1 mediates immediate early gene expression during neuronal activation. J Cell Biol 217:473-481.
293. Gillespie MA, Palii CG, Sanchez-Taltavull D, Shannon P, Longabaugh WJR, Downes DJ, Sivaraman K, Espinoza HM, Hughes JR, Price ND, Perkins TJ, Ranish JA, Brand M. 2020. Absolute Quantification of Transcription Factors Reveals Principles of Gene Regulation in Erythropoiesis. Mol Cell 78:960-974 e11.