UC Irvine
UC Irvine Previously Published Works

Title
Local features determine Ty3 targeting frequency at RNA polymerase III transcription start sites

Permalink
https://escholarship.org/uc/item/02n7k9rd

Journal
Genome Research, 29(8)

ISSN
1088-9051

Authors
Patterson, Kurt
Shavarebi, Farbod
Magnan, Christophe
et al.

Publication Date
2019-08-01

DOI
10.1101/gr.240861.118

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Local features determine Ty3 targeting frequency at RNA polymerase III transcription start sites

Kurt Patterson,[1] Farbod Shavarebi,[1] Christophe Magnan,[2] Ivan Chang,[1] Xiaojie Qi,[1] Pierre Baldi,[2] Virginia Bilanchone,[1] and Suzanne B. Sandmeyer[1]

[1] *Department of Biological Chemistry, School of Medicine, University of California, Irvine, Irvine, California 92697, USA;*
[2] *School of Information and Computer Sciences, University of California, Irvine, Irvine, California 92697, USA*

Retroelement integration into host genomes affects chromosome structure and function. A goal of a considerable number of investigations is to elucidate features influencing insertion site selection. The *Saccharomyces cerevisiae* Ty3 retrotransposon inserts proximal to the transcription start sites (TSS) of genes transcribed by RNA polymerase III (RNAP3). In this study, differential patterns of insertion were profiled genome-wide using a random barcode-tagged Ty3. Saturation transposition showed that tRNA genes (tDNAs) are targeted at widely different frequencies even within isoacceptor families. Ectopic expression of Ty3 integrase (IN) showed that it localized to targets independent of other Ty3 proteins and cDNA. IN, RNAP3, and transcription factor Brf1 were enriched at tDNA targets with high frequencies of transposition. To examine potential effects of *cis*-acting DNA features on transposition, targeting was tested on high-copy plasmids with restricted amounts of 5′ flanking sequence plus tDNA. Relative activity of targets was reconstituted in these constructions. Weighting of genomic insertions according to frequency identified an A/T-rich sequence followed by C as the dominant site of strand transfer. This site lies immediately adjacent to the adenines previously implicated in the RNAP3 TSS motif (CAA). In silico DNA structural analysis upstream of this motif showed that targets with elevated DNA curvature coincide with reduced integration. We propose that integration mediated by the Ty3 intasome complex (IN and cDNA) is subject to inputs from a combination of host factor occupancy and insertion site architecture, and that this results in the wide range of Ty3 targeting frequencies.

[Supplemental material is available for this article.]

Nearly all organisms contain Long Terminal Repeat (LTR) and non-LTR retroelements that replicate via reverse transcription of their genomic RNA into a complementary DNA (cDNA) and integrate into the genomes of host cells. The presence of these elements affects regulatory sequences as well as epigenetic modulation (Sultana et al. 2017; Gilbert and Feschotte 2018; Klein and O'Neill 2018). The genomic distribution of retrotransposons and retroviruses is not random (Kvaratskhelia et al. 2014; Sandmeyer et al. 2015; Lesbats et al. 2016; Sultana et al. 2017). Because sites of integration can be key determinants in pathogenesis or untoward effects of viral vectors, there remains great motivation to understand the basis of retroelement integration site selection (Thomas et al. 2003; Kotterman et al. 2015; Küry et al. 2018).

One of the earliest examples of retroelement insertion specificity was the eponymous Ty3 element of the *Ty3/Gypsy* LTR retrotransposon class of *Saccharomyces cerevisiae* (Hansen et al. 1988; Sandmeyer et al. 2015; Sultana et al. 2017). Ty3 insertion occurs at the narrowly defined, nucleosome-free transcription start sites (TSS) of genes transcribed by RNA polymerase III (RNAP3), including 275 tDNAs, *SNR6* (U6 RNA), and 5S RNA (*RDN5*) genes (Qi et al. 2012). Insertion patterns similar to that of the Ty3 element have also been described in other yeasts (Dujon et al. 2004; Casaregola and Barth 2013; Guo et al. 2015; Magnan et al. 2016) and *Dictyostelium* (Winckler et al. 2005).

The Ty3 life cycle and genome organization are similar to those of simple retroviruses such as Moloney murine leukemia virus (MoMLV) (Sandmeyer et al. 2015). Ty3 expression is induced in haploid mating cells. Polyprotein Gag3 structural and Gag3-Pol3 catalytic protein precursors together with RNA assemble into virus-like particles (VLPs), thereby triggering proteolytic maturation of Ty3 polyproteins into Gag3 structural proteins capsid (CA) and nucleocapsid (NC), as well as Gag3-Pol3 catalytic proteins protease (PR), reverse transcriptase (RT), and integrase (IN). Based on retroviral models, the Ty3 intasome, a multimer of IN in association with the cDNA ends and potentially other proteins, enters the nucleus. Ectopically expressed Ty3 IN can mediate localization of itself and fused heterologous protein domains to the nucleus (Lin et al. 2001) and is required for cDNA processing and strand transfer (Kirchner and Sandmeyer 1996).

Investigations in vivo and in vitro showed that Ty3 integration is dependent on intact promoter elements of tDNAs, implicating transcription factors in targeting (Chalker and Sandmeyer 1992, 1993; Kirchner et al. 1995). Transcription of tDNAs by RNAP3 is mediated by general transcription factors TFIIIC and TFIIIB. TFIIIC associates with internal promoter elements, A- and B-boxes, and directs sequence-independent binding of transcription initiation factor TFIIIB upstream of the TSS (Kassavetis et al. 1990). In vivo genetic evidence and in vitro pulldowns indicated that IN makes direct contact with TFIIIC (Aye et al. 2001). Nevertheless, in vitro Brf1 and Spt15 (also known as Tbp or Tbp1) components of TFIIIB are sufficient for targeting and strand transfer mediated by recombinant IN (Qi and Sandmeyer 2012), implicating these proteins as key IN-interacting partners and potential determinants in insertion site selection. A previous in

Corresponding author: sbsandme@uci.edu

Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.240861.118. Freely available online through the *Genome Research* Open Access option.

© 2019 Patterson et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

vivo mapping study of Ty3 insertions showed a seemingly disparate usage of target genes with identical promoter elements. However, the observed relative usage of target sites was not accurately quantifiable because targeting frequency was determined based on the number of sequencing reads rather than counts of independent insertion events at a given site (Qi et al. 2012). This apparent difference in target usage was particularly interesting because in yeast, RNAP3-transcribed genes are mostly occupied by TFIIIB, TFIIIC, and RNAP3 (Harismendy et al. 2003; Roberts et al. 2003; Moqtaderi and Struhl 2004). This study investigated the apparent differences in usage of similar targets and the features that might account for those differences. Ty3 transposition frequency was quantified genome-wide using a random 8-nt "barcode" (8N)-tagging strategy (Chatterjee et al. 2014) to count multiple independent events at a single position. Independent events within tDNA isoacceptor families vary widely in frequency. We investigated binding of Brf1, RNAP3 subunit Rpc34, and IN and explored the structural properties of upstream sequences to determine the association with transposition frequency. Using insertion sites weighted for frequency of use, a nonpalindromic consensus integration sequence was defined.

Results

Analysis of de novo Ty3 insertion frequency

Retroviruses and retrotransposons integrate at preferred sites (Sultana et al. 2017). However, highly preferred sites are not readily differentiated by counting standard high-throughput sequencing reads because the number of reads at a site cannot distinguish between amplification of individual events and multiple independent events. Amplification bias could arise from clonal expansion of cells with specific insertions (as would not occur in the case for example of ChIP-seq) or could occur through PCR amplification as occurs in standard production of sequencing libraries. Ty3 has the strongest insertion site preference of known retrotransposons, making it critical to distinguish independent events from clonal amplification phenomena. To evaluate features that direct Ty3 insertion, relative use of Ty3 integration sites was quantified using a set of galactose-inducible Ty3 elements with a random 8N barcode inserted into the U5 region of the LTR just upstream of the internal domain (pGAL-Ty3-8N). After induction of transposition and allowing time for integration to occur, chromosomal DNA was extracted, and Ty3-genome joints containing the unique barcodes were preferentially amplified and subjected to Illumina sequencing. Genomic sequence at the integration joint with Ty3 sequence was aligned to the *S. cerevisiae* reference sequence (Supplemental Methods). Sequencing reads with unique barcodes mark independent insertions at a given site and the 5-bp target site duplication (TSD) generated at the ends of the integrated element can be deduced (Methods; Supplemental Fig. S1; Supplemental Table S1). Each of three independent experiments produced between 45 and 85 × 10^6 paired-end reads of which ~57%–78% passed filters and were included in downstream analysis. This analysis of 8N barcodes identified 8401, 11,823, and 12,152 unique insertion events in the three respective experiments. Insertions were normalized based on the total number detected per experiment so that experimental results could be compared for the three experiments. Ty3 inserted at virtually all of the 272 unique RNAP3-transcribed genes as well as the 5S rRNA gene array where only flanking *RDN5* orphan genes can be uniquely mapped. Subsequent analysis focused on insertions at tDNAs rather than all RNAP3 genes. These patterns were highly reproducible between experiments such that pairwise comparisons of experimental results correlated with R^2 values from 0.94 to 0.96. Unique insertions within a 50-bp region upstream of each tDNA were clustered for comparison of tDNA targets (Supplemental Table S2). Normalized insertion cluster counts for 267 tDNAs ranged from 2.75 ± 0.079 at tE(UUC)P to 112.46 ± 8.69 at *CDC65* (Systematic name: tQ(CUG)M). The vast majority of insertions were positioned upstream of tDNAs (encoding mature tRNA sequence) such that the 5-bp Ty3 TSD tDNA proximal ends were typically within 20 bp of the 5′ end of tDNAs (Fig. 1A,B; Supplemental Table S2). Based on thresholds of greater or less than 1.5-fold of the mean use of all tDNA targets, tDNAs were classified as hot (43), average (148), or cold (76). Insertion cluster counts varied among individual tDNAs of different isoacceptors as well as within the same isoacceptor family where they share virtually identical coding sequences (Supplemental Fig. S3). The glutamine (Q) and valine (V) tDNA families contained some of the most extreme ranges of transposition frequency, differing within families by 10.6-fold and 6.0-fold, respectively (Supplemental Table S3). Circos mapping (Krzywinski et al. 2009) of hot, average, and cold insertion clusters showed that these clusters as well as the 42 endogenous Ty3 LTRs were localized to RNAP3 sites throughout the genome (Fig. 1C).

Comparison of Ty3 insertion frequency with Brf1, RNAP3, and IN localization

RNAP3-transcribed genes are mostly occupied by TFIIIB, TFIIIC, and RNAP3 in vivo (Harismendy et al. 2003; Roberts et al. 2003; Moqtaderi and Struhl 2004). In vitro experiments have further shown that components of TFIIIB are essential for Ty3 integration (Qi et al. 2012). We directly tested the extent of correspondence of TFIIIB and RNAP3 complexes at tDNAs with Ty3 targeting frequency. TFIIIB and RNAP3 were localized by ChIP, and their relative association with tDNAs was expressed as fold enrichment (FE) using MACS2 (Methods; Supplemental Methods; Zhang et al. 2008; Feng et al. 2012). ChIP was performed using Brf1 and RNAP3 subunit Rpc34 fused at the amino terminal end to a triple FLAG tag (N-FLAG) expressed from their native genomic context in an exact replacement of the wild-type genes except for addition of the tag and a single *LoxP* site. Recombinant IN mediates strand transfer in vitro in the presence of a recombinant fusion of Brf1 and Spt15 (Qi and Sandmeyer 2012). IN contains nuclear localization signals sufficient for nuclear entry of itself together with heterologous fusion domains (Lin et al. 2001; Sandmeyer et al. 2015). We expressed triple-FLAG-tagged ectopic Ty3 IN using an estradiol-inducible expression system (Supplemental Methods) in the absence of Ty3 cDNA and other Ty3 proteins to determine whether IN is capable of independent association with targets, and if so, how this association relates to targeting frequency.

Analysis showed that Brf1 and Rpc34 were enriched surrounding insertion sites (Pearson correlation coefficient over a 100-bp window surrounding the peak positions, $r = 0.86$) (Fig. 2A,B; Supplemental Fig. S2A,B; Supplemental Table S4). Consistent with known functions, Brf1 and Rpc34 occupancy was offset upstream and downstream, respectively, from the tDNA 5′ end (Fig. 2A,B). IN also mapped to tDNAs, but with a broader distribution than that of Brf1 or Rpc34. Control untagged strains [BY4741 and BY4741 + empty plasmid (pKP3915)] showed no significant enrichment of Brf1, Rpc34, or IN at tDNAs.

A small subset of tDNAs did not exhibit occupancy with our test proteins and therefore warranted closer inspection.

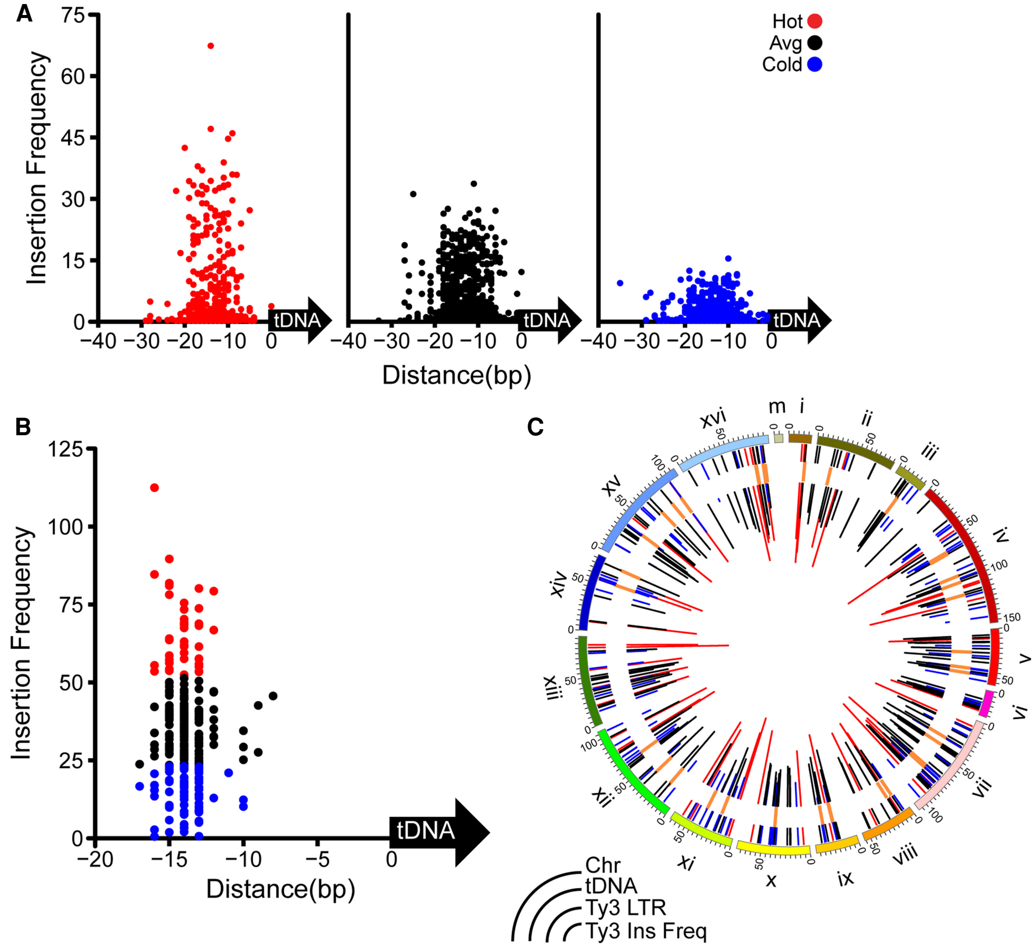

Figure 1. Genome-wide Ty3 insertion mapping using barcode-tagged elements. (*A*) Ty3 insertion sites are plotted with aligned mature tRNA-coding sequence (tDNA) = 0th bp and insertion site defined as the tDNA proximal end of the 5-bp TSD. Each dot represents the normalized unique barcodes of insertions starting at the strand transfer site proximal to the tDNA per 10,000 hits: (red) hot; (black) average; (blue) cold. Note that classification into hot, average, and cold was based on subsequently aggregating insertions per a 50-bp upstream window of the tDNA. (*B*) Insertions from *A* are binned across a 50-bp window upstream of the tDNA (colored as in *A*). (*C*) Circos tracks display chromosomes, tDNAs, Ty3 LTRs, and Ty3 insertion frequency: (*outside* to *inside*) colored as in *A*; (Ty3 LTRs) peach.

tD(GUC)N was previously found to be occupied by TFIIIC subunit Tfc1 (Harismendy et al. 2003), but not occupied by the RNAP3 factor Brf1 in that study or the current study, or by Ty3 IN in the current study. Examination of the sequence containing tD(GUC)N showed that a Ty1 insertion (YNLWTy1-2) occurred within the gene, thereby effectively removing the A-box promoter element. tD(GUC)N sustained few insertions (0.67 ± 0.352) (Supplemental Tables S3, S5, S6). The likely failure of TFIIIB to bind to the disrupted promoter of tD(GUC)N in our study and that of Harismendy et al. (2003) was consistent with the deficiency in transposition. The persistence of a strain that appears unlikely to express tD(GUC)N can be explained by the existence of four additional tD(GUC) genes. These tDNAs—tD(GUC)B, tD(GUC)D, tD(GUC)J2, and tD(GUC)J3—occur immediately downstream from tR(UCU) tDNAs—tR(UCU)B, tR(UCU)D, tR(UCU)J1, and tR(UCU)J2, respectively. Previous in vivo and in vitro evidence suggests that these tDNAs are expressed as dimeric transcripts (Kjellin-Straby et al. 1984; Engelke et al. 1985; Hottinger-Werlen et al. 1985). In each case, the upstream tR(UCU) gene showed a far greater number of insertions than the downstream tD(GUC) gene, and insertions at the tD(GUC) gene could not be specifically assigned. Brf1, Rpc34, and IN were enriched as expected at these dimeric sites. However, mapping had insufficient resolution to distinguish individual members of the pairs. Because only the tR(UCU) genes sustained substantial insertions, Brf1, Rpc34, and IN most likely predominantly bound at the upstream gene. All four tD(GUC) genes were excluded from further analysis of features differentiating hot and cold Ty3 targets (Supplemental Table S6). Four extended, enriched regions spanning positions 0 to +500 were observed in ChIP-seq of Brf1, Rpc34, and IN (Fig. 2A). Closer examination of these regions showed that the first base positions of tDNAs tK(UUU)G2 and tL(GAG)G were only 372 bp apart, and the first base positions of tDNAs tI(AAU)B and tG(GCC)B were only 204 bp apart. In these cases the pair members could be distinguished and led to an extended ChIP-seq pattern (Fig. 2A). Hot targets bound more Brf1, Rpc34, and IN (FE in the order Brf1 > Rpc34 > IN) (Fig. 2A, upper panel) than all other tDNAs (*P*-values = 1.25×10^{-6}, 2.56×10^{-5}, and 4.43×10^{-4}, respectively). However, cold targets averaged lower occupancy and were not as significantly different from other targets (*P*-values = 6.1×10^{-3}, 5.8×10^{-2}, and 1.3×10^{-2}, respectively). On average the FE peaks of Brf1 and IN were displaced upstream of the tDNA relative to the peak of Rpc34 (Fig. 2B).

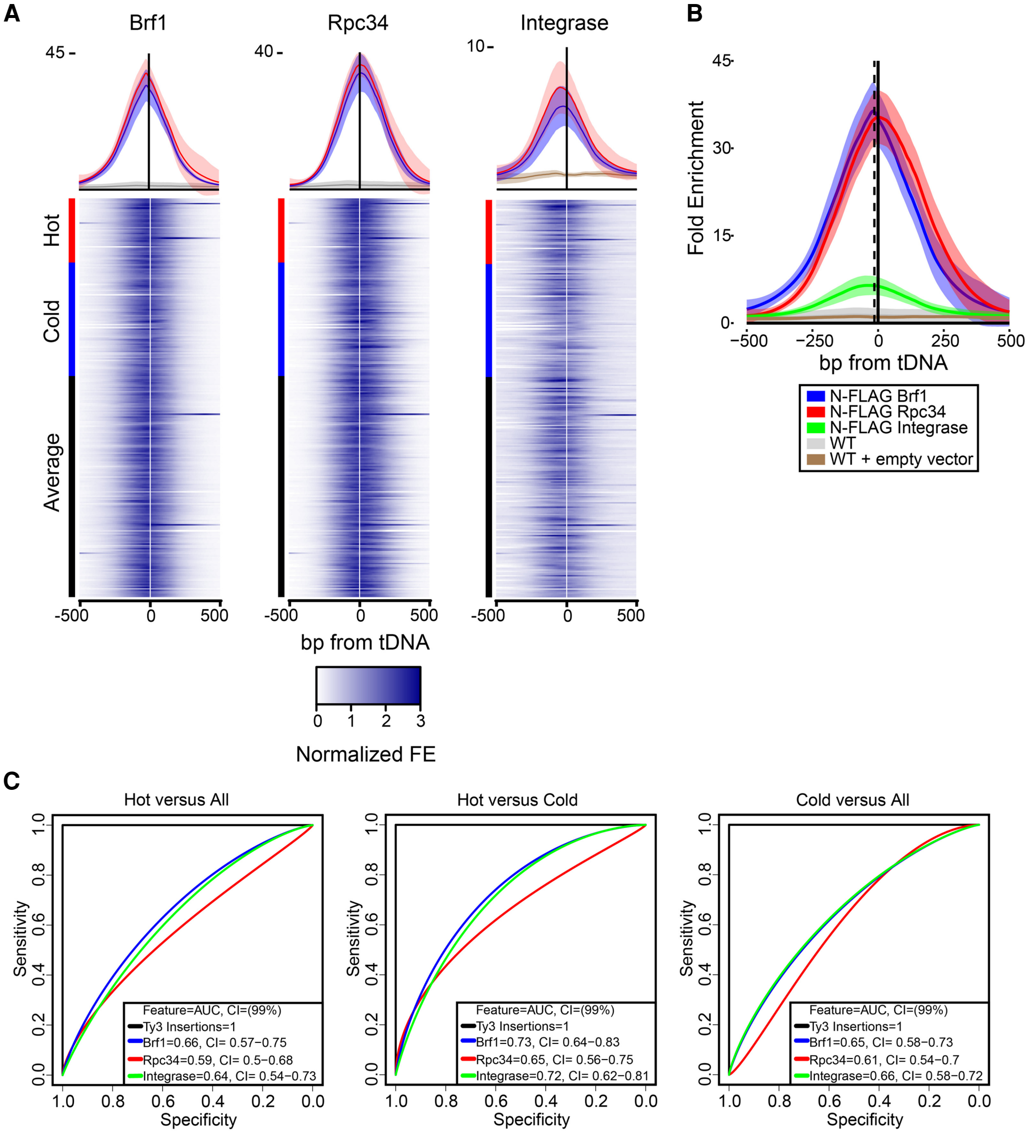

Figure 2. Basal transcription factors, RNAP3, and IN at Ty3 targets. (*A*) ChIP-seq analysis of N-FLAG-tagged Brf1, Rpc34, and IN. Heatmaps showing normalized fold enrichment (FE) for RNAP3-transcribed genes relative to flanking sequence; rows ordered according to Ty3 transposition frequency *top* to *bottom*, high to low, as indicated by the colored bar. Position 0 refers to the nucleotide encoding the 5′ end of the mature tRNA. *Above* the heatmaps is an expanded view of FE for Brf1, Rpc34, and IN for hot and cold genes compared to untagged Brf1 and Rpc34 strains and empty vector strain for IN with standard deviation represented in lower intensity. (*B*) Peak analysis averaged for all tDNAs and traced over a 1-kb surrounding window. Lightened haze around each line indicates standard deviation. WT refers to the untagged parent strain of N-FLAG Brf1 and N-FLAG Rpc34; WT + empty vector (pKP3915) control for N-FLAG IN (WT + pKP4010). The dotted vertical line represents the average Ty3 insertion site position. (*C*) ROC analysis comparing hot versus all other tDNA, hot versus cold, and cold versus all other tDNAs. Legends for each plot show the AUC with 99% confidence intervals (CI). As a positive control for each plot, Ty3 insertion frequency (used to define hot and cold) is included to show perfect association between phenotype classifier and FE predictors (black line).

We further applied Receiver Operator Characteristic (ROC) (Berry et al. 2006; Roth et al. 2011) to assess how well FE predicted hot or cold classification. An area under the curve (AUC) with 99% confidence intervals (CI) showed that FE of Brf1, IN, and Rpc34 had significant predictive power for classification of hot Ty3 targets (Fig. 2C). Equivalent binding of factors would have provided an AUC value of 0.5 for each protein for hot versus cold and all other tDNAs. However, when target sites were classified as either "hot" or "all other tDNAs," AUC values of 0.66, 0.64, and 0.59 were obtained for Brf1, IN, and Rpc34, respectively, and similarly for "cold" versus "all other tDNAs," of 0.65, 0.66, and 0.61, were obtained, respectively. When only hot and cold phenotype groups were compared, ROC analysis for Brf1, IN, and Rpc34 produced AUC values of 0.73, 0.72, and 0.65, respectively. Overall then, hot and cold targets were better predicted by Brf1 and IN than by the RNAP3 subunit Rpc34.

Local DNA sequences confer Ty3 targeting frequency bias

Although this is the first Ty study in which a barcoding strategy rigorously distinguished independent events, previous studies of both Ty1 and Ty3 (Ji et al. 1993; Bachman et al. 2004; Qi et al.

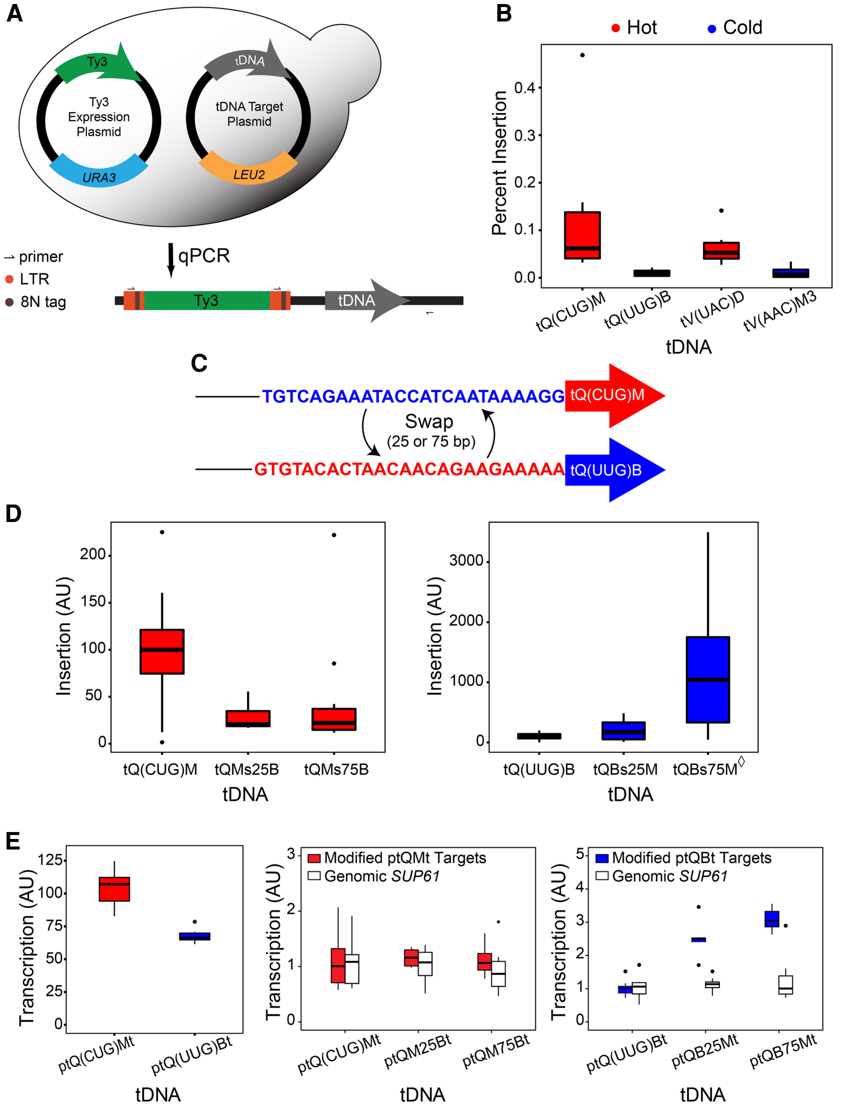

Figure 3. Ty3 insertion frequency into high-copy-number target plasmids. (*A*) Diagram of tDNA target assay. Measurement of insertion frequency of Ty3 into tDNA hot and cold target plasmids by qPCR described in Methods. (*B*) Frequency of insertion (%) into tDNA target plasmids. Biological replicates $N = 4–6$. (*C*) Diagram of hot tQ(CUG)M and cold tQ(UUG)B target tDNAs and sequence swaps. (*D*) Sequence upstream of tDNA mature coding region influences Ty3 insertion frequency. Measurement of insertion frequency into hot and cold tDNA targets with and without sequence swaps described in *C*. Insertion frequency expressed as arbitrary units (AU) normalized to native tDNA targets. Biological replicates $N = 10–12$. The diamond (◊) indicates a data point (14,124) excluded from the box plot. (*E*) Transcription levels from tDNA target plasmids. Box plots show transcription from tQM and tQB (*left*), or tQMt target plasmids with 25 or 75 bp of upstream sequence from tQBt (*middle*) and vice versa (*right*). White boxes represent transcription from genomic *SUP61*, an intron containing tRNA where nascent pre-tRNA was measured as a proxy for transcription. Values expressed as normalized arbitrary units (AU).

2012) inferred preferential use of targets by relying on data sets from individual loci or high-throughput sequencing lacking barcoding quantitation (Cheung et al. 2018). Given the clearly differing behavior of similar targets, we sought to delimit the region that confers the characteristic targeting property. Individual tDNAs (mature tRNA-coding sequence) plus 120 bp of upstream flanking sequence were subcloned into yeast shuttle vectors to create target plasmids (Methods). Briefly, cells were transformed with target plasmid and donor plasmids carrying an inducible Ty3 (Fig. 3A). After a 24-h induction of Ty3 expression, genomic DNA was harvested, and the percentage of Ty3 insertion events into the plasmid out of total was determined by qPCR (Methods; Supplemental Methods). Comparisons of the mean insertion percentages observed for plasmids containing hot targets [tQ(CUG)M, tV(UAC)D] over that of cold [tQ(UUG)B, tV(AAC)M3] targets was 136.4-fold and 2.89-fold, respectively. Although the absolute insertion percentages differed, use of hot and cold targets in the plasmid target assay mirrored the trends observed for chromosomal insertion events (Fig. 3B).

To further delimit the region to which differences might be attributed, a pair of hot (tQ(CUG)M) and cold (tQ(UUG)B) isoacceptor tDNAs of the glutamine tDNA family were used as targets in the plasmid assay. These contain identical tDNAs, but differ in flanking sequence so that the flanking 75-bp sequence is only 21.3% identical. The contribution of the 75 or 25 bp immediately upstream of the tDNA to differences in targeted insertion was assessed by swapping the designated regions (−1 to −26 or −1 to −76) and measuring insertion frequency into the different targets. Insertion frequency was expressed in arbitrary units (AU) relative to the original sequence as 100% (Fig. 3C,D). Transposition frequency was reduced by ∼3.5-fold and approximately twofold when 25 bp (ptQM25B) and 75 bp (ptQM75B) of upstream sequence from cold target ptQB was swapped into the ptQM context (Wilcoxon rank-sum test $P = 1.04 \times 10^{-2}$ and 4.00×10^{-2}, respectively). In the converse experiment, when 25 bp of upstream sequence from hot target tQM was swapped into the cold tQB, target frequency was elevated by approximately twofold (ptQB25M) ($P = 0.291$). A swap of 75 bp of upstream sequence from tQM into the cold target increased transposition by ∼23-fold (ptQB75M) ($P = 1.44 \times 10^{-4}$) (Fig. 3D).

These results argue that the immediate upstream region of tDNA is a significant contributor to the differential observed within tDNA families that have common promoter elements. To further examine the nature of this contribution, the region spanning the upstream and tDNA of either hot or cold gene targets was analyzed using MEME Suite (Bailey et al. 2009). Analysis readily identified the highly conserved internal A- and B-box promoter element motifs and, with the exception of a bias for R = G in the A-box promoter element position 8, hot and cold genes were essentially indistinguishable (Supplemental Fig. S3A). For example, tQ and tY families, which showed roughly five- to 10-fold differences in Ty3 insertion frequency across members (Supplemental Fig. S3B), exhibited relatively little intra-family difference in overall sequence, including within the A-box sequences where both families encoded G in position 8 of the A-box promoter element (Supplemental Fig. S3C).

The differential occupancy of Brf1 and RNAP3 subunit Rpc34, albeit modest, between hot and cold targets suggested that targeting has parallels with transcription. However, earlier in vitro experiments suggested that RNAP3 competes with the integration complex (Connolly and Sandmeyer 1997). To attempt to directly address the relationship between transcription and integration in vivo, hot and cold plasmid tDNA targets were modified by inserting a unique primer binding site just upstream of the terminator sequence to enable measurement of plasmid-specific pre-tRNA levels via qPCR (Methods; Supplemental Methods). Assuming this tail is processed as an early step in maturation, the level of pre-tRNA detected would reflect newly made transcripts. However, consistent differences were not observed between hot and cold tDNA pairs with respect to this reporter for transcriptional activity (Fig. 3E).

Ty3 strand transfer occurs at the site of RNAP3 initiation and is consistent with the retrovirus central YR rule

The mechanisms underlying upstream sequence effects on Ty3 targeting motivated closer examination of the region immediately upstream of the tDNA to identify potential sequence motifs associated with either hot or cold Ty3 targets. Retroviral and LTR retrotransposon IN proteins are members of the D, D (X_{35}), E superfamily of polyesterases and share a conserved core domain. The intasome contains multimeric retroviral IN proteins in complex with cDNA (Engelman and Cherepanov 2014; Ballandras-Colas et al. 2016). IN introduces a strand transfer of cDNA ends to host sequence staggered across the helix by 4–6 bp, thereby generating, after repair of the two gaps, the characteristic TSD. Recent structural analysis of the retroviral intasome has highlighted a structural role for a flexible YR (pyrimidine/purine) dinucleotide in enabling IN active site target access (Hare et al. 2010; Maertens et al. 2010; Serrao et al. 2014). To determine characteristic features of the Ty3 insertion site, our analysis focused specifically on the 11-bp window containing the Ty3 5-bp TSD (positions 0–4) plus 3 bp of flanking sequence on each side. The tRNAs are synthesized with 10–12 nt of 5′ precursor sequence that is trimmed during maturation, but because of the similarity of tRNA isoacceptor sequences, the TSS for specific loci are not well defined. The majority of Ty3 insertions position the tDNA proximal end of the TSD within 20 bp upstream of the mature tRNA-coding sequence and therefore presumably close to the TSS (Fig. 1B; Supplemental Table S3). After excluding 36 tDNAs that have preexisting Ty3 insertions, a total of 236 tDNA insertion site sequences were analyzed. Target sequences aligned to the TSD on the non-tRNA template strand were weighted by transposition frequency and analyzed with WebLogo (Crooks et al. 2004). In all sequences, C and A were overrepresented at positions 4 and 6, respectively, and G was depleted throughout (Fig. 4A). Cold targets were additionally enriched for T/A at nucleotide positions 0–3.

To assess the potential role of flexible dinucleotides in Ty3 integration, dinucleotide analysis of the 11-bp window surrounding and including Ty3 insertion sites was also performed (Fig. 4B). The 11-bp window representing hot sites showed peaks of the flexible YR dinucleotide in bins −2, 1, and 5. In contrast to hot targets, cold targets showed only a strong YR peak at dinucleotide bin 4. Hot and cold targets differed with respect to the relatively inflexible RR dinucleotide, with hot target peaks at −2 and 2 and cold targets showing a gradient from high upstream to a low point at bin 4.

The context of Ty3 insertion sites is unique because of its dual roles as a target of Ty3 transposition as well as a "target" for transcription initiation. Previous studies of tDNA transcription showed that the preferred start site is at "A" centered at approximately position −13 upstream of the mature tRNA-coding sequence (Giuliodori et al. 2003; Yukawa et al. 2011). A common motif "TCAACA" spanning the TSS on the nontemplate strand was found in 31 genes studied by in vitro analysis where the first "A" is the most common initiating base (Yukawa et al. 2011). To obtain an unbiased candidate motif, we used MEME Suite (Bailey et al. 2009) and searched for 6-bp motifs within a 23-bp window directly upstream of mature tRNA-coding sequences. The strongest motif was ttCAan (Fig. 4C), which overlapped the previously identified TSS, "TCAACA" (Giuliodori et al. 2003; Yukawa et al. 2011). The MEME-predicted ttCAan motif also resembles the sequence contained in the 11-bp window when examining TSDs of all targets (Fig. 4A, left), specifically nucleotide positions 2–7. This underscores the likelihood that universal YR bias at the tDNA proximal end of the Ty3 TSD is coincident with the TSS (Kassavetis et al. 2001; Grove et al. 2002). This would position the Ty3 TSD YR dinucleotide within the transcription initiation DNA open complex (Kassavetis et al. 2001, 2003). Although relative nucleotide frequencies at the "CAA" positions between the two WebLogo analyses display clear differences (Fig. 4A versus Fig. 4C), we note that in Figure 4A, the MEME motif was weighted by the total number of Ty3 insertions at each tDNA target, whereas in Figure 4C the tDNA MEME motif WebLogo was not weighted.

We next investigated the relationship of the Ty3 insertion site to RNAP3 TSS. Comparison of Ty3 insertion sites in this study to the tDNA TSS determined in the previous study (Yukawa et al. 2011) showed that 25 hotspots for independent Ty3 insertions detected by random barcoding were within 1 nt of the dominant TSS identified for 29 tDNAs (Fig. 4D). To test whether the discovered MEME Suite motif was a reliable TSS predictor for tDNAs, we measured the distance between the strongest TSS of the 29 tDNAs reported by Yukawa et al. (2011) and the MEME-predicted motif upstream of the same tDNAs (Fig. 4D). These distances were measured between the first "C" nucleotide in both motifs, that is, ttCAan in MEME and TCAACA in Yukawa et al. (2011). At 20 of 29 tDNAs, MEME motifs overlapped with empirical motifs exactly at the "C" position (bp difference of 0), and an additional two tDNAs were within 2 bp of a perfect overlap (Fig. 4D). MEME-discovered motifs were therefore deemed as a suitable TSS predictor upstream of tDNAs, and the distances of the Ty3 TSD to these predicted TSS motifs were calculated to determine whether Ty3 insertion density around the TSS region of hot, average, and cold tDNAs displayed any marked differences (Fig. 4D). These distances were measured from the fifth position of the TSD to the aforementioned "C" nucleotide of the predicted TSS (position 3 in 6-bp TSS motif). Mean differences for hot, average, and cold tDNAs were 0.46 bp, −0.82 bp, and −1.41 bp, respectively, underscoring the proximity of tDNA proximal Ty3 strand transfer sites to the TSS and further suggesting that Ty3 insertions at or upstream of the TSS "A" might be favored more strongly than insertions downstream from this base.

DNA curvature corresponds to low-frequency targets

Transposable element (TE) insertion sites that contain nucleotide sequences that are flexible can facilitate TE integration (Repanas

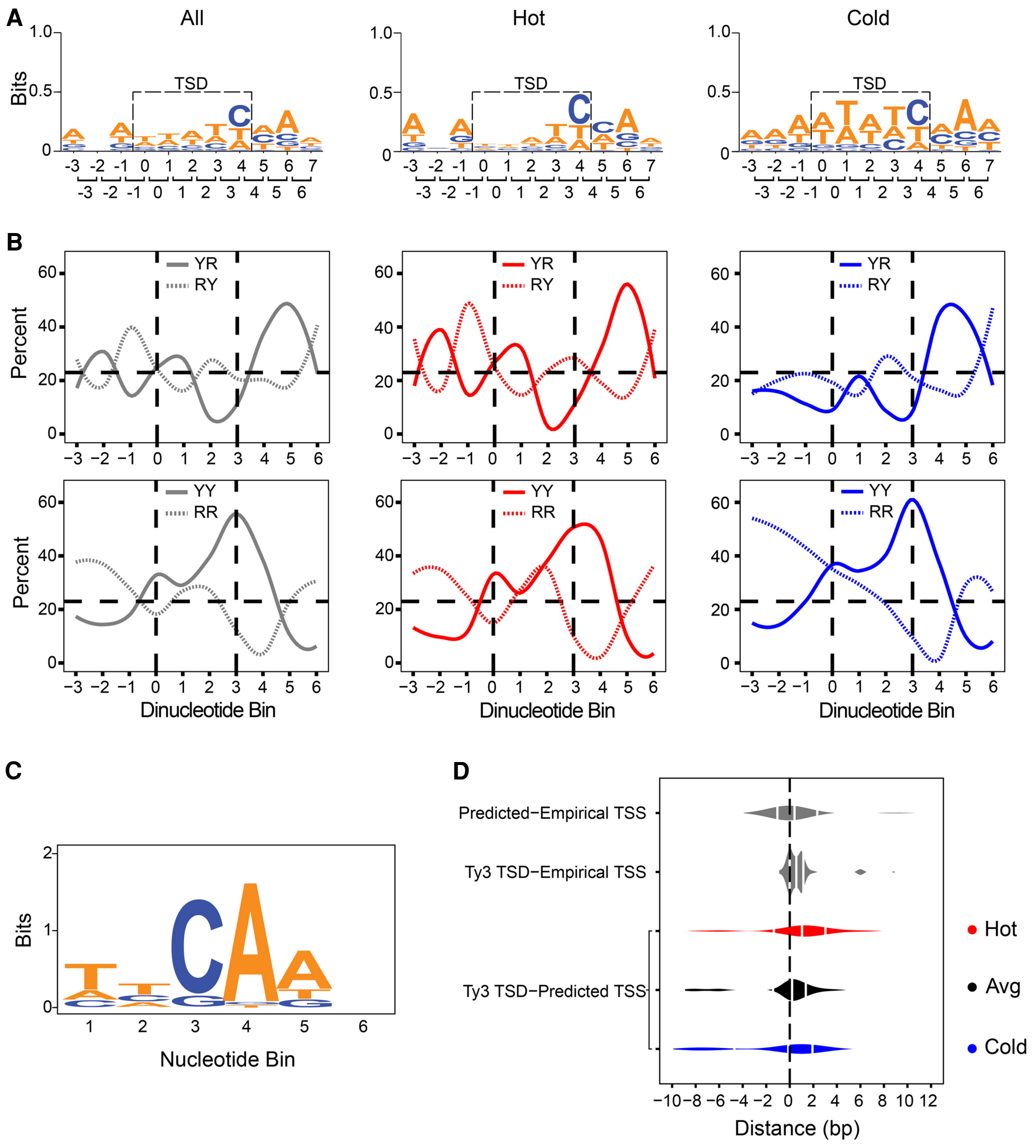

Figure 4. Ty3 insertion site analysis upstream of tDNA genes. (*A*) WebLogo analysis of the 11 bp comprised of Ty3 5-bp target site duplication (TSD) and ±3 bp flanking sequence. Each TSD was weighted by the total number of Ty3 insertions at that site. Brackets indicate the corresponding nucleotide positions (*top* row of numbers) assigned to each dinucleotide bin (*bottom* row of numbers). (*B*) Plots of dinucleotide frequency determined from sequences shown in *A*. Dinucleotide starts at position indicated ("0" = dinucleotide at positions 0 and 1 of TSD, etc.). YR/RY (*top*) and YY/RR (*bottom*) plots of TSD and flanking sequence shown in *A*. Vertical dashed lines mark the dinucleotide bins representing the borders of the TSD. Horizontal dashed line represents the random frequency of the YR dinucleotide in the *S. cerevisiae* genome (23.25%). Random frequency of all dinucleotide bins in the *S. cerevisiae* genome are 23.25% (YR), 23.25% (RY), 26.71% (YY), and 26.79% (RR). (*C*) WebLogo analysis of conserved motifs within a 23-bp window upstream of all tDNA genes by MEME Suite. All four DNA nucleotides occur at roughly the same frequency at position 6. (*D*) Distance analysis of Ty3 TSD to MEME-predicted and empirically determined TSS upstream of tDNAs. From *top* to *bottom*: distance between MEME-predicted TSS and TSS of 29 tDNAs empirically determined by Yukawa et al. (2011); distance between Ty3 TSD and empirical TSS; distance between Ty3 TSD and MEME-predicted TSS of all tDNAs in this study categorized by hot, average, and cold phenotypes. For all comparisons, distance is measured from the fifth base of the TSD to the first conserved "C" nucleotide in both MEME-predicted motifs and empirically determined motifs (see text for detailed explanation). The first, second, and third quartiles of each data set are denoted by white lines on each violin plot.

et al. 2007; Maertens et al. 2010; Voigt et al. 2016). In the DNA elastic rod model, DNA bendability is determined by local sequence contribution to predicted flexibility of the DNA helix, whereas curvature is determined over a longer sequence and is associated with spacing of A-rich tracks that collectively result in a convex surface (Trifonov and Sussman 1980; Munteanu et al. 1998; Nov Klaiman et al. 2009; Bansal et al. 2014). Curvature and bendability were estimated upstream of all tDNAs from position −130 to +30 relative to position 1 of the mature tRNA-coding sequence using the bend.it tool (Fig. 5A; Vlahovicek et al. 2003). DNA curvature within the −26 to −1 region averaged 2.22°/10.5 bp turn less at all hot targets compared to cold ($P = 6.8 \times 10^{-7}$). Bendability was not found to be significantly correlated with transposition (Supplemental Fig. S4). Pearson correlation analysis did, however, show a significant negative relationship between DNA curvature and Ty3 insertion ($r = -0.30$, P-value = 3.0×10^{-6}) (Fig. 5B). This prediction is consistent with the enrichment of RR dinucleotides across the upstream flanking and TSD sequences of cold tDNA targets.

Curvature analysis of upstream sequences in specific hot and cold plasmid targets showed a similar pattern. Hot target ptQM had less DNA curvature than cold target ptQB. Swapping hot/cold upstream sequences in the ptQ plasmid targets that dramatically changed insertion frequency (Fig. 3C,D) also influenced curvature (Fig. 5C,D). These results suggest that increasing curvature antagonizes integration in the TSS context. To the extent that curvature is responsible for modulating targeting, it might also influence the stability of the nucleosome directly upstream of the tDNA and therefore the ability of such nucleosomes to destabilize TFIIIB binding, therefore indirectly but significantly affecting targeting.

Discussion

The barcode labeling of the Ty3 yeast retrotransposon enabled us to show that despite exquisite specificity for highly similar gene targets, the frequency of insertions varies over a broad range together with the presence of Brf1, RNAP3 subunit Rpc34, and IN. In specific test cases, we found that immediate upstream sequence composition showed differences between hot and cold genes in dinucleotide composition and DNA curvature.

In our study, we did not directly evaluate a role for nucleosome occupancy. However, because TFIIIB and nucleosomes do not co-occupy the sequence upstream of the tDNA TSS, the effect of nucleosomes competing with TFIIIB or the effect of curvature could have been reflected in the TFIIIB occupancy at target genes. Comparisons between nucleosome maps based on previous studies (Nagarajavel et al. 2013; Cole et al. 2014) and our studies of Brf1 and integrase or integration frequency were not possible owing to differences in culture conditions.

We also did not evaluate a role for nuclear localization in targeting. In *S. cerevisiae*, nuclear localization of tDNAs varies among genes and is also dependent on transcriptional activity across the cell cycle (Mavrich et al. 2008; Duan et al. 2010; Brogaard et al. 2012; Kumar and Bhargava 2013). In the case of yeast Ty1, which targets the tDNA proximal upstream nucleosome rather than the TSS (Baller et al. 2012), disruption of the nuclear pore is associated with increased targeting to chromosomal ends relative to tDNAs (Manhas et al. 2018). Recent work from the Engelman laboratory shows that CPSF6 bound to CA essentially acts to shield chromatin at the nuclear periphery from HIV-1 integration so that intasome activity is distributed throughout the nucleus (Achuthan et al. 2018; Engelman and Singh 2018).

Intasome and target site selection

Our previous studies showed that IN is necessary and sufficient for nuclear localization in vivo (Lin et al. 2001), and in vitro recombinant IN is sufficient for position-specific integration into targets bound by a recombinant fusion of Brf1 and Spt15 (Qi and Sandmeyer 2012). In this study, we showed that ectopic IN expressed in the absence of other Ty3 components, localizes to tDNA targets.

The Ty3/Gypsy family found in plants and fungi includes Ty3-like and Chromoviridae retroelements that differ in the IN domain (Malik and Eickbush 1999). The latter class targets insertion into heterochromatin by interaction between the IN chromo domain and histone marks, H3K9me2 and H3K9me3 (Gao et al. 2008). *S. cerevisiae* lacks H3K9 methylation (Grunstein and Gasser 2013) and perhaps as a consequence, the carboxyl-terminal portion of the IN protein differs between the two groups of elements.

Retrovirus insertion preferences are more widely distributed than those of the Ty elements. Similar to the Chromoviridae, insertion bias has been extensively linked to interactions with epigenetic features. For example, MLV IN interacts with bromodomain BET proteins, BRD2, BRD3, and BRD4 associated with promoter regions (Lewinski et al. 2006; Felice et al. 2009; De Rijck et al. 2013; Aiyer et al. 2014). HIV IN interacts with LEDGF/p75, which in turn associates with H3K36-methylated nucleosomes (Cherepanov et al. 2003; Pradeepa et al. 2012; Eidahl et al. 2013). Structural studies of the prototype foamy virus (PFV) strand transfer complex (STC) showed that the bias for nucleosomal DNA reflects contributions from IN contacts with histones as well as the presence of flexible YR dinucleotides within the TSD (Maertens et al. 2010; Serrao et al. 2014; Maskell et al. 2015).

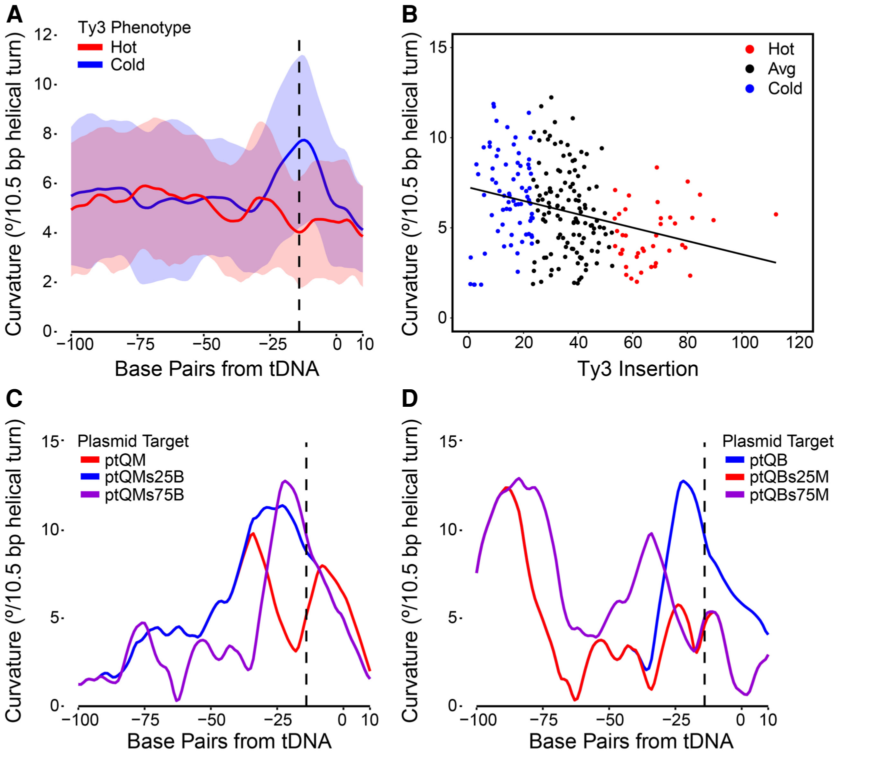

Figure 5. Role of curvature in target site determination. (*A*) Plot of curvature as determined by bend.it analysis (Vlahovicek et al. 2003) of 100 bp upstream of tDNA for hot (tQM) and cold (tQB) targets. (*B*) Linear regression of curvature versus insertion frequency. (*C*,*D*) Plots of curvature of hot and cold plasmid targets and swapped sequences as described in legend of Figure 3C.

Roles of DNA sequence and conformation in Ty3 targeting

Swapping upstream regions of 25–75 bp between hot tQM and cold tQB targets significantly affected insertion frequency. This observation focused our attention on the context of strand transfer and flanking DNA. Examination of the Ty3 5-bp TSD consensus showed that it has a nonpalindromic YR enrichment at a position corresponding or proximal to the TSS CA of transcription initiation. The pattern of symmetric T/A enrichment outside of the TSD seen in retroviruses (Serrao et al. 2014) is not observed for Ty3. This is likely because our analysis oriented the insertions relative to the tDNA. Previous work from our laboratory indicated that Ty3 displays insertion bias that can only occur if the target is asymmetric (Aye et al. 2001). Recent studies from the Bangham laboratory showed that flanking T/A palindromic sequences are likely a feature of a population-based retroviral integration site consensus rather than actual individual insertion sites (Kirk et al. 2016). Ty3 appears to have adapted to target a sequence evolved for enhanced flexibility for transcription initiation. Our dinucleotide analysis showed that flexible YR dinucleotides identified at the strand transfer site also coincide with the RNAP3 TSS. If retrovirus TSD are similarly examined as individual events rather than composites of all events on both strands, a strong bias for the nonpalindromic sequence 5′-T(N1/2)[C(N0/1)T|(W1/2)C]CW-3′, in which square brackets represent the start and end of the TSD, W denotes A or T, and | represents the axis of symmetry, emerges at the TSD-genome junction (Kirk et al. 2016). The Ty3 TSD-genome junction we found at all targets is consistent with this nonpalindromic consensus (ANW [NNNWC]C/AA).

The precision of Ty3 integration allows additional speculation about the nature of the DNA distortion that occurs at the TSD. TFIIIB maps between −40 and −5 bp relative to the tDNA TSS (Kassavetis et al. 1989). The TFIIIB intermediate complex composed of Spt15 and Brf1 binds largely through Spt15 contacts, causing widening of the narrow groove of the DNA and introducing a sharp bend in the helix at this position (Juo et al. 2003). TFIIIB binding also specifically facilitates opening of the transcription bubble roughly in the TSS region −9 to +11, a function that can be substituted by synthetic single-stranded regions to compensate for Bdp1 deletion mutants (Kassavetis et al. 1999, 2003; Kassavetis and Geiduschek 2006). Recent cryo-EM modeling of transcribing RNAP3 (Hoffmann et al. 2015) combined with cross-linking studies of RNAP3 complexes (Wu et al. 2012; Male et al. 2015) show that the opening in the initiation complex bends the DNA centered at the TSS. Our analysis shows that Ty3 insertion frequency corresponds to the position of this bend, and that cold targets displayed substantially more DNA stiffness or curvature in regions upstream of the TSS (Fig. 5; Supplemental Fig. S4). Together these observations suggest that tDNA structure evolved to allow opening at the TSS by transcription intiation (Fig. 6A) and that structure has been hijacked by the Ty3 intasome to facilitate integration at a position that minimizes disruption to the host genome (Fig. 6B).

tDNA targeting by retrotransposons has evolved independently several times in eukaryotic cells with relatively compact genomes. In the case of Ty3 tDNA targeting, not only do insertions avoid disruption of upstream promoter elements, but the ends of the Ty3 element even mimic the RNAP3 TSS itself. Dissection of the nonpalindromic Ty3 TSD consensus in this work is consistent with the new view that retrovirus insertion sites are nonpalindromic. Altogether these findings also underscore the preservation of common target features of retrotransposition and the transcription open complex.

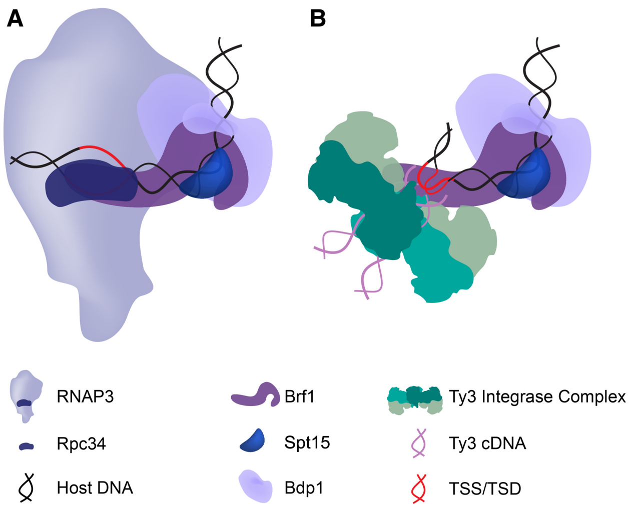

Figure 6. Modeled comparison of RNAP3 TSS and Ty3 integration TSD. (*A*) RNAP3 complex is recruited by the TFIIIB complex containing Brf1, Spt15, and Bdp1 that binds ~20–40 bases upstream of the tDNA. RNAP3 subunit Rpc34 is positioned near the active site of the TSS and transcription bubble (Wu et al. 2012). (*B*) Suggested model of Ty3 integration into target DNA at or near the TSS. Ty3 integration complex is recruited to target sites via interactions with Brf1. Ty3 cDNA integration may require bending of the host DNA into the integrase complex active site to facilitate strand transfer with Ty3 cDNA ends. Model suggests that flexibility of this region contributes to integration, whereas stiffness corresponding to curvature does not.

Methods

Strains and plasmids

The reference yeast strain BY4741 and its derivatives are described in Supplemental Table S1. Cells were cultured at 30°C (unless otherwise indicated) in medium as described in Supplemental Methods. Plasmids and primers used are listed in Supplemental Table S1. Manipulation of strains and plasmids used standard techniques. The construction of pGAL-Ty3-8N plasmid library, β-estradiol-inducible Ty3 IN plasmid pKP4010, and FLAG-tagged Brf1 and Rpc34 is described in Supplemental Methods. The transposition efficiency of the pGAL-Ty3-8N for a mixture of 800 isolates compared to the Ty3 lacking the barcode was 88%±10%. Thus, the efficiency of the tagged element was generally comparable to wild type.

Retro-seq

In three independent experiments, the pGAL-Ty3-8N library was transformed into *S. cerevisiae* strain BY4741. Approximately 15,000–17,000 transformants were pooled together and mixed vigorously for each replicate. Transformant pools were induced for Ty3 expression and plated on 5-FOA medium to select for cells that had lost the donor plasmid. Colonies were combined for gDNA isolation and sequencing library preparation. After sequencing, reads were mapped to the sacCer3 *S. cerevisiae* genome (Supplemental Methods) using Bowtie (Langmead et al. 2009). Scripts developed for cataloging unique Ty3 integration event are described in Supplemental Code.zip.

Chromatin immunoprecipitation

Chromatin Immunoprecipitation (ChIP) was performed as described elsewhere (Kuras and Struhl 1999; Boukaba et al. 2004; Zentner et al. 2013; Zentner and Henikoff 2013a,b; Kasinathan et al. 2014). Briefly, cells were grown as described above for transposition frequency sequencing, except without 5-FOA selection. Chromatin was harvested from cultures and FLAG-tagged Brf1, Rpc34, or IN was immunoprecipitated with Anti-FLAG M2 Magnetic Beads (MilliporeSigma M8823). Eluted DNA was prepared for DNA sequencing or used directly for ChIP-qPCR (Supplemental Methods).

DNA sequencing library preparation

Preparation of DNA sequencing libraries for Illumina sequencing was adapted from the Illumina Multiplex Sample Preparation Guide. Additional details for sequencing library preparations and data analysis are provided in Supplemental Methods.

Ty3 plasmid transposition assay

The pGal-Ty3-8N plasmid and target tDNA plasmid were cotransformed into BY4741. Transformed cells were induced for Ty3 expression for 24 h at 23°C. Plasmid DNA was extracted (GeneJET Plasmid Miniprep kit; Thermo Fisher Scientific) and assayed for transposition by qPCR (Supplemental Methods).

Transcription assay from plasmid-borne tDNAs

BY4741 cultures with both Ty3 expression plasmid and various modified tDNA plasmids (Supplemental Table S1) were grown as plasmid transposition experiments except that cells were induced for Ty3 expression for ∼12 h of SG induction (OD_{600} of ∼1.0); Extracted RNA (YeaStar RNA isolation kit; Zymo Research) was DNase I-treated (Thermo Fisher Scientific), and purified (RNA Clean and Concentrator kit; Zymo Research). cDNA was made (iScript Reverse Transcription Supermix; Bio-Rad Laboratories) and were used for qPCR as described (Supplemental Methods).

Quantitative PCR

All qPCR experiments were performed using the CFX96 C1000 Touch (Bio-Rad). Data analysis for qPCR experiments is described in Supplemental Methods and primers are described (Supplemental Table S1).

Data analysis

Figures were prepared as follows: All box plots, scatter plots, bar graphs, violin plots, and protein occupancy graphs were made using ggplot2 (Wickham 2016); heatmaps were made with gplot (https://cran.r-project.org/web/packages/gplots/gplots.pdf) and lattice (Sarkar 2008); pROC was used for ROC analysis (Robin et al. 2011), and ade4 for tri-plots (Dray and Dufour 2007). The Circos diagram was made with Circos software (Krzywinski et al. 2009). DNA curvature and bendability were measured using the bend.it tool (Vlahovicek et al. 2003) along with custom scripts for batch analysis (Supplemental Code). The bend.it tool was also applied to determine curvature on the A-phased template DNA described in Pasi et al. (2016). Unless otherwise noted, computer software generated in house for analysis was made in Perl (https://www.perl.org/).

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE97894.

Acknowledgments

This work was funded by Public Health Services grant GM33281 and the Grace Beekhuis Bell Chair of Biological Chemistry to S.B.S. We thank the University of California, Irvine, Genomic High-Throughput Facility supported by National Cancer Institute Cancer Center Support Grant (CA-62203, R. Van Etten, PI, University of California, Irvine) and National Institute of General Medical Sciences Shared Instrumentation grants (1S10RR025496-01 and 1S10OD010794-01 to S.B.S.). K.P. was supported by NIAID T32 AI–07319 (PI, B. Semler, University of California, Irvine). We thank J. Yu of the Sandmeyer laboratory for helpful discussions.

Author contributions: K.P., X.Q., and S.B.S. designed the study. K.P. conducted the Ty3 transposition; Retro-seq; Host factor ChIP-seq mapping; Brf1, Rpc34, and IN ChIP-seq experiments and data analysis; DNA curvature and bendability analysis; and figure preparation. F.S. conducted transposition qPCR assays, distance analysis of motifs, and aided in figure preparation. I.C. was responsible for Circos mapping of Ty3 targets. C.M. and P.B. designed and scripted software to catalog initial Ty3 genome transposition events. K.P. and S.B.S. wrote the manuscript with assistance from V.B., F.S., and C.M.

References

Achuthan V, Perreira JM, Sowd GA, Puray-Chavez M, McDougall WM, Paulucci-Holthauzen A, Wu X, Fadel HJ, Poeschla EM, Multani AS, et al. 2018. Capsid-CPSF6 interaction licenses nuclear HIV-1 trafficking to sites of viral DNA integration. *Cell Host Microbe* **24:** 392–404.e8. doi:10.1016/j.chom.2018.08.002

Aiyer S, Swapna GV, Malani N, Aramini JM, Schneider WM, Plumb MR, Ghanem M, Larue RC, Sharma A, Studamire B, et al. 2014. Altering murine leukemia virus integration through disruption of the integrase and BET protein family interaction. *Nucleic Acids Res* **42:** 5917–5928. doi:10.1093/nar/gku175

Aye M, Dildine SL, Claypool JA, Jourdain S, Sandmeyer SB. 2001. A truncation mutant of the 95-kilodalton subunit of transcription factor IIIC reveals asymmetry in Ty3 integration. *Mol Cell Biol* **21:** 7839–7851. doi:10.1128/MCB.21.22.7839-7851.2001

Bachman N, Eby Y, Boeke JD. 2004. Local definition of Ty1 target preference by long terminal repeats and clustered tRNA genes. *Genome Res* **14:** 1232–1247. doi:10.1101/gr.2052904

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37:** W202–W208. doi:10.1093/nar/gkp335

Ballandras-Colas A, Brown M, Cook NJ, Dewdney TG, Demeler B, Cherepanov P, Lyumkis D, Engelman AN. 2016. Cryo-EM reveals a novel octameric integrase structure for betaretroviral intasome function. *Nature* **530:** 358–361. doi:10.1038/nature16955

Baller JA, Gao J, Stamenova R, Curcio MJ, Voytas DF. 2012. A nucleosomal surface defines an integration hotspot for the *Saccharomyces cerevisiae* Ty1 retrotransposon. *Genome Res* **22:** 704–713. doi:10.1101/gr.129585.111

Bansal M, Kumar A, Yella VR. 2014. Role of DNA sequence based structural features of promoters in transcription initiation and gene expression. *Curr Opin Struct Biol* **25:** 77–85. doi:10.1016/j.sbi.2014.01.007

Berry C, Hannenhalli S, Leipzig J, Bushman FD. 2006. Selection of target sites for mobile DNA integration in the human genome. *PLoS Comput Biol* **2:** e157. doi:10.1371/journal.pcbi.0020157

Boukaba A, Georgieva EI, Myers FA, Thorne AW, López-Rodas G, Crane-Robinson C, Franco L. 2004. A short-range gradient of histone H3 acetylation and Tup1p redistribution at the promoter of the *Saccharomyces cerevisiae SUC2* gene. *J Biol Chem* **279:** 7678–7684. doi:10.1074/jbc.M310849200

Brogaard K, Xi L, Wang JP, Widom J. 2012. A map of nucleosome positions in yeast at base-pair resolution. *Nature* **486:** 496–501. doi:10.1038/nature11142

Casaregola S, Barth G. 2013. *Transposable elements and their activities in Yarrowia lipolytica*. Springer, Berlin.

Chalker DL, Sandmeyer SB. 1992. Ty3 integrates within the region of RNA polymerase III transcription initiation. *Genes Dev* **6:** 117–128. doi:10.1101/gad.6.1.117

Chalker DL, Sandmeyer SB. 1993. Sites of RNA polymerase III transcription initiation and Ty3 integration at the U6 gene are positioned by the TATA box. *Proc Natl Acad Sci* **90:** 4927–4931. doi:10.1073/pnas.90.11.4927

Chatterjee AG, Esnault C, Guo Y, Hung S, McQueen PG, Levin H. 2014. Serial number tagging reveals a prominent sequence preference of retrotransposon integration. *Nucleic Acids Res* **42:** 8449–8460. doi:10.1093/nar/gku534

Cherepanov P, Maertens G, Proost P, Devreese B, Van Beeumen J, Engelborghs Y, De Clercq E, Debyser Z. 2003. HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J Biol Chem* **278:** 372–381. doi:10.1074/jbc.M209278200

Cheung S, Manhas S, Measday V. 2018. Retrotransposon targeting to RNA polymerase III-transcribed genes. *Mob DNA* **9:** 14. doi:10.1186/s13100-018-0119-2

Cole HA, Ocampo J, Iben JR, Chereji RV, Clark DJ. 2014. Heavy transcription of yeast genes correlates with differential loss of histone H2B relative to H4 and queued RNA polymerases. *Nucleic Acids Res* **42:** 12512–12522. doi:10.1093/nar/gku1013

Connolly CM, Sandmeyer SB. 1997. RNA polymerase III interferes with Ty3 integration. *FEBS Lett* **405:** 305–311. doi:10.1016/S0014-5793(97)00200-7

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* **14:** 1188–1190. doi:10.1101/gr.849004

De Rijck J, de Kogel C, Demeulemeester J, Vets S, El Ashkar S, Malani N, Bushman FD, Landuyt B, Husson SJ, Busschots K, et al. 2013. The BET family of proteins targets moloney murine leukemia virus integration near transcription start sites. *Cell Rep* **5:** 886–894. doi:10.1016/j.celrep.2013.09.040

Dray S, Dufour AB. 2007. The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw* **22:** 1–20. doi:10.18637/jss.v022.i04

Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, Lee C, Shendure J, Fields S, Blau CA, Noble WS. 2010. A three-dimensional model of the yeast genome. *Nature* **465:** 363–367. doi:10.1038/nature08973

Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuveglise C, Talla E, et al. 2004. Genome evolution in yeasts. *Nature* **430:** 35–44. doi:10.1038/nature02579

Eidahl JO, Crowe BL, North JA, McKee CJ, Shkriabai N, Feng L, Plumb M, Graham RL, Gorelick RJ, Hess S, et al. 2013. Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. *Nucleic Acids Res* **41:** 3924–3936. doi:10.1093/nar/gkt074

Engelke DR, Gegenheimer P, Abelson J. 1985. Nucleolytic processing of a tRNA^{Arg}-tRNA^{Asp} dimeric precursor by a homologous component from *Saccharomyces cerevisiae*. *J Biol Chem* **260:** 1271–1279.

Engelman A, Cherepanov P. 2014. Retroviral integrase structure and DNA recombination mechanism. *Microbiol Spectr* **2:** 1–22. doi:10.1128/microbiolspec.MDNA3-0024-2014

Engelman AN, Singh PK. 2018. Cellular and molecular mechanisms of HIV-1 integration targeting. *Cell Mol Life Sci* **75:** 2491–2507. doi:10.1007/s00018-018-2772-5

Felice B, Cattoglio C, Cittaro D, Testa A, Miccio A, Ferrari G, Luzi L, Recchia A, Mavilio F. 2009. Transcription factor binding sites are genetic determinants of retroviral integration in the human genome. *PLoS One* **4:** e4571. doi:10.1371/journal.pone.0004571

Feng J, Liu T, Qin B, Zhang Y, Liu XS. 2012. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7:** 1728–1740. doi:10.1038/nprot.2012.101

Gao X, Hou Y, Ebina H, Levin HL, Voytas DF. 2008. Chromodomains direct integration of retrotransposons to heterochromatin. *Genome Res* **18:** 359–369. doi:10.1101/gr.7146408

Gilbert C, Feschotte C. 2018. Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. *Curr Opin Genet Dev* **49:** 15–24. doi:10.1016/j.gde.2018.02.007

Giuliodori S, Percudani R, Braglia P, Ferrari R, Guffanti E, Ottonello S, Dieci G. 2003. A composite upstream sequence motif potentiates tRNA gene transcription in yeast. *J Mol Biol* **333:** 1–20. doi:10.1016/j.jmb.2003.08.016

Grove A, Adessa MS, Geiduschek EP, Kassavetis GA. 2002. Marking the start site of RNA polymerase III transcription: the role of constraint, compaction and continuity of the transcribed DNA strand. *EMBO J* **21:** 704–714. doi:10.1093/emboj/21.4.704

Grunstein M, Gasser SM. 2013. Epigenetics in *Saccharomyces cerevisiae*. *Cold Spring Harb Perspect Biol* **5:** a017491. doi:10.1101/cshperspect.a017491

Guo Y, Singh PK, Levin HL. 2015. A long terminal repeat retrotransposon of *Schizosaccharomyces japonicus* integrates upstream of RNA pol III transcribed genes. *Mob DNA* **6:** 19. doi:10.1186/s13100-015-0048-2

Hansen LJ, Chalker DL, Sandmeyer SB. 1988. Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. *Mol Cell Biol* **8:** 5245–5256. doi:10.1128/MCB.8.12.5245

Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P. 2010. Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* **464:** 232–236. doi:10.1038/nature08784

Harismendy O, Gendrel CG, Soularue P, Gidrol X, Sentenac A, Werner M, Lefebvre O. 2003. Genome-wide location of yeast RNA polymerase III transcription machinery. *EMBO J* **22:** 4738–4747. doi:10.1093/emboj/cdg466

Hoffmann NA, Jakobi AJ, Moreno-Morcillo M, Glatt S, Kosinski J, Hagen WJ, Sachse C, Müller CW. 2015. Molecular structures of unbound and transcribing RNA polymerase III. *Nature* **528:** 231–236. doi:10.1038/nature16143

Hottinger-Werlen A, Schaack J, Lapointe J, Mao J, Nichols M, Söll D. 1985. Dimeric tRNA gene arrangement in *Schizosaccharomyces pombe* allows increased expression of the downstream gene. *Nucleic Acids Res* **13:** 8739–8747. doi:10.1093/nar/13.24.8739

Ji H, Moore DP, Blomberg MA, Braiterman LT, Voytas DF, Natsoulis G, Boeke JD. 1993. Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. *Cell* **73:** 1007–1018. doi:10.1016/0092-8674(93)90278-X

Juo ZS, Kassavetis GA, Wang J, Geiduschek EP, Sigler PB. 2003. Crystal structure of a transcription factor IIIB core interface ternary complex. *Nature* **422:** 534–539. doi:10.1038/nature01534

Kasinathan S, Orsi GA, Zentner GE, Ahmad K, Henikoff S. 2014. High-resolution mapping of transcription factor binding sites on native chromatin. *Nat Methods* **11:** 203–209. doi:10.1038/nmeth.2766

Kassavetis GA, Geiduschek EP. 2006. Transcription factor TFIIIB and transcription by RNA polymerase III. *Biochem Soc Trans* **34:** 1082–1087. doi:10.1042/BST0341082

Kassavetis GA, Riggs DL, Negri R, Nguyen LH, Geiduschek EP. 1989. Transcription factor IIIB generates extended DNA interactions in RNA polymerase III transcription complexes on tRNA genes. *Mol Cell Biol* **9:** 2551–2566. doi:10.1128/MCB.9.6.2551

Kassavetis GA, Braun BR, Nguyen LH, Geiduschek EP. 1990. *S. cerevisiae* TFIIIB is the transcription initiation factor proper of RNA polymerase III, while TFIIIA and TFIIIC are assembly factors. *Cell* **60:** 235–245. doi:10.1016/0092-8674(90)90739-2

Kassavetis GA, Letts GA, Geiduschek EP. 1999. A minimal RNA polymerase III transcription system. *EMBO J* **18:** 5042–5051. doi:10.1093/emboj/18.18.5042

Kassavetis GA, Letts GA, Geiduschek EP. 2001. The RNA polymerase III transcription initiation factor TFIIIB participates in two steps of promoter opening. *EMBO J* **20:** 2823–2834. doi:10.1093/emboj/20.11.2823

Kassavetis GA, Han S, Naji S, Geiduschek EP. 2003. The role of transcription initiation factor IIIB subunits in promoter opening probed by photochemical cross-linking. *J Biol Chem* **278:** 17912–17917. doi:10.1074/jbc.M300743200

Kirchner J, Sandmeyer SB. 1996. Ty3 integrase mutants defective in reverse transcription or 3′-end processing of extrachromosomal Ty3 DNA. *J Virol* **70:** 4737–4747.

Kirchner J, Connolly CM, Sandmeyer SB. 1995. Requirement of RNA polymerase III transcription factors for in vitro position-specific integration of a retroviruslike element. *Science* **267:** 1488–1491. doi:10.1126/science.7878467

Kirk PD, Huvet M, Melamed A, Maertens GN, Bangham CR. 2016. Retroviruses integrate into a shared, non-palindromic DNA motif. *Nat Microbiol* **2:** 16212. doi:10.1038/nmicrobiol.2016.212

Kjellin-Straby K, Engelke DR, Abelson J. 1984. Homologous *in vitro* transcription of linear DNA fragments containing the tRNA^{Arg}-tRNA^{Asp} gene pair from *Saccharomyces cerevisiae*. *DNA* **3:** 167–171. doi:10.1089/dna.1984.3.167

Klein SJ, O'Neill RJ. 2018. Transposable elements: genome innovation, chromosome diversity, and centromere conflict. *Chromosome Res* **26:** 5–23. doi:10.1007/s10577-017-9569-5

Kotterman MA, Chalberg TW, Schaffer DV. 2015. Viral vectors for gene therapy: translational and clinical outlook. *Annu Rev Biomed Eng* **17:** 63–89. doi:10.1146/annurev-bioeng-071813-104938

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19:** 1639–1645. doi:10.1101/gr.092759.109

Kumar Y, Bhargava P. 2013. A unique nucleosome arrangement, maintained actively by chromatin remodelers facilitates transcription of yeast tRNA genes. *BMC Genomics* **14:** 402. doi:10.1186/1471-2164-14-402

Kuras L, Struhl K. 1999. Binding of TBP to promoters *in vivo* is stimulated by activators and requires Pol II holoenzyme. *Nature* **399:** 609–613. doi:10.1038/21239

Küry P, Nath A, Créange A, Dolei A, Marche P, Gold J, Giovannoni G, Hartung HP, Perron H. 2018. Human endogenous retroviruses in neurological diseases. *Trends Mol Med* **24:** 379–394. doi:10.1016/j.molmed.2018.02.007

Kvaratskhelia M, Sharma A, Larue RC, Serrao E, Engelman A. 2014. Molecular mechanisms of retroviral integration site selection. *Nucleic Acids Res* **42:** 10209–10225. doi:10.1093/nar/gku769

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25. doi:10.1186/gb-2009-10-3-r25

Lesbats P, Engelman AN, Cherepanov P. 2016. Retroviral DNA integration. *Chem Rev* **116:** 12730–12757. doi:10.1021/acs.chemrev.6b00125

Lewinski MK, Yamashita M, Emerman M, Ciuffi A, Marshall H, Crawford G, Collins F, Shinn P, Leipzig J, Hannenhalli S, et al. 2006. Retroviral DNA integration: viral and cellular determinants of target-site selection. *PLoS Pathog* **2:** e60. doi:10.1371/journal.ppat.0020060

Lin SS, Nymark-McMahon MH, Yieh L, Sandmeyer SB. 2001. Integrase mediates nuclear localization of Ty3. *Mol Cell Biol* **21:** 7826–7838. doi:10.1128/MCB.21.22.7826-7838.2001

Maertens GN, Hare S, Cherepanov P. 2010. The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* **468:** 326–329. doi:10.1038/nature09517

Magnan C, Yu J, Chang I, Jahn E, Kanomata Y, Wu J, Zeller M, Oakes M, Baldi P, Sandmeyer S. 2016. Sequence assembly of *Yarrowia lipolytica* strain W29/CLIB89 shows transposable element diversity. *PLoS One* **11:** e0162363. doi:10.1371/journal.pone.0162363

Male G, von Appen A, Glatt S, Taylor NM, Cristovao M, Groetsch H, Beck M, Müller CW. 2015. Architecture of TFIIIC and its role in RNA polymerase III pre-initiation complex assembly. *Nat Commun* **6:** 7387. doi:10.1038/ncomms8387

Malik HS, Eickbush TH. 1999. Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. *J Virol* **73:** 5186–5190.

Manhas S, Ma L, Measday V. 2018. The yeast Ty1 retrotransposon requires components of the nuclear pore complex for transcription and genomic integration. *Nucleic Acids Res* **46:** 3552–3578. doi:10.1093/nar/gky109

Maskell DP, Renault L, Serrao E, Lesbats P, Matadeen R, Hare S, Lindemann D, Engelman AN, Costa A, Cherepanov P. 2015. Structural basis for retroviral integration into nucleosomes. *Nature* **523:** 366–369. doi:10.1038/nature14495

Mavrich TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I, Pugh BF. 2008. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res* **18:** 1073–1083. doi:10.1101/gr.078261.108

Moqtaderi Z, Struhl K. 2004. Genome-wide occupancy profile of the RNA polymerase III machinery in *Saccharomyces cerevisiae* reveals loci with incomplete transcription complexes. *Mol Cell Biol* **24:** 4118–4127. doi:10.1128/MCB.24.10.4118-4127.2004

Munteanu MG, Vlahovicek K, Parthasarathy S, Simon I, Pongor S. 1998. Rod models of DNA: sequence-dependent anisotropic elastic modelling of local bending phenomena. *Trends Biochem Sci* **23:** 341–347. doi:10.1016/S0968-0004(98)01265-1

Nagarajavel V, Iben JR, Howard BH, Maraia RJ, Clark DJ. 2013. Global 'bootprinting' reveals the elastic architecture of the yeast TFIIIB–TFIIIC transcription complex *in vivo*. *Nucleic Acids Res* **41:** 8135–8143. doi:10.1093/nar/gkt611

Nov Klaiman T, Hosid S, Bolshoy A. 2009. Upstream curved sequences in *E. coli* are related to the regulation of transcription initiation. *Comput Biol Chem* **33:** 275–282. doi:10.1016/j.compbiolchem.2009.06.007

Pasi M, Mornico D, Volant S, Juchet A, Batisse J, Bouchier C, Parissi V, Ruff M, Lavery R, Lavigne M. 2016. DNA minicircles clarify the specific role of DNA structure on retroviral integration. *Nucleic Acids Res* **44:** 7830–7847. doi:10.1093/nar/gkw651

Pradeepa MM, Sutherland HG, Ule J, Grimes GR, Bickmore WA. 2012. Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing. *PLoS Genet* **8:** e1002717. doi:10.1371/journal.pgen.1002717

Qi X, Sandmeyer SB. 2012. *In vitro* targeting of strand transfer by the Ty3 retroelement integrase. *J Biol Chem* **287:** 18589–18595. doi:10.1074/jbc.M111.326025

Qi X, Daily K, Nguyen K, Wang H, Mayhew D, Rigor P, Forouzan S, Johnston M, Mitra RD, Baldi P, et al. 2012. Retrotransposon profiling of RNA polymerase III initiation sites. *Genome Res* **22:** 681–692. doi:10.1101/gr.131219.111

Repanas K, Zingler N, Layer LE, Schumann GG, Perrakis A, Weichenrieder O. 2007. Determinants for DNA target structure selectivity of the human LINE-1 retrotransposon endonuclease. *Nucleic Acids Res* **35:** 4914–4926. doi:10.1093/nar/gkm516

Roberts DN, Stewart AJ, Huff JT, Cairns BR. 2003. The RNA polymerase III transcriptome revealed by genome-wide localization and activity–occupancy relationships. *Proc Natl Acad Sci* **100:** 14695–14700. doi:10.1073/pnas.2435566100

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, Müller M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12:** 77. doi:10.1186/1471-2105-12-77

Roth SL, Malani N, Bushman FD. 2011. Gammaretroviral integration into nucleosomal target DNA *in vivo*. *J Virol* **85:** 7393–7401. doi:10.1128/JVI.00635-11

Sandmeyer S, Patterson K, Bilanchone V. 2015. Ty3, a position-specific retrotransposon in budding yeast. *Microbiol Spectr* **3:** MDNA3-0057-2014. doi:10.1128/microbiolspec.MDNA3-0057-2014

Sarkar D. 2008. *Lattice: multivariate data visualization with R*. Springer-Verlag, New York.

Serrao E, Krishnan L, Shun MC, Li X, Cherepanov P, Engelman A, Maertens GN. 2014. Integrase residues that determine nucleotide preferences at sites of HIV-1 integration: implications for the mechanism of target DNA binding. *Nucleic Acids Res* **42:** 5164–5176. doi:10.1093/nar/gku136

Sultana T, Zamborlini A, Cristofari G, Lesage P. 2017. Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat Rev Genet* **18:** 292–308. doi:10.1038/nrg.2017.7

Thomas CE, Ehrhardt A, Kay MA. 2003. Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet* **4:** 346–358. doi:10.1038/nrg1066

Trifonov EN, Sussman JL. 1980. The pitch of chromatin DNA is reflected in its nucleotide sequence. *Proc Natl Acad Sci* **77:** 3816–3820. doi:10.1073/pnas.77.7.3816

Vlahovicek K, Kaján L, Pongor S. 2003. DNA analysis servers: plot.it, bend.it, model.it and IS. *Nucleic Acids Res* **31:** 3686–3687. doi:10.1093/nar/gkg559

Voigt F, Wiedemann L, Zuliani C, Querques I, Sebe A, Mátés L, Izsvák Z, Ivics Z, Barabas O. 2016. Sleeping Beauty transposase structure allows rational design of hyperactive variants for genetic engineering. *Nat Commun* **7:** 11126. doi:10.1038/ncomms11126

Wickham H. 2016. *ggplot2: elegant graphics for data analysis*. Springer-Verlag, New York.

Winckler T, Szafranski K, Glöckner G. 2005. Transfer RNA gene-targeted integration: an adaptation of retrotransposable elements to survive in the compact *Dictyostelium discoideum* genome. *Cytogenet Genome Res* **110:** 288–298. doi:10.1159/000084961

Wu CC, Herzog F, Jennebach S, Lin YC, Pai CY, Aebersold R, Cramer P, Chen HT. 2012. RNA polymerase III subunit architecture and implications for open promoter complex formation. *Proc Natl Acad Sci* **109:** 19232–19237. doi:10.1073/pnas.1211665109

Yukawa Y, Dieci G, Alzapiedi M, Hiraga A, Hirai K, Yamamoto YY, Sugiura M. 2011. A common sequence motif involved in selection of transcription start sites of Arabidopsis and budding yeast tRNA genes. *Genomics* **97:** 166–172. doi:10.1016/j.ygeno.2010.12.001

Zentner GE, Henikoff S. 2013a. Mot1 redistributes TBP from TATA-containing to TATA-less promoters. *Mol Cell Biol* **33:** 4996–5004. doi:10.1128/MCB.01218-13

Zentner GE, Henikoff S. 2013b. Regulation of nucleosome dynamics by histone modifications. *Nat Struct Mol Biol* **20:** 259–266. doi:10.1038/nsmb.2470

Zentner GE, Tsukiyama T, Henikoff S. 2013. ISWI and CHD chromatin remodelers bind promoters but act in gene bodies. *PLoS Genet* **9:** e1003317. doi:10.1371/journal.pgen.1003317

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9:** R137. doi:10.1186/gb-2008-9-9-r137

Received August 10, 2018; accepted in revised form June 12, 2019.