RESEARCH   **Open Access**

CrossMark

Analysis of 3D genomic interactions identifies candidate host genes that transposable elements potentially regulate

Ramya Raviram[1,2,3†], Pedro P. Rocha[1,4†], Vincent M. Luo[1,2], Emily Swanzey[5], Emily R. Miraldi[2,6,7,8],
Edward B. Chuong[9], Cédric Feschotte[10], Richard Bonneau[2,6,7] and Jane A. Skok[1*] ⓘD

Abstract

Background: The organization of chromatin in the nucleus plays an essential role in gene regulation. About half of the mammalian genome comprises transposable elements. Given their repetitive nature, reads associated with these elements are generally discarded or randomly distributed among elements of the same type in genome-wide analyses. Thus, it is challenging to identify the activities and properties of individual transposons. As a result, we only have a partial understanding of how transposons contribute to chromatin folding and how they impact gene regulation.

Results: Using PCR and Capture-based chromosome conformation capture (3C) approaches, collectively called 4Tran, we take advantage of the repetitive nature of transposons to capture interactions from multiple copies of endogenous retrovirus (ERVs) in the human and mouse genomes. With 4Tran-PCR, reads are selectively mapped to unique regions in the genome. This enables the identification of transposable element interaction profiles for individual ERV families and integration events specific to particular genomes. With this approach, we demonstrate that transposons engage in long-range intra-chromosomal interactions guided by the separation of chromosomes into A and B compartments as well as topologically associated domains (TADs). In contrast to 4Tran-PCR, Capture-4Tran can uniquely identify both ends of an interaction that involve retroviral repeat sequences, providing a powerful tool for uncovering the individual transposable element insertions that interact with and potentially regulate target genes.

Conclusions: 4Tran provides new insight into the manner in which transposons contribute to chromosome architecture and identifies target genes that transposable elements can potentially control.

Keywords: Transposons, Chromosome conformation capture, Endogenous retroviruses, Nuclear organization, Solo LTRs, Enhancers, Loops, Gene regulation

Background

The structural organization of the genome is regulated at different levels to establish a functional framework that facilitates cellular processes such as gene expression and programmed somatic recombination. Fluorescent in situ hybridization (FISH) studies revealed that chromosomes occupy discrete territories with very little intermingling between them. With the development of chromosome conformation capture (3C) techniques that rely on cross-linking of chromatin in close spatial proximity, additional

levels of organization have been described. There are many different 3C-based variants, with Hi-C being the most comprehensive, in that it can potentially identify all pair-wise chromatin interactions in a given population of cells [1–5]. The first Hi-C study revealed that each chromosome is divided into active (A) or inactive (B) compartments that range in size from ~ 5 to 10 MB in mammalian cells [6]. Furthermore, these analyses demonstrate that regions on the chromosome belonging to the same compartment preferentially interact with each other. With improved Hi-C sequencing depth, the presence of topologically associated domains (TADs) were defined [7]. The latter consist of highly self-interacting regions separated from each other by insulated boundaries. Unlike

Raviram *et al. Genome Biology*   (2018) 19:216

Page 2 of 19

A/B compartments, which are cell type specific, TADs are for the most part invariant across cell types and orthologous genomic regions of different species. It is thought that the main function of TADs is to restrict the influence of enhancers to genes found in the same domain. Indeed, approximately 90% of promoter-enhancer interactions occur between elements in the same TAD [8–12]. Further support for TAD restricted regulation comes from several studies in which disruption of TAD borders has been shown to lead to aberrant gene expression through exposure to previously insulated enhancers [13–16].

A large portion of a typical mammalian genome is comprised of transposable elements (TEs) however they are typically ignored in high-throughput sequencing-based studies due to their repetitive nature [17, 18]. As a consequence of this, few studies have analyzed how TEs influence, and are influenced by nuclear organization [19, 20]. These mobile elements, which have been propagated in the genomes of all eukaryotic species, can be classified as either DNA transposons or retrotransposons, depending on their mode of transposition. DNA transposons propagate via a cut and paste mechanism while retrotransposons use a copy and paste mode of action that relies on an RNA intermediate [21]. TEs have largely been considered as inert elements that remain silenced, except during a short temporal window during germ cell development. However, this view is changing with increasing evidence demonstrating that not all elements are permanently repressed in the genome. For example, in mammals, transcription of endogenous retroviruses and LINE-1 elements has been shown to occur during early embryogenesis [22–25] and transposition of L1 elements has been observed in human neurons [26].

Recent studies have suggested that TEs might act as enhancers capable of influencing expression of endogenous genes [25, 27–32]. This is not surprising, as TEs contain *cis*-regulatory elements such as those found at the 5′ UTR of long interspersed nuclear elements (LINEs) and long terminal repeats (LTRs) of endogenous retroviruses (ERVs) [33–36]. These *cis*-regulatory elements have evolved to control TE transcription but they can also influence expression of adjacent "host" genes [37, 38]. In fact, this idea was first put forward by Barbara McClintock who discovered transposons and described them as "controlling elements" because of their ability to influence the expression of maize genes during development [39, 40]. Moreover, genome-wide profiling of transcription factor (TF) binding in human and mouse cells has revealed that TEs contribute a significant proportion of sites [41–44]. In addition, a recent study [45] observed enriched binding of STAT1 in primate-specific ERVs called MER41 after interferon gamma (IFNG) treatment in HeLa and other human cells. STAT1-bound MER41 elements were enriched near innate immunity genes and in particular those that respond to interferon treatment. CRISPR deletion of selected MER41 elements led to downregulation of the interferon-driven transcriptional response of genes in close proximity. This indicates that MER41 retroviral insertions can act as enhancers, which control interferon-inducible expression of neighboring genes. Thus, not all TEs are silenced and inert, and, under particular conditions, they can act as regulatory elements, capable of influencing the expression of surrounding genes. However, only a subset of TEs is predicted to have such effects [46]. Therefore, it is essential to develop techniques that determine the identity of potential target genes that come into contact with these elements in order to ascertain which integration events have potential regulatory function. Follow-up targeting experiments can then be used to assess the extent to which they contribute to host gene regulation.

3C-based techniques have become instrumental in the identification of enhancers and the genes they regulate [47–50]. Here, we used two variants of 3C, tailored to identify different aspects of murine and human transposon-mediated interactions. We first describe 4Tran-PCR, which was used to obtain an interaction profile for different ERV families in murine cells by mapping all interactions from a specific ERV family to unique sequences within the mouse genome. This approach effectively identified strain-specific, polymorphic insertion sites. In addition, 4Tran-PCR revealed that the interaction profile of mouse ERVs follows a similar pattern to that of host genes, such that if an element is located in an active (A) or inactive (B) compartment, it will preferentially contact other loci within the same compartment, regardless of whether the element itself is decorated with active or inactive marks. Furthermore, 4Tran-PCR demonstrates that TE-mediated interactions are locally restricted to the same TAD.

Capture-4Tran uses biotinylated DNA oligonucleotides (oligos) to capture TEs in conjunction with surrounding uniquely mappable sequences, thereby enabling the identification of the interaction profile of a specific transposon copy, unlike the PCR variant of this technique. This provides a powerful tool for uncovering the individual insertion sites that interact with, and potentially regulate target genes. Here we designed an oligonucleotide that captures interactions involving human MER41 elements. We found that a significant fraction of MER41 interactions occur with promoters, STAT1 binding sites and histone modifications indicative of active transcription. We also observed that the loops are preformed and can be detected in the absence of STAT1 binding, prior to IFNγ induction, indicating that the contacts are not dependent on the presence of this transcription factor. Thus, 4Tran provides new insight

into the role of TEs in shaping genome organization and regulating cellular processes in mammalian cells.

Results

Chromosomal interactions involving TEs can be analyzed using 4Tran-PCR and Capture-4Tran

To investigate different aspects of murine and human retrotransposon-mediated interactions, we used two variants of 4Tran. The first, based on circular chromosome conformation capture (4C-Seq), was called 4Tran-PCR (Additional file 1: Figure S1). 4C-Seq captures the frequency with which a bait (or viewpoint) physically contacts other locations across the genome within a population of cells. Regions surrounding the bait interact at high frequency due to the polymer nature of chromosomes and the 3D organization of the genome into TADs. As such, 4C-Seq baits are characterized by a single region of strong signal that decays with increasing distance from the bait. Our strategy to probe TE interaction frequency across the genome, consists of using 4C-Seq with primers that hybridize to the repetitive region of transposons. Interactions are mapped to unique genomic sequences, while reads that map to multiple genomic locations, including TEs with low-sequence divergent copies are discarded as we cannot identify which insertion site these emanate from. With this method, several bait-like profiles with a strong accumulation of reads surrounding the integration site of a particular TE type are detected. This approach is best suited for analysis of TEs with relatively low copy number where it can be assumed that interactions are captured from the nearest insertion site.

As an alternative approach, we used a 4Tran variant named Capture-4Tran (Additional file 1: Figure S1). As the name implies, this is a capture based approach that uses two rounds of DNA capture with 120 bp biotinylated DNA probes to pull down interactions with specific regions of interest [51]. As in Capture-C, we performed sonication instead of the second digestion step used in 4C-Seq. This substantially improves our ability to distinguish between PCR duplicates and unique interactions [51]. Additionally, as in Capture Hi-C, we used Hi-C libraries [47] instead of the 3C libraries used in Capture-C. The primary difference between the two approaches is the use of biotin/streptavidin beads in the Hi-C library, which enrich for true ligation junctions formed due to 3D proximity. This increases the number of informative reads obtained at the end of the experiment, which is essential when enriching for interactions from thousands of capture regions. By combining paired-end sequencing with biotinylated oligonucleotides that hybridize to the 5′ or 3′ end of a TE integration site, isolated fragments contain the end of a transposon and the region neighboring the integration site as well as

any interacting fragments (Additional file 1: Figure S1). This approach increases the chances that uniquely mappable regions are captured with interactions involving TEs and therefore, each interaction can be assigned to a specific TE locus.

4Tran-PCR detects annotated ERV integration sites

The workflow for 4Tran-PCR shown in Additional file 1: Figure S2a depicts the sequential process from primer design (following 4C template generation) to identification of interactions that we used in 4Tran-PCR. Briefly, we design baits matching the consensus sequence of a TE family (Fig. 1a), as described in Repbase [52]. We predict which regions of the genome should be amplified with these primers and then intersect these locations with all annotated TEs to ensure that only the TE of interest is captured. Primer pairs that pass all of the above criteria can then be used to detect bait profiles based on enrichment of 4C-Seq signal.

To implement 4Tran-PCR in mouse cells, we focused on TEs classified as endogenous retrovirus (ERVs), which are known to be particularly diverse, recent, and active in the murine genome [33]. We selected the following murine ERVs: IAPEz, ETnERV, RLTR6, and MuLV-int/RLTR4 as these are among the youngest murine TE elements and they have retained the ability to transpose [53]. RLTR6 is the family name designated by Repbase/dfam [52] for the LTR of prototypical VL30 elements, a well-characterized group of active LTR retroelements in the mouse [54]. Distinct insertions belonging to the same family are expected to exhibit high sequence similarity, which facilitates repetitive bait design. To analyze their interaction profiles, we designed 4C-like baits for each repeat type, based on the consensus sequence for a set of primary and secondary restriction enzymes. Primer design rules are largely similar to those used in traditional 4C-Seq [55], except that primers are not excluded if they are located in repetitive genomic regions (details in the "Methods" section).

Primer pairs were selected based on the following criteria: location within the TE, number of integration copies that can be potentially captured and specificity to the transposon of interest. In addition, we selected baits located close to the 5′ and 3′ ends of TEs where *cis*-regulatory elements recognized by transcription factors tend to occur [56]. It is important to note that due to truncations, the majority of TE insertions are not representative of the full-length element (Fig. 1b and Additional file 1: Figure S2b). For ERVs, these include a high frequency of solitary LTRs, which result from non-allelic recombination events between the LTRs of full-length proviral integrants [33]. Since LTRs are so numerous, we avoided placing baits on them in order to keep the copy number to a manageable number. In addition, they are smaller in size making it
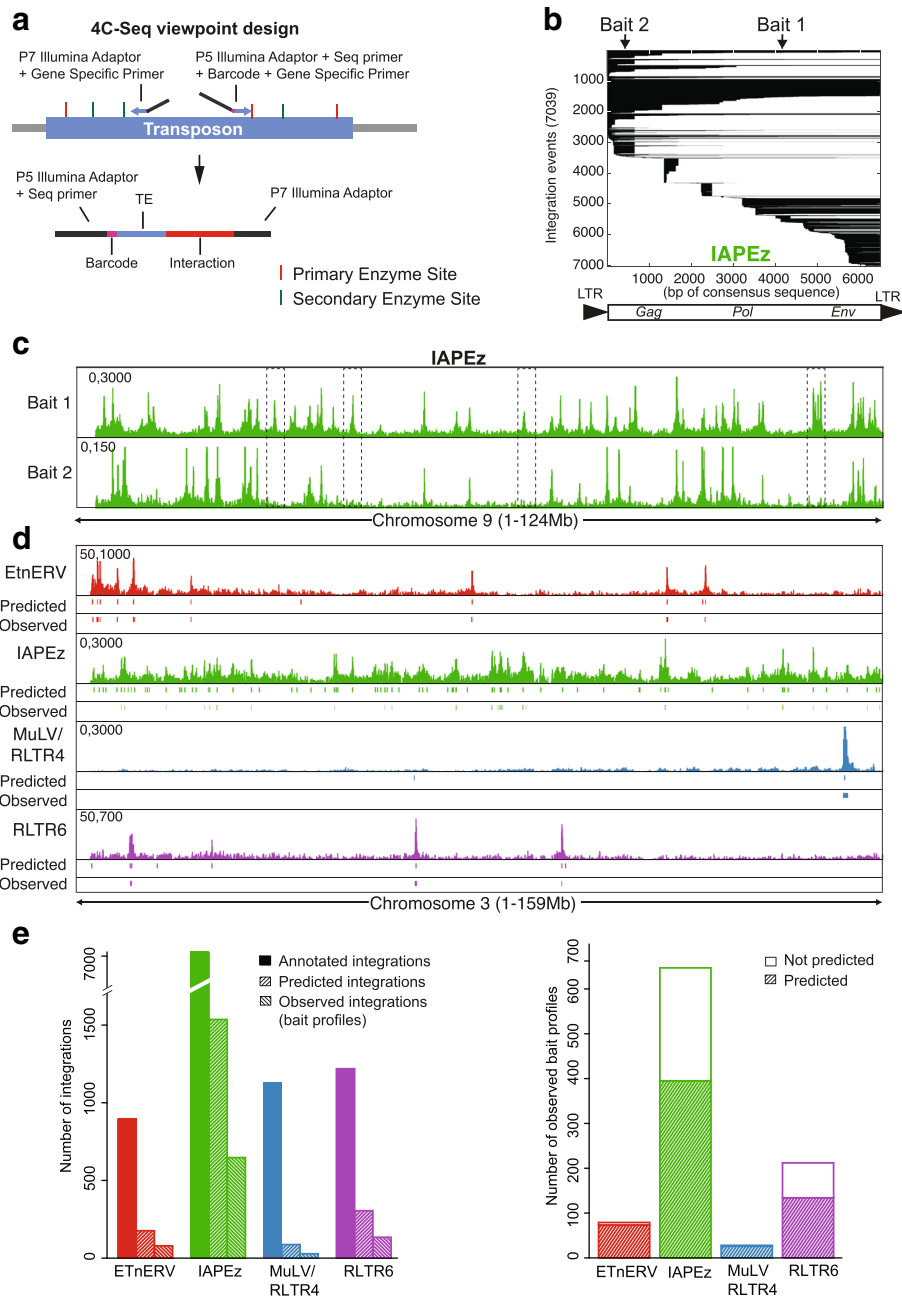
Fig. 1 Chromosomal interactions involving TEs can be analyzed using 4Tran-PCR. **a** Design of 4C-Seq primers to capture interactions from multiple transposon integration sites. The scheme shows how baits are designed within the transposon as well as the potential location of the two primers and the restriction enzyme sites necessary for amplification of a 4Tran-PCR library. The bottom of the scheme displays an interaction identified by Illumina sequencing containing the barcode, the sequence corresponding to the TE bait and the interaction fragment captured with the bait. **b** Schematic representation of all IAPEz integration events. Each line represents a different integration and the black lines show which part of the consensus sequence (shown under the plot) is retained by each integration. Integration events are sorted by 5′ position on the consensus sequence and by size of integration. Arrows represent the two locations tested for IAPEz baits. **c** 4Tran data in ES cells for the two IAPEz baits. Boxes represent three integration events detected by bait 1 that are not captured by bait 2. **d** 4Tran data for four different mouse ERVs on chromosome 3. Regions predicted to form bait-like profiles based on the presence of primers sequences (Predicted) and our algorithm based on bait like profiles (Observed) are shown under each 4Tran signal plot. **e** The left plot shows the number of annotated integrations in the mm10 genome for each ERV (annotated), the number of predicted integrations and the number of observed bait-like profiles. The plot on the right shows the number of observed and predicted bait-like profiles

Raviram *et al. Genome Biology*   (2018) 19:216

Page 5 of 19

difficult to design a pair of 4C-Seq primers. Finally, to ensure that no other TE families would be amplified by our bait sequences, we adapted the UCSC in silico tool to predict bait location and cross-referenced this to the annotation of all murine TEs. We applied these criteria to the baits designed for the four TEs, IAPEZ, ETnERV, RLTR6, and MuLV/RLTR4. Figure 1b shows the location of each bait relative to the consensus sequence and integration events of all IAPEz elements. The primer location for other ERVs can be found in Additional file 1: Figure S2b.

To test the primer pairs selected for each ERV, we performed 4C-Seq in mouse embryonic stem cells using DpnII and Csp6I as the primary and secondary restriction enzymes. As predicted, 4Tran-PCR primers located within sequences of endogenous retroviruses generated several bait-like profiles, instead of a single bait profile as in conventional 4C-Seq (Fig. 1c and Additional file 1: Figure S2d). It should be noted that the different primers for the same transposon did not capture all the same integration events, but rather captured only integration events that retained the primer sequence matching the consensus sequence (Fig. 1c). Since only unique reads are mapped, we detect interactions involving the regions surrounding the TE, while those involving only the repetitive regions of the transposon are discarded. Additional file 1: Figure S2c shows a zoomed in view of an annotated ERV bait showing the 4Tran-PCR signal derived from unique sequence reads of surrounding regions and an absence of TE reads in the center.

To determine if 4Tran-PCR can identify annotated ERV integration events, we selected preferred primer pairs based on their proximity to LTRs and their ability to generate efficient 4C signal amplification. As the mm10 reference genome from the Genome Reference Consortium is based on the C57/Bl6 strain, we performed 4Tran-PCR using template prepared from murine ex vivo derived splenic resting B cells from mice of this strain. B cells were chosen for their easy availability in our lab and because high-resolution Hi-C datasets that we could use as validations were restricted to B cells at the time of these experiments. Our first goal was to test whether the observed 4Tran bait profiles align with known ERV integration sites (Fig. 1d). To identify these regions in an unbiased manner, we binned the genome into 200 kb windows and compared the 4Tran signal in each window to a background distribution. Windows that contained enriched signal were called "observed baits." Additionally, we defined "predicted baits" as ERV integration events that retain the region of the consensus sequence matching the bait primers (see the "Methods" section for details).

The ERV family with the highest number of predicted integration events is IAPEz, followed by RLTR6, ETnERV, and MuLV. The number of observed baits for each of these ERVs also followed the same trend (Fig. 1e left). For
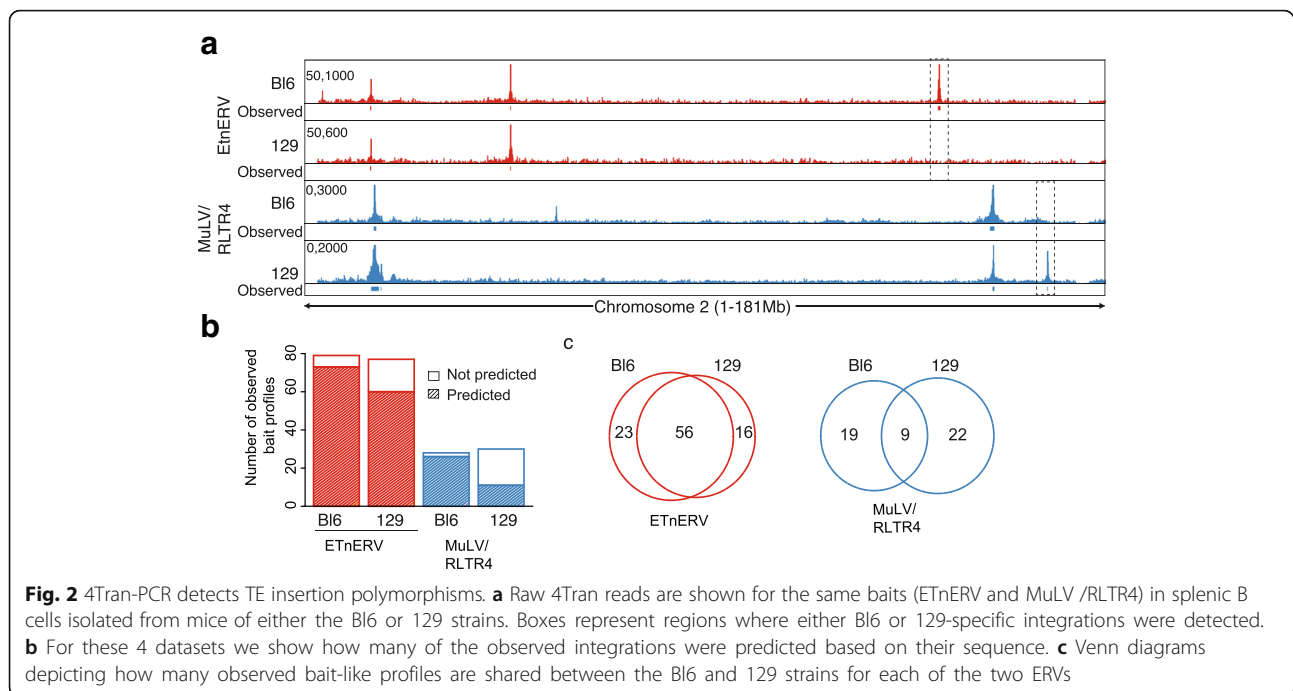
all four ERVs analyzed, we detected fewer observed baits than predicted. This could either be due to emergence of mutations that are not annotated in the reference genome, or to polymorphisms in our C57/Bl6 colony. Another possibility is that our method may not distinguish some of the signals as true 4Tran bait profiles due to low signal. This is particularly relevant for samples with a high frequency of integration sites where the higher density of bait-like profiles could obscure the distinction between bait signal and regions of high interactions. Another potential source for the disparity between the number of "observed" and "predicted" baits is that multiple integrations in close proximity cannot be detected as single copies.

Finally, we asked whether the locations we identified as observed baits had been predicted. (Fig. 1e right). The ERV elements with fewer observed baits, ETnERV and MuLV/RLTR4 displayed an almost perfect correspondence between observed and predicted baits. In contrast, the RLTR6 and IAPEz ERVs, had a higher number of observed baits, many of which were not predicted. As mentioned above, this discrepancy could be due to the high density of bait-like profiles, which blur the distinction between bait signal and regions of high interactions. Thus, the use of 4Tran is best suited to ERVs with a lower number of integration sites. Alternatively, primers that detect fewer integration sites of a specific TE family could be preferentially used instead of primers that detect all hybridization sites.

4Tran-PCR detects TE insertion polymorphisms

As 4Tran is able to detect the location of ERV integration events through enrichment of their local chromatin interactions, we next asked whether differences in integration sites of mobile ERVs could be detected between mouse strains as has been previously reported [57, 58]. Since the four ERV elements we analyzed are among the youngest in the murine lineage, we tested this hypothesis by performing 4Tran in ex vivo derived resting splenic B cells from the C57/Bl6 and 129S6 mouse strains. For this, we used the same primer sets described above to analyze integration of ETnERV and MuLV/RLTR4 and only retained the observed baits that overlapped between replicates for downstream analysis (Additional file 1: Figure S3a). We found many observed baits that were shared between mouse strains as well as C57/Bl6 and 129-specific integrations (Fig. 2a). As expected, the C57/Bl6 strain on which the mm10 reference genome is built performed better in identifying observed versus predicted baits compared to the 129S6 strain (Fig. 2b).

The Mouse Genomes Project (sanger.ac.uk/science/data/mouse-genomes-project) database, which generated whole-genome sequencing of 18 different mouse lab strains [59] was used to confirm the presence of structural variations in non-overlapping regions. For example,

Fig. 2 4Tran-PCR detects TE insertion polymorphisms. **a** Raw 4Tran reads are shown for the same baits (ETnERV and MuLV /RLTR4) in splenic B cells isolated from mice of either the Bl6 or 129 strains. Boxes represent regions where either Bl6 or 129-specific integrations were detected. **b** For these 4 datasets we show how many of the observed integrations were predicted based on their sequence. **c** Venn diagrams depicting how many observed bait-like profiles are shared between the Bl6 and 129 strains for each of the two ERVs

in the top panel of Fig. 2a, an ETnERV integration is detected only in the C57/Bl6 strain. When this bait location was compared between mouse strains in the database, we found a deletion that coincides precisely with the C57/Bl6 annotated ETnERV integration in the 129 mouse strain. Conversely, in the bottom panel of Fig. 2a, we show an MuLV/RLTR4 bait detected in 129S6 but not in C57/Bl6 cells, consistent with the annotation of a polymorphic insertion at this location in the database. These results indicate that differences in observed baits across strains correspond to strain-specific ERV insertions.

When differences between the strains were quantified genome-wide, ETnERV showed a higher level of conservation between 129 and C57/Bl6 than MuLV/RLTR4. As shown in Fig. 2c, up to 70% of the MuLV integrations were strain-specific, in contrast to the ETnERV family where 70% of integration sites are shared between the two strains. We verified sequence variations of all bait regions specific to a particular strain in the Mouse Genomes Project database and identified 28/38 differences in MulV/RLTR4 and 26/36 differences in ETnERV associated with a variation (Additional file 2: Table S3). These frequencies are in line with previous estimates for each of these ERV families of the numbers of elements that have polymorphisms across different mouse sub-strains [60].

Finally, we tested 4Tran-PCR in different sub-strains of 129 mice using the ETnERV bait, and found that polymorphic insertions could be detected between sub-strains. In the left panel of Additional file 1: Figure S3b, an integration event is captured in 2 out of 3 129 sub-strains while in the right panel, another integration

is captured in only 1 out of 3. We cross-referenced the regions of differential integrations to the Mouse Genomes Project database and found that the annotations matched our findings. Surprisingly, we also detected differences in integration events between littermates using the MuLV/RLTR4 bait and an integration on chromosome 2 was identified in one mouse that was not detected in its littermate (Additional file 1: Figure S3c). This transposition event is not annotated in the reference C57/Bl6 genome and likely arose in the germ cells of one of the parents. In sum, these examples show that 4Tran-PCR is capable of detecting differences in TE integration sites that are present in a population of cells. While other techniques solely focused on detecting new TE integrations have been developed [61, 62], a more comparative analysis is needed to assess the performance of this aspect of 4Tran-PCR.

ERV interactions are constrained by compartments and local domain structure

Having determined that 4Tran-PCR identifies chromosomal interactions of ERV elements we looked at how transposon contacts are influenced by the different levels of nuclear organization and asked whether ERV-mediated long-range interactions are influenced by chromosome compartmentalization. Compartment formation (A and B) seems to be mostly guided by the propensity of large chromatin fragments with similar histone and DNA modification patterns to share the same physical space. Formation of compartment A, for example, results from the spatial clustering of genomic regions enriched for histone

post-translational modifications associated with transcriptional activity, such as H3K4me1/3 or H3K27ac [5, 8]. Therefore, we asked whether full-length ERV copies known to be decorated with the heterochromatin mark H3K9me3 [56, 63] but located in A compartment regions would frequently contact regions in the B compartment that share H3K9me3 enrichment, or rather if they would evade compartmentalization because of the influence of their local A compartment neighborhood.

To address this question, we used Hi-C and ChIP-seq data from a splenic murine lymphoma IgM+ B cell line, CH12 that is used as a model for B cells in the same developmental stage as those we use here [64]. We performed a principal component analysis on 200 Kb-binned Hi-C data and classified each bin as compartments A and B depending on a positive or negative principal component score, respectively [6] (Fig. 3a). We then focused on

data generated using the RLTR4 bait where our primers detected a single integration site on chromosome 3, which is located in a region that belongs to compartment A (Fig. 3a). It should be noted that only single integration events on individual chromosomes can be used for this analysis because if multiple integration events were present it would not be possible to determine which site the interactions were emanating from. Full-length insertions of this ERV are silenced in B cells and decorated with the heterochromatin mark H3K9me3 (Additional file 1: Figure S4). Visual inspection of the interactions of this RLTR4 copy with the rest of the genome revealed that the regions of higher interaction belonged to the A compartment. The 4C-ker pipeline [65] was used to quantify this and determine which regions on chromosome 3 interact at high frequency with the MuLV integration site. As shown in Fig. 3b, interacting regions have positive PC
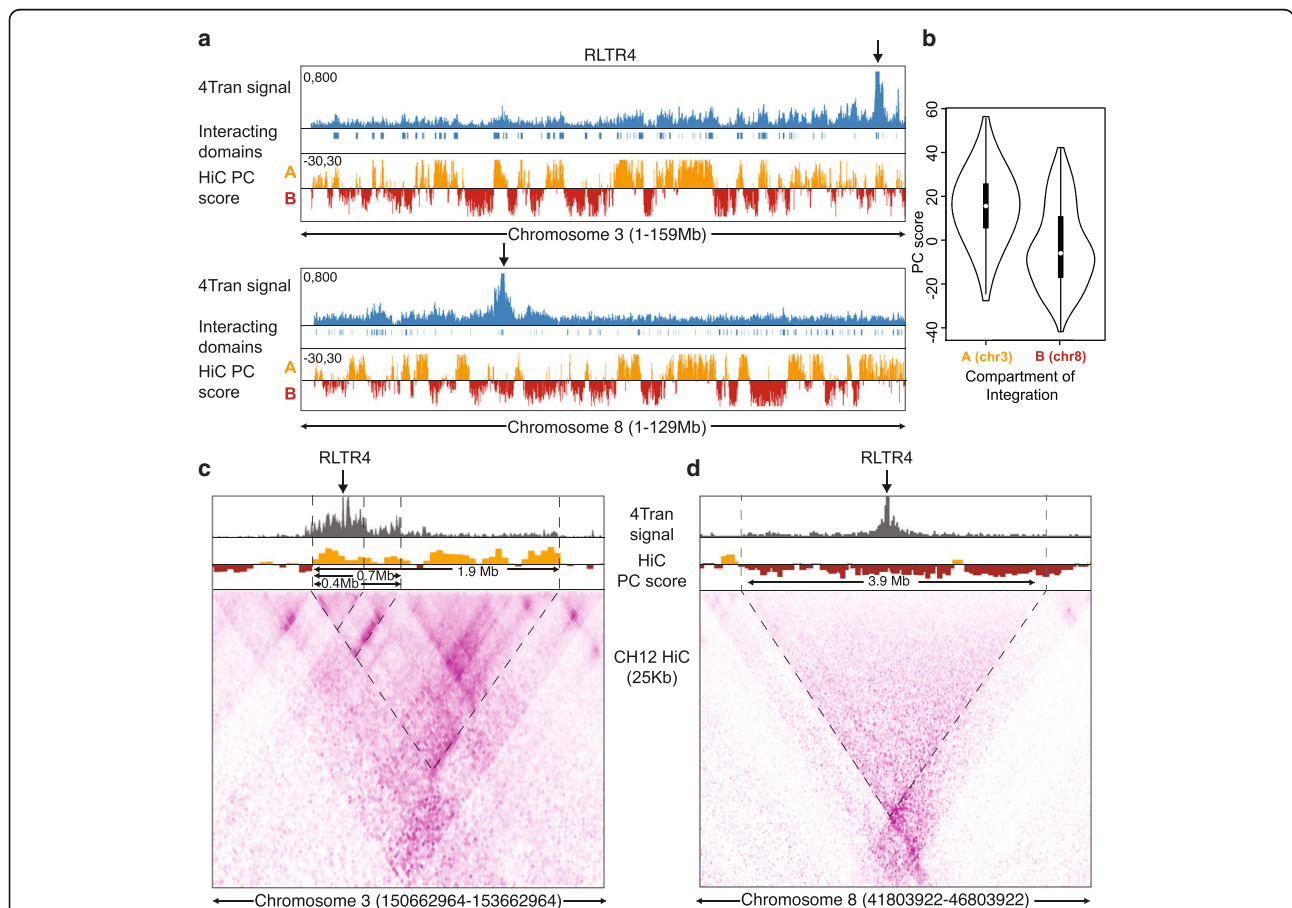

Fig. 3 ERV interactions are constrained by the different levels of nuclear organization **a** Whole chromosome view of 4Tran-PCR signal for RLTR4 integrations on chromosomes 3 and 8. The single integrations shown for these chromosomes are highlighted with an arrow over the plot and the regions identified as significantly interacting with these sites are shown under 4Tran-PCR signal as boxes. Hi-C data is represented by the PC-score calculated for each 50 kb bin. A positive PC score is characteristic of A regions, while a negative score is associate with B regions. **b** Violin plots representing the PC score for all regions identified as interacting *in cis* with the RLTR4 integration in chromosomes 3 and 8. An integration in compartment A leads to contacts with other compartment A regions, while the reverse is true for an integration in compartment B on chromosome 8. **c**, **d** High-resolution 4Tran-PCR data is shown together with Hi-C from Ch12 B cells. Hi-C is shown using both principal component score and 25 Kb-bins. Dashed lines highlight regions of high 4Tran-PCR signal and borders of domains, as described in the "Results" section

score values, indicating that this ERV integration preferentially interacts with other compartment A regions on chromosome 3. This data demonstrates that, just like non-repetitive loci, TE long-range interactions are guided by their compartment status and determined by the overall epigenetic status of their compartment rather than their specific chromatin marks. As a validation, interactions of an RLTR4 copy integrated on chromosome 8 in a B compartment region were examined. Again, these reveal that most interactions occurred with regions in the same compartment (Fig. 3a, b). Given the limitation of having only one integration per chromosome, there are not many examples we can analyze. However, three additional examples of single RLTR4 integration events on chromosomes 14, 15, and 19 follow the same trend whereby interactions are restricted to regions within the same A compartment that they are integrated in, rather than to regions with the same epigenetic state (Additional file 1: Figure S5).

We next examined whether the local interactions of transposable elements are confined within smaller architectural structures such as TADs [7, 66]. Here, we took advantage of the high-resolution Hi-C CH12 data to directly compare with 4Tran-PCR signal. Hi-C data revealed that the RLTR4 integration on chromosome 3 element is integrated in a small domain of approximately 0.4 Mb which is insulated from its immediate upstream region and is a substructure of two wider (0.7 and 1.9 Mb) nested domains that expand downstream of the MuLV integration site (Fig. 3c). 4Tran-PCR signal portrays the same architecture, with the strongest signal located within the 0.4 Mb domain. Interactions are dramatically reduced in the regions immediately upstream, indicative of strong insulation of the 0.4 Mb domain from its flanking regions. Furthermore, even though the regions upstream of the RLTR4 integration are closer on the linear chromosome, 4Tran signal is much stronger downstream of the RLTR4 element and constrained by the two subdomains of the larger 1.9 Mb domain. In contrast, the RLTR4 integration on chromosome 8 is much less structured, likely because it is located in a large 3.9 Mb compartment B domain (Fig. 3d). Similarly, 4Tran signal from this RLTR4 integration decays linearly and symmetrically with distance from the ERV site and is constrained by the borders of the domain detected by Hi-C. Taken together, these results indicate that both local domain structure and organization of the genome into compartments influences the manner in which TEs interact with other loci. It is important to note that low sequencing depth 4Tran signal (2–10 million reads) achieves a similar resolution to billions of Hi-C reads. Thus, 4Tran-PCR can be used to easily assess both local and long-range interactions from transposable elements.

Identifying interactions from individual TEs using 4Tran-PCR works optimally if insertion sites are well separated on the linear chromosome. However, 4Tran-PCR can confound the characterization of interactions from TE integrations in close proximity and it is thus best suited for analysis of TEs with relatively low copy number integrations. Under these circumstances, it can be assumed that interactions are captured from the nearest site, as by definition regions in the same domain interact at higher frequency than with the rest of the genome. A further drawback of 4Tran-PCR is that it is less efficient at amplifying interactions from families with highly divergent sequences, as it relies on the presence of two PCR primers and two restriction enzyme sites. To address these issues, we developed a second variant of 4Tran, Capture-4Tran.

Capture-4Tran identifies pairwise interactions from TEs

Capture-4Tran overcomes the problems related to repetitive sequence mappability by combining paired-end sequencing with biotinylated oligonucleotides that hybridize to the 5′ or 3′ end of a TE integration site. It can therefore be used to uniquely identify interactions from individual TE insertions. To test the feasibility of this approach, we first performed Capture-4Tran for the IAPEz family, which is one of the youngest and more homogeneous ERVs in the mouse genome. We performed a pilot experiment at low-sequencing depth using a probe based on the consensus sequence of the LTR that is most commonly associated with the IAPEz element, IAPLTR1a. As shown in Additional file 1: Figure S6, this strategy allowed unique mapping of interactions from IAPLTR1a elements, consisting of either a solo LTR or full-length insertions. Up to 30.5% of all interactions that we uniquely aligned in our Capture-4Tran experiment have at least one side mapped to an IAPLTR element demonstrating the ability of our probe to enrich for the desired fragments.

We next tested if Capture-4Tran is able to identify interactions from slightly older TEs that are therefore less homogeneous in sequence, but may have greater potential for being coopted for adjacent gene regulation [36, 42, 67]. For this we focused on the MER41 LTR elements in the human genome. MER41 represents the LTR of an endogenized gamma-retrovirus that entered the anthropoid primate lineage 45–60 million years ago [45]. In the human genome there are currently 7190 MER41 copies divided into 6 subfamilies (A–E, G), which range on average from 83 to 91% nucleotide similarity to their respective consensus sequence [52]. To identify MER41-mediated interactions, we designed a probe of 120 nucleotides around DpnII restriction enzyme sites in the MER41B consensus sequence (Fig. 4a) (a subfamily of MER41 that contains STAT1 binding sites). HeLa cells treated with interferon gamma (IFNγ) were selected to test the ability of this probe to enrich for MER41 interactions as this system was previously used to demonstrate the involvement of a subset of MER41 elements as interferon-inducible enhancers regulating adjacent immune genes [45]. Using Blastn, we predicted

that 3452 (49.8%) individual MER41 copies would be captured by our probes (1790 MER41A, 1609 MER41B, and 180 MER41C) (Fig. 4b).

In addition to the Capture-4Tran library, we sequenced a Hi-C library for comparison. Capture-4Tran yielded ~ 291 times more sequencing reads associated with MER41B-mediated interactions than the Hi-C approach (5 million unique contacts out of 19.0 million versus 55.0 thousand out of 61 million contacts in Hi-C). The reads generated a contact matrix that is clearly distinct from a Hi-C matrix in that TADs and compartments are no longer distinctly visible. Instead, strong horizontal and vertical lines stand out that are centered at the diagonal of the matrix and emanate from MER41 elements (Fig. 4c). To determine whether the MER41 probe hybridized to DpnII fragments not predicted by Blastn analysis, we calculated the number of reads per DpnII fragment and considered only those belonging to the top 0.005% (see the "Methods" section for details). Altogether, 1766 DpIIn fragments representing 1687 distinct MER41 LTR elements were detected. In addition, 272 fragments that were not predicted by Blastn were identified, of which 183 contained MER41 elements and the remaining 86 (representing less than 5% of all captured fragments) corresponded to regions without annotated MER41 elements (Fig. 4b). Thus, using a single oligonucleotide we are able to detect a substantial fraction of copies (24.2%) from an abundant and relatively ancient TE family with a high rate of specificity (95%). The small percentage of off-target enrichment suggests that inclusion of a few extra probes hybridizing to other regions of the consensus sequence, or targeting the consensus sequence of other MER41 subtypes (here an oligo targeting only MER41A, B and C was used) would easily allow capture of the remaining family of MER41 LTR elements.

Characterization of significant interactions from MER41 elements

The Chicago pipeline was used to identify specific pairwise interactions between MER41 elements and other genomic regions [68]. Using a Chicago score threshold of 7.5, 4913 interactions from 943 MER41 elements were detected in HeLa cells treated with IFNγ and 2107 interactions from 461 MER41 elements in control HeLa cells. A mean value of 6.3 interactions (median value of 2) per bait was identified (Additional file 1: Figure S7a). The difference in the absolute number of interactions identified could be due to the sequencing depth of the samples, therefore to account for this we performed differential analysis based on the read count at the interactions using DESeq2. We detected 9 interactions with differential contact frequency (Additional file 2: Table S4) after IFNγ induction (see the "Methods" section for details). Since the majority of interactions are stable between the control and treatment cells,

we merged the interactions that were not called as differential (6004 unique contacts) for downstream analysis.

To characterize the MER41 elements captured by our probe, we obtained all ChIP-Seq datasets available for HeLa cells from ENCODE and overlapped enrichment peaks with the MER41 elements captured by our probe. Of the 327 MER41 elements that overlapped with a peak, the majority were bound by STAT1. (Fig. 5a). Although most of the MER41 elements were not enriched for any ChIP-Seq signal (Fig. 5b), we did not observe a difference in the mappability score or the number of interactions of these elements from those that did overlap with detected ChIP-Seq peaks (Additional file 1: Figure S7c, d), which suggests that lack of ChIP-Seq signal is not related to problems arising from alignment caused by MER41 repetitiveness.

We verified the position of the interactions relative to TADs in Capture-4Tran data from both control and IFNγ treated cells and as expected, we observed that ~ 75% of interactions occurred between loci within the same TAD as the MER41 element (Fig. 5c). In line with this observation, the median distance between MER41 capture-baits and their *cis* interactions was found to be ~ 162.6 kb (less than the average size of TADs). (Fig. 5d). Next, we asked whether loci that interact with MER41 are significantly enriched for STAT1 binding and open chromatin marks as assessed by FAIRE-Seq. The overlap of interactions with STAT1 peaks, FAIRE-Seq and gene bodies was found to be significant for all these features compared to randomized backgrounds (Fig. 5e). MER41 elements also interacted more frequently with gene-rich regions than expected by chance (Additional file 1: Figure S7b). Furthermore, a small number of interactions occur between MER41 elements (Additional file 1: Figure S7e).

We next assessed MER41 interactions with the promoter region of genes (TSS ± 2 kb) and focused on the most frequent contacts (Chicago interaction score > 10 and a linear distance less than 5 Mb from the MER41 anchor) and identified 377 interactions (Additional file 2: Table S5). In general, each MER41 element contacted a single promoter; however, some elements interacted with 2 or more promoters (Fig. 5f). In addition, most promoters were found in contact with 1 or 2 MER41 elements, with 38 out of 287 contacting multiple MER41 elements (Fig. 5f). The majority of the interactions that overlapped within 2 kb of a TSS site have either a STAT1 or FAIRE-Seq peak (Fig. 5g). Of the 175 genes induced upon interferon gamma (IFNγ) treatment identified by microarray analysis (see the "Methods" section for details) [7], only three genes, IFI6, IFITM1, and IL4R, significantly interacted with a MER41 LTR element (Fig. 5h). Other genes that are not affected by IFNγ treatment such as SERPINB6, also interacted with MER41B elements (Additional file 1: Figure S8a) suggesting that these
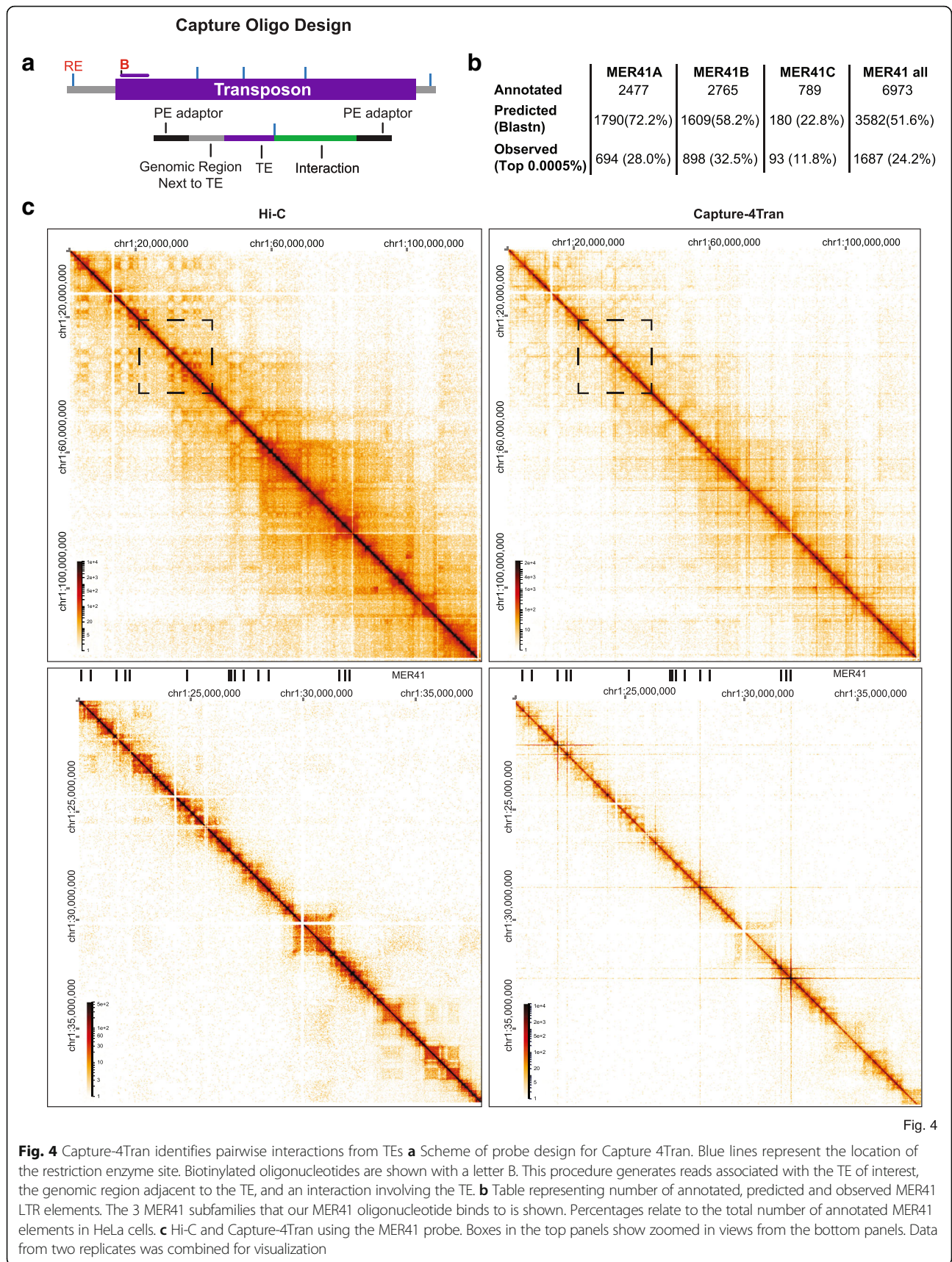
Fig. 4 Capture-4Tran identifies pairwise interactions from TEs **a** Scheme of probe design for Capture 4Tran. Blue lines represent the location of the restriction enzyme site. Biotinylated oligonucleotides are shown with a letter B. This procedure generates reads associated with the TE of interest, the genomic region adjacent to the TE, and an interaction involving the TE. **b** Table representing number of annotated, predicted and observed MER41 LTR elements. The 3 MER41 subfamilies that our MER41 oligonucleotide binds to is shown. Percentages relate to the total number of annotated MER41 elements in HeLa cells. **c** Hi-C and Capture-4Tran using the MER41 probe. Boxes in the top panels show zoomed in views from the bottom panels. Data from two replicates was combined for visualization
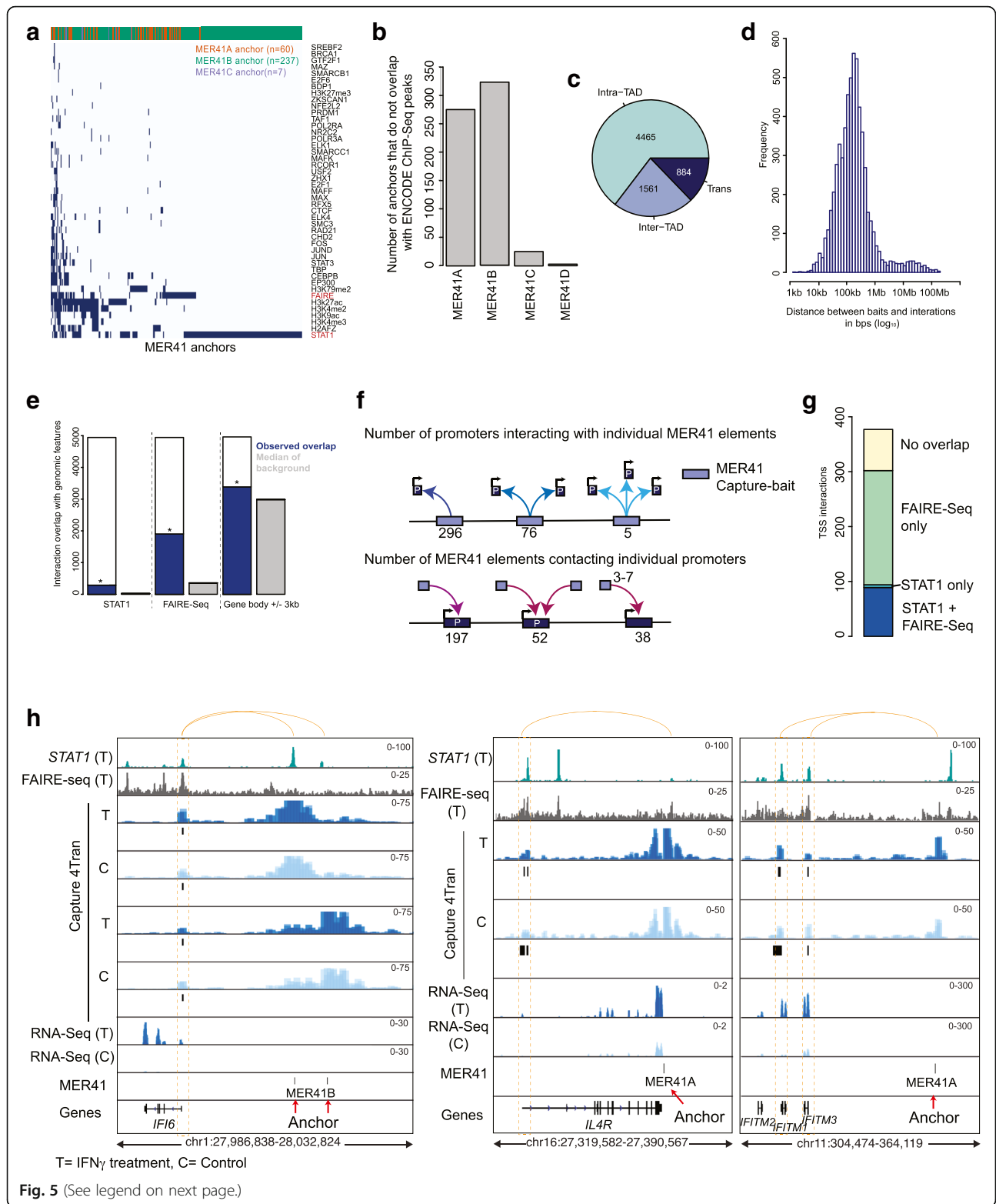
Raviram *et al. Genome Biology*　　(2018) 19:216

Page 11 of 19

Fig. 5 (See legend on next page.)

(See figure on previous page.)

Fig. 5 Interactions involving MER41 elements. **a** Overlap of MER41 anchors with ChIP-Seq peaks from ENCODE. Only STAT1 and FAIRE-Seq data were derived from HeLa cells treated with interferon-γ. **b** Breakdown of the MER41 anchors that do not overlap with any ChIP-Seq peaks. **c** Interactions are classified whether they are in *cis* (same chromosome) or *trans* (other chromosomes) and *cis* interactions are further divided based on whether they are in the same TAD as the MER41 anchor element. **d** Histogram of the distance between MER41 anchors and their *cis* interactions. **e** Overlap of interactions (± 1 kb) with STAT1 ChIP-Seq peaks, FAIRE-Seq peaks and gene bodies ± 3 kb. The background regions were generated based on randomly shuffling the position of the interactions on the same chromosome and calculated the overlap with each feature (median of 1000 iterations displayed). Significance was calculated based on the number of times an overlap with the randomized interaction positions is greater than observed and divided by 1000. **f** Representation of the number of promoters each MER41 element was found to interact with (top panel) and the number of MER41 elements contacting each promoter (bottom panel). **g** Number of TSS interactions that overlap with STAT1 or FAIRE-Seq peaks. **h** Genomic tracks of STAT1 ChIP-Seq signal, FAIRE-Seq signal and Capture 4Tran data from HeLa cells with (T) and without IFNγ treatment. Black boxes below Capture 4Tran data represent interactions identified by Chicago (C). Red arrow represents the MER41B anchor and the orange dotted rectangle represents the interaction at the IFI6, IL4R, IFITM1, and IFITM3 promoter

interactions might be mediated by other transcription factors or require additional factors for gene activation.

It has previously been established that a MER41B element located ~ 20 kb downstream of *IFI6* contributes to its IFNγ-mediated transcriptional upregulation. Specifically, deletion of this element leads to an approximately twofold reduction in the level of IFNγ-mediated IFI6 transcripts [45]. Capture-4Tran data clearly indicates that this MER41B LTR integration is engaged in an interaction with the promoter of *IFI6* (Fig. 5h). Moreover, 4Tran identified a second MER41B element inserted further upstream in contact with the *IFI6* promoter, suggesting that this element could also contribute to the regulation of *IFI6*.

4Tran also captured an interaction between the *IFIMT1* and *IFITM3* genes and a MER41A element downstream of these genes (Fig. 5h). This element is within a cluster of several MER41 elements and is immediately upstream of a STAT1 binding site. It is of note that a previous study aimed at identifying enhancers of the *IFITM* genes, used CRISPR to delete sequences in the region encompassing the MER41A element and found that those overlapping the STAT1 binding site act as an important enhancer of the *IFITM1* cluster of genes [69]. Thus, it is possible that the interaction between the MER41A and the IFITM genes is responsible for, or contributes to the contacts between the STAT1 binding region and the IFITM promoters. However, although both *IFI6* and *IFITM1* are strongly up-regulated by IFNγ in HeLa cells, the contacts between MER41 elements and the promoters are established prior to treatment, indicating that STAT1 binding is not responsible for initiating the physical regulatory interactions at these loci.

A subset of genes known to be regulated by MER41 elements in response to IFNγ in HeLa cells such as *AIM2* [45] was not identified by Capture-4Tran. The most likely explanation is that these elements are too close to their target promoters on the linear chromosome (< 5 Kb) to be detected by this assay. Furthermore, our probe does not hybridize to some of the genes shown to be

regulated by elements of MER41 subfamilies (e.g., the element regulating *AIM2* belongs to the family MER41E, which was not targeted by our probe). In sum, our analyses suggests that multiple MER41 elements in the genome might function as regulatory elements, and that Capture-4Tran can be used to identify these elements as well as a list of target genes that could be further validated using other experimental approaches.

MER41 elements are involved in several potential regulatory interactions

As many interactions do not involve STAT1 (Fig. 5b), we also investigated whether other transcription factors bind to MER41 interacting regions. We divided the interactions into those that overlap TSS sites (± 3 kb), gene bodies or inter-genic regions. Most interactions at TSS sites overlapped with active histone marks (Fig. 6a), suggesting the genes they are associated with are active genes. In addition, we observed several clusters of interactions at TSSs gene bodies and intergenic regions that are bound by common factors such as a cluster of interactions bound by STAT3, EP300, JUN, and JUND or that overlapped with binding of the architectural proteins CTCF, SMC3, and RAD21(Fig. 6a).

We highlight an example of an interaction occurring with the *MYPN* gene, which encodes a protein specifically expressed in heart and skeletal muscle (Additional file 1: Figure S8b) [70]. This gene is not affected by IFNγ treatment, and its TSS is not bound by STAT1 (Fig. 6b). The MER41 element upstream of the gene promoter is bound by STAT1 and interacted with regions that are bound by other transcription factors. The majority of interactions in this region seem to be restricted by a CTCF binding sites ~ 10 kb upstream of the MER41 elements and another one in an intron of the *MYPN* gene, suggesting that the regulation of this gene is most likely through an element in this domain. The examples provided here show that we can identify the transposable elements that engage in long-range genomic interactions as well as the target genes that they might regulate.
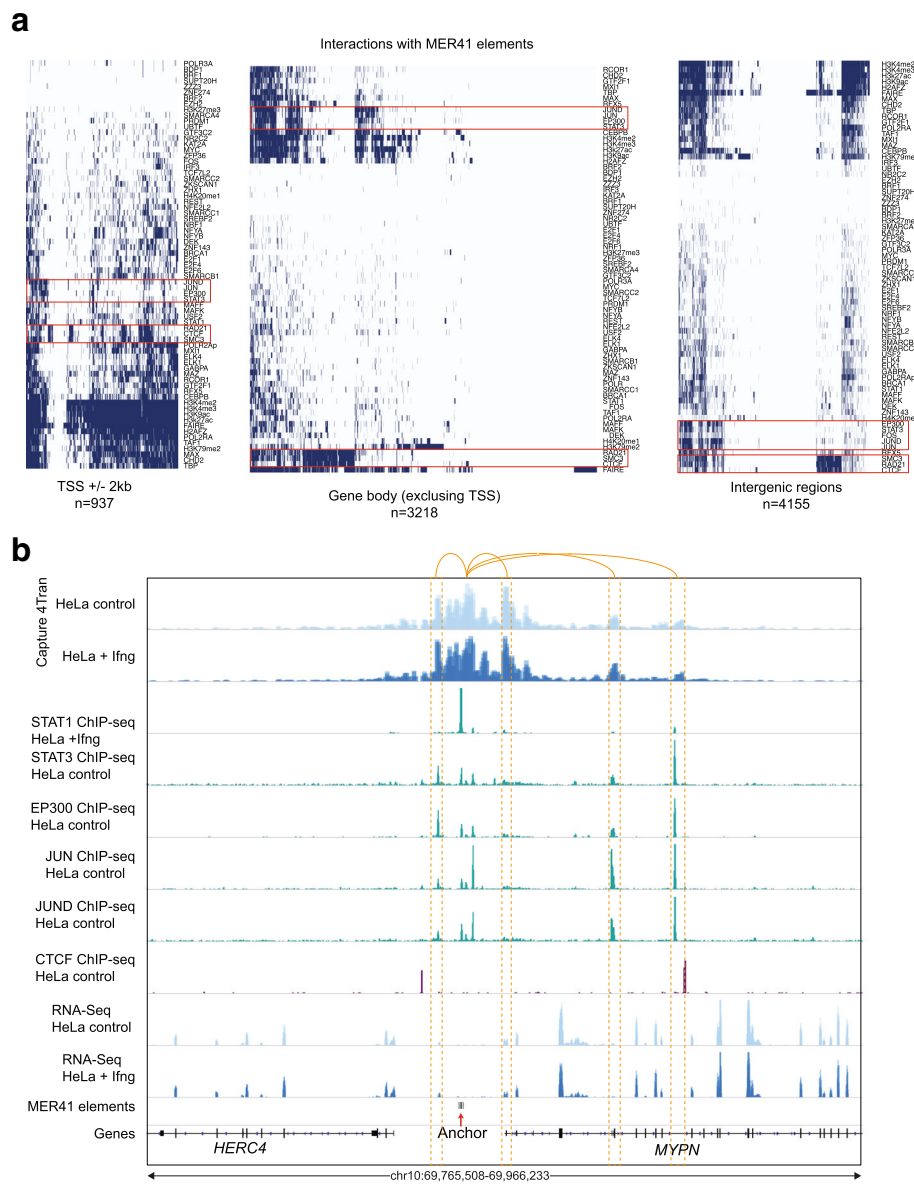
Fig. 6 Interactions between MER41 elements and their overlap with transcription factor binding. **a** Overlap of interactions (± 1 kb) with several ChIP-Seq peaks from ENCODE. Interactions are separated based on whether they are at TSS sites, gene body or intergenic regions. **b** Genomic tracks of ChIP-Seq signal and 4Tran data from control and IFNγ treated HeLa cells. Red arrow represents the MER41B anchor and the orange dotted rectangle represents the interactions

Discussion

We utilized a set of 3C-based tools that takes advantage of the repetitive nature of transposons to study how they impact nuclear organization. Our findings provide insight into how transposons interact with other components of the genome, including host genes, and are themselves influenced by chromosome organization. Specifically, we established two different iterations of chromosome conformation capture, a PCR and a Capture-based variant that we collectively call 4Tran. We applied these tools to endogenous retroviruses that have retained the ability to

replicate and insert themselves elsewhere in the genome and show that 4Tran can detect both fixed insertions and those that remain polymorphic in the population. Interestingly, young ERVs that have integrated into an active A compartment engage in long-range interactions with other A compartment loci of the genome, despite being heterochromatically silenced (as judged by H3K9me3 enrichment). In addition, we find that chromatin interactions of ERVs are mostly delimited by the organization of chromosomes into smaller, highly self-interacting domains and their chromosomal looping is largely restricted to these

regions. Finally, we demonstrate that Capture-4Tran can be used to identify TEs that establish direct looping interaction with adjacent host gene promoters and thereby likely modulate host gene transcription.

The choice between use of the PCR-based and the capture-based variants of 4Tran will depend on the biological question. 4Tran-PCR works best with younger TE families that have a moderate number (up to a few hundred) of integration events with high sequence similarity between different copies. It provides high-depth sequencing data at low cost, is easier to implement than Capture-4Tran, and can be used to characterize long-range TE contacts and boundaries of domains that restrict their local interaction. Capture-4Tran, on the other hand, can be used for analysis of younger TEs, as well as older families with high sequence divergence. Indeed, it bypasses most of the limitations of 4Tran-PCR, including small size and it is highly specific for the elements it captures. As a result, it can be used to study high copy number transposons such as SINEs, LINEs, and solo LTRs.

A number of recent studies have focused on the role of transposons in gene regulation. In the absence of a chromosome conformation capture tool, TE influence has been examined in the limited context of the nearest gene. However, TEs can participate in both long- and short-range interactions and could potentially be involved in the regulation of multiple target loci. 4Tran therefore provides a tool for more precisely defining the extent of this control and identifying TE-mediated interactions with specific promoter regions.

Capture-4Tran has a clear advantage over promoter Capture Hi-C [47], which enriches for all DNA interactions involving gene promoters. For example, it will only enrich for promoter interactions if a gene is in contact with the TE of interest. This results in reduced numbers of uninformative reads and thus a lower sequencing depth is required, making it more affordable to analyze replicates and compare different conditions. Moreover, by designing a capture probe next to the TE integration site, it is possible to detect exactly which repetitive element is engaged in an interaction with a gene promoter, while analysis of promoter-based contacts is less likely to identify interactions with the repetitive portion of younger TEs with a low level of sequence divergence between copies.

Capture-4Tran results are consistent with a model in which a subset of MER41 elements act as interferon-inducible enhancers that contribute to STAT1-mediated transcriptional activation of interferon sensitive genes in response to pathogen infection. Our results show that MER41 elements function as a prototypical enhancer by making contact with the promoters of flanking genes. Furthermore, our analysis uncovers many direct MER41-gene associations that further support the notion that these primate-specific endogenous retroviruses have a substantial impact on gene regulation in human cells and represent an important force driving the regulatory evolution of the primate innate immune response. These experiments add to a growing body of evidence demonstrating two types of enhancer-promoter contacts: stable and dynamic. Stable contacts, such as the majority of enhancer-promoter contacts identified during a TNFα response, are formed prior to signaling [71]. These are consistent with our observations for MER41 in IFNγ stimulated cells. Similarly, genes activated under conditions of hypoxia by the HIF transcription factor are engaged in stable DNA contacts with their enhancer elements [72]. In contrast, DNA enhancer promoter loops can be dynamically rewired during differentiation and cell-type-specific DNA contacts established only when enhancers are activated by binding of cell-type-specific transcription factors [66, 73–78] . The situation in *Drosophila* contrasts with the findings from mammalian studies showing that pre-established loops connect enhancers and promoters even before gene activation in early development [79]. The fact that both interferon and TNFα responsive genes establish loops with similar dynamics suggests that there is a fundamental difference between transcriptional responses to stress/pathogens and those underlying developmental transitions. It is tempting to speculate that the interferon response is similar to TNF-α signaling and that there is a fundamental difference between transcriptional responses to stress/pathogens and those underlying developmental transitions.

Despite major advances in the field since the development of chromosome conformation, there is still a lot to be learned about the role of nuclear architecture in gene regulation. In particular, our knowledge of the contribution of TEs in these processes is lagging behind because of an absence of tools amenable to their analysis. Here, we provide the first 3C-based approach for addressing this question. 4Tran is important for learning how TEs such as MER41 influence transcription of host genes, and this in turn provides new insight into the mechanisms by which TEs contribute to the evolution of gene regulatory networks. Furthermore, 4Tran can be used to study the impact of TEs in disease settings such as cancer where the methylation status of the genome is altered and transposons become activated [80, 81].

Conclusions

It is well established that TEs represent an important source of genetic variation [42]. These mobile elements were first identified in the late 1940s by Barbara McClintock, who named them controlling elements for their ability to affect gene expression in maize [39, 40]. However, it has only recently become apparent that TEs bound by transcription factors [35, 41] can have a major impact on gene regulation [42, 45]. Indeed, the spreading of multiple TF motifs in the genome by transposons

Raviram *et al. Genome Biology*   (2018) 19:216

Page 15 of 19

is thought to be important in driving the evolution of gene regulatory networks. In addition, TEs have contributed to the organization of mammalian genomes by propagating binding sites, for the architectural protein, CTCF [35, 82]. However, it should be noted that not all TE integrations influence chromosome folding and gene expression as some do not harbor binding sites for regulatory factors and those that do, are often actively repressed [21] and/or lose these sequences over time [46]. We used 4Tran to characterize interactions between TEs and neighboring regions and to determine how they are physically organized in the nucleus. Here we show the usefulness of this approach in identifying contacts between transposons and the promoters of genes whose expression these elements have the capacity to control.

Methods

Mice
All mice used here were of wild-type inbred strains. Animal care was approved by Institutional Animal Care and Use Committee. Protocol number is 160704-02 (NYU School of Medicine). Young mice of less than 12 weeks were used in all experiments.

Preparation of template for 4Tran-PCR
4C-Seq material was prepared from the following murine cells: embryonic stem cells, embryonic fibroblasts, and splenic B cells. Starting material for each sample was 10 million cells. The 4C template material prepared from mouse embryonic stem cells was previously described [15, 83]. These ES cells are from the ATCC cell line #CRL-1821 and were obtained from a 129/Ola mouse strain. Mouse embryonic fibroblasts were prepared from E13.5 embryos of the 129S1 strain. Embryos were isolated from timed-matings and dissociated using trypsin after removal of internal organs and decapitation. Cells were cultured for two passages before 4C material preparation. Resting mature splenic B cells were isolated as previously described [84] using magnetic beads for CD43 depletion from either C57/Bl6 Taconic mice (C57BL/6NTac) or from 129S6 mice also from Taconic (129S6/SvEvTac). Processing of 4C material was performed as described previously [64] using DpnII and Csp6I as enzymes for DNA digestion. Briefly, cells were fixed with formaldehyde and digested in situ with DpnII. Following cell-lysis DNA fragments were diluted to favor proximity-mediated ligation. Concatemers of DNA fragment ligations were subsequently digested with Csp6I and religated upon dilution. 4C template was then amplified by 30 PCR cycles using 1 μg of DNA divided by ten 50 μl PCR reactions. Data is not shown here but we have also successfully tested in situ ligation by simply omitting the SDS treatment step following digestion with the primary restriction enzyme. Library amplification was

done using Illumina single end adaptors, however, the protocol also works with paired-end adaptors which are necessary for Illumina machines (Nextseq and Miseq for example). A list of samples generated in the study can be found in Additional file 2: Table S1 and a list of primers used for amplification can be found in Additional file 2: Table S2.

Bait design and 4Tran-PCR
For each TE, 4C-Seq primers were designed to the consensus sequence obtained from Repbase. We then designed our algorithm to find all possible primer pairs based on the desired enzyme digestion. To be considered as a potential bait, the following criteria are required: 300 bp separating primary enzyme restriction sites and 150 bp between primary and secondary enzyme restriction site. For these locations a 20-nucleotide primer is designed that includes the primary restriction enzyme site and the second primer is designed using the default parameters of Primer 3. We used annotations from Repeatmasker based on the "-int" sequence for all repeats except RLTR4 where both the "-int" and solo LTR sequences were included since our probe was able to capture many solo LTRs.

4Tran-PCR analysis
Single-end reads were mapped to a reduced genome of unique 25 bp fragments adjacent to DpnII sites (mm10) using bowtie2 [85] as per the details described in 4C-ker [65]. Reads that map to the Encode blacklist regions were removed from downstream analysis [86]. To remove fragments that arise from self-ligation or incomplete digestion, only counts below a quantile of 99.9% were retained. For visualization, we generated counts in 100 kb windows, overlapping by 25 kb. In addition, to decrease the effect of PCR artifacts, counts in each window that are greater than the 75% quantile were reduced to that value. To define "observed baits," we used read counts obtained in 100 kb non-overlapping windows. A z-score was then defined for each window based on the mean and standard deviation across all windows in the genome. Finally, an FDR adjusted p value was calculated based on fitting to a normal distribution and enriched windows were defined based on p values smaller than 0.05. Windows separated by 100 kb were merged. Only those windows identified in both replicates were called as observed baits. To determine the location of predicted baits, we used the UCSC in silico PCR tool. One of the primers for each bait was input as the reverse complement and the lowest stringency mismatch option (15 nucleotides) was chosen. To determine the regions interacting most frequently with specific TE integration events (Fig. 3), we used the 4C-ker pipeline by performing cis analysis using a k of 10.

Preparation of Hi-C and Capture-4Tran material

Murine B cell material was prepared as described in the section above. Human HeLa F2 cells were generous gifts from Cedric Feschotte and Edward Chuong. Cells were treated for 24 h with 1000 U/ml of recombinant human IFNγ (cat# 11500-2, PBL assay science). Following treatment cells were fixed with Formaldehyde for 4Tran-Capture as well as Hi-C. To verify that cells responded to interferon treatment, RNA was collected and expression of the *IFI6* gene was measured by reverse transcriptase quantitative PCR. Primers can be found in Additional file 2: Table S2.

Capture-4Tran and Hi-C

Our approach was adapted from two published protocols. The Hi-C part of the protocol was performed mostly as described in [87]. Ligation of adaptors and Ilummina indices as well as DNA capture were performed as detailed in [51]. Briefly, 10 million HeLa cells per replicate and condition were washed with PBS, trypsinized and spun down before resuspension in media containing 2% formaldehyde. Fixation was done for 10 min at room temperature and stopped by adding Glycine. Nuclei were obtained after lysis on ice for 30 min with 10 mM Tris-HCl pH 8, 10 mM NaCl, 0.2% Igepal CA-630 and 1 Roche Complete EDTA-free tablet. After 10 min treatment with 0.1% SDS at 37 °C, nuclei were incubated with Triton 10% for 10 min also at 37 °C. Nuclei were then suspended in 1X DpnII buffer (NEB) and chromatin digested overnight with 400 U of DpnII (NEB) at 37 °C. Following an additional 4-h incubation with extra 400 U of DpnII, digestion was stopped by 65 °C incubation for 30 min. DNA ends were then filled in with dATP-Biotin using polymerase klenow for 60 min at 37 °C. Ligation was performed in situ overnight (with nuclei intact) using 2000 U of T4 DNA ligase from NEB (M0202) at 16 °C. Decrosslinking was performed overnight with proteinase K at 65 °C. Following two phenol chloroform extractions the efficiency of biotin fill in was verified using PCR and digestion with ClaI (primers in Additional file 2: Table S2). Forty micrograms of unligated biotin ends was removed by incubation with T4 DNA polymerase for 4 h at 20 °C. DNA fragments were then sonicated using an LE Covaris 220 instrument using 450 PIP, 30% duty factor, 200 cycles per burst for 60 s. Following sonication, DNA fragments were end-repaired using T4 DNA polymerase, T4 PNK and Klenow polymerase (all from NEB). DNA fragments were then bound to Streptavidin C1 beads and washed following manufacturer's instructions and using 150 μl per 2.5 μg of sonicated DNA. A-tailing and adaptor ligation was then performed with DNA fragments bound to beads. A-tailing was done using Klenow polymerase without 3′ exonuclease activity. For ligation, adaptors

from the NEBNext DNA library prep reagent set (E6040) were used and ligated overnight. Following ligation, enzyme USER treatment was used to open the stem loop. Hi-C library amplification was then amplified using 7 PCR cycles, with Phusion polymerase and PCR primers from NEBNext Multiplex Oligos for Illumina (E7335) and resulted in approximately 1 μg of Hi-C material. For the Hi-C sample, preparation was stopped here and the sample corresponding to material treated with Interferonγ sequenced. For Capture, we used the Nimblegen SeqCap EZ hybridization and wash kit (05634261001), Nimblegen SeqCap EZ accessory kit v2 and Nimblegen SeqCap EZ HE-oligo kit. Briefly, 1 μg of Hi-C material was incubated at 42 °C for 48 h with 2.89 μmol of a MER41 biotinylated DNA oligo according to Nimblegen's instructions. Following recommended washes, beads were amplified using 8 cycles, with the Post-LM-PCR Oligos 1 & 2 oligos described in Nimblegen's kit and Kappa 2× PCR master mix. This yielded approximately 700 ng of DNA, which was used for a second round of capture as described above for 24 h. Final amplification of libraries following second capture was done using 5 PCR cycles. In total 20 PCR cycles were used to amplify the libraries and approximately 100 ng of material were prepared for sequencing using Illumina Nextseq paired-end 50 base pair reads.

Hi-C and Capture 4Tran-capture analysis

Processing of Hi-C and Capture-4Tran reads was performed using Hicup [88]. Reads were mapped to the hg19 genome. Prediction of sites that hybridize to the MER41 probe was performed using Blastn from ensemble (ensembl.org) with default parameters, without Repeatmasker filtering and allowing for up to 5000 hits. To detect fragments with high number of long-range interactions, each end of a mapped read-pair was first assigned to a restriction fragment in the genome using juicertools. Any read pairs that were within 100 fragments were removed. For the remaining interactions, the number of unique reads was counted and a quantile cut-off of 0.0005 was used to select the observed fragments with a high number of reads. This number was selected based on the percentage of DpnII fragments that Blastn predicts which are enriched by our oligonucleotide. Chicago [68] was used to identify interactions using the default parameters with a score cut-off of 7.5 on 2 replicates. The DpnII fragments with high number of reads were used as the baits (anchors) in Chicago and any fragment that did not overlap with a MER41 element was removed from downstream analysis (48/992). We used Cooler (github.com/mirnylab/cooler) to process the data and HiGlass (higlass.io/) to visualize Capture-4Tran and Hi-C results. Normalization of the Capture-4Tran data was performed using DESeq2 on read

counts in 1 kb bins overlapping by 500 bp in a 1 MB region around the bait. Normalized bedGraph files were used for visualization of data in IGV.

Differential analysis for the 6026 interactions was performed using DESeq2. The input for the analysis was the counts in the interacting region for each sample (2 replicated per condition) and adjusted p value cut off of 0.01 and a log2 fold change of 2 was used to call significant interactions.

Published Hi-C data analysis and Capture-4Tran analysis
Paired-end reads were mapped separately using bowtie 2.0.1. Downstream analysis was performed using Homer default parameters to obtain the count matrix and PCA scores to determine A and B compartment composition.

Published transcriptional data analysis of interferon treatment
To identify genes upregulated by interferon treatment in HeLa cells we used a microarray dataset [89] as this study contained three different replicates for each condition and was therefore more reliable. To identify upregulated genes we used the GEO2R tool from the NCBI GEO website (ncbi.nlm.nih.gov/geo) and used as criteria a log fold change of 1 and an adjusted p-value of 0.05. To identify all genes that are potentially expressed in HeLa cells we used RNA-seq data (SRR2992615-48Hr Control, SRR2992616-48Hr IFNγ, SRR2992619-72Hr Control, SRR2992620-72Hr IFNγ) [90]. Single-end RNA-Seq data was mapped to the hg19 genome using Tophat v2.1.1 (--no-coverage-search −N 0 --b2-very-sensitive). Htseq-counts was used to count the reads per gene using the hg19 Refseq annotation of genes. The counts were then normalized using DESeq2 for sequencing depth and further for gene length by dividing the normalized count by gene length and multiplying by 1e3.

Additional files

Additional file 1: Supplementary Figures. This file contains all the supplementary figures for this paper. (PDF 4494 kb)

Additional file 2: Supplementary Tables. This file contains 5 supplementary tables for this paper. **Table S1.** List of 4Tran samples. **Table S2.** List of oligos used for 4Tran. **Table S3.** Integration differences between mouse strains with annotations of sequence variants from the Mouse Genome Project Database. **Table S4.** Differential interactions. **Table S5.** List of Potential Regulatory Interactions. (XLSX 151 kb)

Availability of data and materials
All massively parallel sequencing-based datasets have been deposited in GEO under the accession number GSE122977 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi) [91]. Scripts for bait design can be found at https://github.com/rr1859/4Tran under the GPL-3.0 license and on zenodo at https://zenodo.org/record/1479954 [92].

Authors' contributions
RR, PPR, RB, and JAS conceived the study and designed the experiments. RR, PPR, VML, and ES performed experiments. RR and PPR analyzed the data. ERM, EBC, and CF helped with initial planning of experiments and data analysis. RR, PPR, and JAS wrote the manuscript. All authors read, edited, and approved the final manuscript.

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Department of Pathology, New York University School of Medicine, New York, NY 10016, USA. [2]Department of Biology, New York University, New York, NY 10003, USA. [3]Ludwig Institute for Cancer Research, La Jolla, CA, USA. [4]Division of Developmental Biology, Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, Bethesda, MD 20892, USA. [5]Department of Developmental Genetics, New York University School of Medicine, New York, NY 10016, USA. [6]Department of Computer Science, Courant Institute of Mathematical Sciences, New York, NY 10003, USA. [7]Simons Center for Data Analysis, New York, NY 10010, USA. [8]Divisions of Immunobiology and Biomedical Informatics, Cincinnati Children's Hospital, Cincinnati, OH 45229, USA. [9]BioFrontiers Institute, Department of Molecular, Cellular, and Developmental Biology, University of Colorado Boulder, Boulder, CO 80309, USA. [10]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14850, USA.

References
1. Rocha PP, Raviram R, Bonneau R, Skok JA. Breaking TADs: insights into hierarchical genome organization. Epigenomics. 2015;7:523–6.
2. Bonev B, Cavalli G. Organization and function of the 3D genome. Nat Rev Genet. 2016;17:661–78.
3. Fraser J, Williamson I, Bickmore WA, Dostie J. An overview of genome organization and how we got there: from FISH to hi-C. Microbiol Mol Biol Rev. 2015;79:347–72.
4. Nicodemi M, Pombo A. Models of chromosome structure. Curr Opin Cell Biol. 2014;28:90–5.
5. Denker A, de Laat W. The second decade of 3C technologies: detailed insights into nuclear organization. Genes Dev. 2016;30:1357–82.
6. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326:289–93.
7. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature. 2012;485:376–80.
8. Dekker J, Mirny L. The 3D genome as moderator of chromosomal communication. Cell. 2016;164:1110–21.

Raviram *et al. Genome Biology*   (2018) 19:216

Page 18 of 19

9.   Nora EP, Dekker J, Heard E. Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods? Bioessays. 2013;35:818–28.

10. Schmitt AD, Hu M, Ren B. Genome-wide mapping and analysis of chromosome architecture. Nat Rev Mol Cell Biol. 2016;17:743–55.

11. Krijger PH, de Laat W. Regulation of disease-associated gene expression in the 3D genome. Nat Rev Mol Cell Biol. 2016;17:771–82.

12. Hnisz D, Day DS, Young RA. Insulated neighborhoods: structural and functional units of mammalian gene control. Cell. 2016;167:1188–200.

13. Corces MR, Corces VG. The three-dimensional cancer genome. Curr Opin Genet Dev. 2016;36:1–7.

14. Lupianez DG, Spielmann M, Mundlos S. Breaking TADs: how alterations of chromatin domains result in disease. Trends Genet. 2016;32:225–37.

15. Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, Reinberg D. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science. 2015;347:1017–21.

16. Hanssen LLP, Kassouf MT, Oudelaar AM, Biggs D, Preece C, Downes DJ, Gosden M, Sharpe JA, Sloane-Stanley JA, Hughes JR, et al. Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. Nat Cell Biol. 2017;19(8):952–61.

17. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 2011;13:36–46.

18. Huang CR, Burns KH, Boeke JD. Active transposition in genomes. Annu Rev Genet. 2012;46:651–75.

19. Byrd K, Corces VG. Visualization of chromatin domains created by the gypsy insulator of Drosophila. J Cell Biol. 2003;162:565–74.

20. Cournac A, Koszul R, Mozziconacci J. The 3D folding of metazoan genomes correlates with the association of similar repetitive elements. Nucleic Acids Res. 2016;44:245–55.

21. Friedli M, Trono D. The developmental control of transposable elements and the evolution of higher species. Annu Rev Cell Dev Biol. 2015;31:429–51.

22. Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, Martin L, Ware CB, Blish CA, Chang HY, et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature. 2015;522:221–5.

23. Klawitter S, Fuchs NV, Upton KR, Munoz-Lopez M, Shukla R, Wang J, Garcia-Canadas M, Lopez-Ruiz C, Gerhardt DJ, Sebe A, et al. Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. Nat Commun. 2016;7:10286.

24. Jachowicz JW, Bing X, Pontabry J, Boskovic A, Rando OJ, Torres-Padilla ME. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. Nat Genet. 2017.

25. Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature. 2012;487:57–63.

26. Upton KR, Gerhardt DJ, Jesuadian JS, Richardson SR, Sanchez-Luque FJ, Bodea GO, Ewing AD, Salvador-Palomeque C, van der Knaap MS, Brennan PM, et al. Ubiquitous L1 mosaicism in hippocampal neurons. Cell. 2015;161:228–39.

27. Robbez-Masson L, Rowe HM. Retrotransposons shape species-specific embryonic stem cell gene expression. Retrovirology. 2015;12:45.

28. Goke J, Ng HH. CTRL+INSERT: retrotransposons and their contribution to regulation and innovation of the transcriptome. EMBO Rep. 2016;17:1131–44.

29. Mita P, Boeke JD. How retrotransposons shape genome regulation. Curr Opin Genet Dev. 2016;37:90–100.

30. Trizzino M, Park Y, Holsbach-Beltrame M, Aracena K, Mika K, Caliskan M, Perry GH, Lynch VJ, Brown CD. Transposable elements are the primary source of novelty in primate gene regulation. Genome Res. 2017;27(10):1623–33.

31. Sundaram V, Choudhary MN, Pehrsson E, Xing X, Fiore C, Pandey M, Maricque B, Udawatta M, Ngo D, Chen Y, et al. Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. Nat Commun. 2017;8:14550.

32. Rebollo R, Farivar S, Mager DL. C-GATE - catalogue of genes affected by transposable elements. Mob DNA. 2012;3:9.

33. Mager DL, Stoye JP. Mammalian endogenous retroviruses. Microbiol Spectr. 2015;3 MDNA3–0009-2014.

34. Kazazian HH Jr, Moran JV. The impact of L1 retrotransposons on the human genome. Nat Genet. 1998;19:19–24.

35. Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH, Liu ET. Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. 2008;18:1752–62.

36. Jacques PE, Jeyakani J, Bourque G. The majority of primate-specific regulatory sequences are derived from transposable elements. PLoS Genet. 2013;9:e1003504.

37. Goke J, Lu X, Chan YS, Ng HH, Ly LH, Sachs F, Szczerbinska I. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell Stem Cell. 2015;16:135–41.

38. Philippe C, Vargas-Landin DB, Doucet AJ, van Essen D, Vera-Otarola J, Kuciak M, Corbin A, Nigumann P, Cristofari G. Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. Elife. 2016;5. https://doi.org/10.7554/eLife.13926.

39. McClintock B. Controlling elements and the gene. Cold Spring Harb Symp Quant Biol. 1956;21:197–216.

40. McClintock B. Chromosome organization and genic expression. Cold Spring Harb Symp Quant Biol. 1951;16:13–47.

41. Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T. Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome Res. 2014;24:1963–76.

42. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet. 2017;18:71–86.

43. Sundaram V, Wang T. Transposable element mediated innovation in gene regulatory landscapes of cells: re-visiting the "gene-battery" model. Bioessays. 2018;40. https://doi.org/10.1002/bies.201700155.

44. Rebollo R, Romanish MT, Mager DL. Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet. 2012;46:21–42.

45. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science. 2016;351:1083–7.

46. Simonti CN, Pavlicev M, Capra JA. Transposable element exaptation into regulatory regions is rare, influenced by evolutionary age, and subject to pleiotropic constraints. Mol Biol Evol. 2017;34:1–14.

47. Mifsud B, Tavares-Cadete F, Young AN, Sugar R, Schoenfelder S, Ferreira L, Wingett SW, Andrews S, Grey W, Ewels PA, et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. Nat Genet. 2015;47:598–606.

48. Aranda-Orgilles B, Saldana-Meyer R, Wang E, Trompouki E, Fassl A, Lau S, Mullenders J, Rocha PP, Raviram R, Guillamot M, et al. MED12 regulates HSC-specific enhancers independently of mediator kinase activity to control hematopoiesis. Cell Stem Cell. 2016;19:784–99.

49. Javierre BM, Burren OS, Wilder SP, Kreuzhuber R, Hill SM, Sewitz S, Cairns J, Wingett SW, Varnai C, Thiecke MJ, et al. Lineage-specific genome architecture links enhancers and non-coding disease variants to target gene promoters. Cell. 2016;167:1369–1384 e1319.

50. Symmons O, Pan L, Remeseiro S, Aktas T, Klein F, Huber W, Spitz F. The Shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. Dev Cell. 2016;39:529–43.

51. Davies JO, Telenius JM, McGowan SJ, Roberts NA, Taylor S, Higgs DR, Hughes JR. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. Nat Methods. 2016;13:74–80.

52. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6:11.

53. Stocking C, Kozak CA. Murine endogenous retroviruses. Cell Mol Life Sci. 2008;65:3383–98.

54. Markopoulos G, Noutsopoulos D, Mantziou S, Gerogiannis D, Thrasyvoulou S, Vartholomatos G, Kolettas E, Tzavaras T. Genomic analysis of mouse VL30 retrotransposons. Mob DNA. 2016;7:10.

55. van de Werken HJ, Landan G, Holwerda SJ, Hoichman M, Klous P, Chachik R, Splinter E, Valdes-Quezada C, Oz Y, Bouwman BA, et al. Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat Methods. 2012;9:969–72.

56. Bulut-Karslioglu A, De La Rosa-Velazquez IA, Ramirez F, Barenboim M, Onishi-Seebacher M, Arand J, Galan C, Winter GE, Engist B, Gerle B, et al. Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. Mol Cell. 2014;55:277–90.

57. Zhang Y, Maksakova IA, Gagnier L, van de Lagemaat LN, Mager DL. Genome-wide assessments reveal extremely high levels of polymorphism of two active families of mouse endogenous retroviral elements. PLoS Genet. 2008;4:e1000007.

58. Stoye JP, Coffin JM. Polymorphism of murine endogenous proviruses revealed by using virus class-specific oligonucleotide probes. J Virol. 1988;62:168–75.

59. Keane TM, Goodstadt L, Danecek P, White MA, Wong K, Yalcin B, Heger A, Agam A, Slater G, Goodson M, et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature. 2011;477:289–94.

60. Nellaker C, Keane TM, Yalcin B, Wong K, Agam A, Belgard TG, Flint J, Adams DJ, Frankel WN, Ponting CP. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. Genome Biol. 2012;13:R45.

61. Sanchez-Luque FJ, Richardson SR, Faulkner GJ. Retrotransposon capture sequencing (RC-Seq): a targeted, high-throughput approach to resolve somatic L1 Retrotransposition in humans. Methods Mol Biol. 2016;1400:47–77.

62. Henssen AG, Henaff E, Jiang E, Eisenberg AR, Carson JR, Villasante CM, Ray M, Still E, Burns M, Gandara J, et al. Genomic DNA transposition induced by human PGBD5. Elife. 2015;4. https://doi.org/10.7554/eLife.10565.001.

63. Collins PL, Kyle KE, Egawa T, Shinkai Y, Oltz EM. The histone methyltransferase SETDB1 represses endogenous and exogenous retroviruses in B lymphocytes. Proc Natl Acad Sci U S A. 2015;112:8367–72.

64. Rocha PP, Raviram R, Fu Y, Kim J, Luo VM, Aljoufi A, Swanzey E, Pasquarella A, Balestrini A, Miraldi ER, et al. A damage-independent role for 53BP1 that impacts break order and Igh architecture during class switch recombination. Cell Rep. 2016;16:48–55.

65. Raviram R, Rocha PP, Muller CL, Miraldi ER, Badri S, Fu Y, Swanzey E, Proudhon C, Snetkova V, Bonneau R, Skok JA. 4C-ker: a method to reproducibly identify genome-wide interactions captured by 4C-Seq experiments. PLoS Comput Biol. 2016;12:e1004780.

66. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 2014;159:1665–80.

67. Lowe CB, Bejerano G, Haussler D. Thousands of human mobile element fragments undergo strong purifying selection near developmental genes. Proc Natl Acad Sci U S A. 2007;104:8005–10.

68. Cairns J, Freire-Pritchett P, Wingett SW, Varnai C, Dimond A, Plagnol V, Zerbino D, Schoenfelder S, Javierre BM, Osborne C, et al. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. Genome Biol. 2016;17:127.

69. Li P, Shi ML, Shen WL, Zhang Z, Xie DJ, Zhang XY, He C, Zhang Y, Zhao ZH. Coordinated regulation of IFITM1, 2 and 3 genes by an IFN-responsive enhancer through long-range chromatin interactions. Biochim Biophys Acta. 1860;2017:885–93.

70. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015;347:1260419.

71. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature. 2013;503:290–4.

72. Platt JL, Salama R, Smythies J, Choudhry H, Davies JO, Hughes JR, Ratcliffe PJ, Mole DR. Capture-C reveals preformed chromatin interactions between HIF-binding sites and distant promoters. EMBO Rep. 2016;17:1410–21.

73. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y, Xu X, Lv X, Hugnot JP, Tanay A, Cavalli G. Multiscale 3D genome rewiring during mouse neural development. Cell. 2017;171:557–572 e524.

74. Freire-Pritchett P, Schoenfelder S, Varnai C, Wingett SW, Cairns J, Collier AJ, Garcia-Vilchez R, Furlan-Magaril M, Osborne CS, Fraser P, et al. Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. Elife. 2017;6. https://doi.org/10.7554/eLife.21926.001.

75. Rubin AJ, Barajas BC, Furlan-Magaril M, Lopez-Pajares V, Mumbach MR, Howard I, Kim DS, Boxer LD, Cairns J, Spivakov M, et al. Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. Nat Genet. 2017;49:1522–8.

76. Siersbaek R, Madsen JGS, Javierre BM, Nielsen R, Bagge EK, Cairns J, Wingett SW, Traynor S, Spivakov M, Fraser P, Mandrup S. Dynamic rewiring of promoter-anchored chromatin loops during adipocyte differentiation. Mol Cell. 2017;66:420–435 e425.

77. Phanstiel DH, Van Bortle K, Spacek D, Hess GT, Shamim MS, Machol I, Love MI, Aiden EL, Bassik MC, Snyder MP. Static and dynamic DNA loops form AP-1-bound activation hubs during macrophage development. Mol Cell. 2017;67:1037–1048 e1036.

78. Snetkova V, Skok JA. Enhancer talk. Epigenomics. 2018.

79. Ghavi-Helm Y, Klein FA, Pakozdi T, Ciglar L, Noordermeer D, Huber W, Furlong EE. Enhancer loops appear stable during development and are associated with paused polymerase. Nature. 2014;512:96–100.

80. Babaian A, Mager DL. Endogenous retroviral promoter exaptation in human cancer. Mob DNA. 2016;7:24.

81. Burns KH. Transposable elements in cancer. Nat Rev Cancer. 2017;17:415–24.

82. Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell. 2012;148:335–48.

83. van de Werken HJ, de Vree PJ, Splinter E, Holwerda SJ, Klous P, de Wit E, de Laat W. 4C technology: protocols and data analysis. Methods Enzymol. 2012; 513:89–112.

84. Rocha PP, Micsinai M, Kim JR, Hewitt SL, Souza PP, Trimarchi T, Strino F, Parisi F, Kluger Y, Skok JA. Close proximity to Igh is a contributing factor to AID-mediated translocations. Mol Cell. 2012;47:873–85.

85. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

86. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.

87. Belaghzal H, Dekker J, Gibcus JH. Hi-C 2.0: an optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. Methods. 2017;123:56–65.

88. Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, Andrews S. HiCUP: pipeline for mapping and processing Hi-C data. F1000Res. 2015;4:1310.

89. Wandel MP, Pathe C, Werner EI, Ellison CJ, Boyle KB, von der Malsburg A, Rohde J, Randow F. GBPs inhibit motility of Shigella flexneri but are targeted for degradation by the bacterial ubiquitin ligase IpaH9.8. Cell Host Microbe. 2017;22:507–518 e505.

90. Timosenko E, Ghadbane H, Silk JD, Shepherd D, Gileadi U, Howson LJ, Laynes R, Zhao Q, Strausberg RL, Olsen LR, et al. Nutritional stress induced by tryptophan-degrading enzymes results in ATF4-dependent reprogramming of the amino acid transporter profile in tumor cells. Cancer Res. 2016;76:6193–204.

91. Raviram R, Rocha PP, Luo VM, Swanzey E, Miraldi ER, Chuong EB, Feschotte C, Bonneau R, Skok JA: Analysis of 3D genomic interactions identifies candidate host genes that transposable elements potentially regulate. 2018. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi. Accessed 8 Nov 2018.

92. Raviram R, Rocha PP, Luo VM, Swanzey E, Miraldi ER, Chuong EB, Feschotte C, Bonneau R, Skok JA. Analysis of 3D genomic interactions identifies candidate host genes that transposable elements potentially regulate. Zenodo. 2018; https://zenodo.org/record/1479954. Accessed 8 Nov 2018.