Cell PRESS
Open ACCESS
Cell Reports
Report

First Exon Length Controls Active Chromatin Signatures and Transcription

Nicole I. Bieberstein,[1,2] Fernando Carrillo Oesterreich,[1,2] Korinna Straube,[1] and Karla M. Neugebauer[1,*]
[1]Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany
[2]These authors contributed equally to this work and are listed alphabetically
*Correspondence: neugebauer@mpi-cbg.de
http://dx.doi.org/10.1016/j.celrep.2012.05.019

SUMMARY

Here, we explore the role of splicing in transcription, employing both genome-wide analysis of human ChIP-seq data and experimental manipulation of exon-intron organization in transgenic cell lines. We show that the activating histone modifications H3K4me3 and H3K9ac map specifically to first exon-intron boundaries. This is surprising, because these marks help recruit general transcription factors (GTFs) to promoters. In genes with long first exons, promoter-proximal levels of H3K4me3 and H3K9ac are greatly reduced; consequently, GTFs and RNA polymerase II are low at transcription start sites (TSSs) and exhibit a second, promoter-distal peak from which transcription also initiates. In contrast, short first exons lead to increased H3K4me3 and H3K9ac at promoters, higher expression levels, accuracy in TSS usage, and a lower frequency of antisense transcription. Therefore, first exon length is predictive for gene activity. Finally, splicing inhibition and intron deletion reduce H3K4me3 levels and transcriptional output. Thus, gene architecture and splicing determines transcription quantity and quality as well as chromatin signatures.

INTRODUCTION

It is common knowledge that transgenes lacking introns are poorly expressed compared to those containing introns (Brinster et al., 1988), and removal of introns from endogenous genes markedly reduces transcription (Furger et al., 2002). How introns contribute to transcription remains an open question. One suggestion is that this phenomenon is due to sequence-specific elements present in introns (Parra et al., 2011). A distinct possibility is that splicing regulates transcription through positive feedback. Splicing could reinforce transcription if splicing factors bound to nascent RNA interact with RNA polymerase II (Pol II). For example, spliceosomal U1 snRNP associates with Pol II and has been implicated in enhancement of pre-initiation complex formation (Damgaard et al., 2008; Das et al., 2007). Other studies have implicated snRNPs and SR proteins in stimulating transcription elongation (Fong and Zhou, 2001; Lin et al., 2008). Recently, roles for chromatin in the regulation of splicing have emerged (Carrillo Oesterreich et al., 2011; Luco et al., 2011), suggesting additional modes of communication between splicing and transcription in vivo.

The stimulatory effect of intron insertion on transcription is dependent on promoter proximity (Furger et al., 2002; Rose, 2004). Thus, potential gene regions and features that might reveal relationships between splicing and transcription may be positioned near promoters. We became interested in the "activating" H3K4me3 and H3K9ac marks, which are concentrated globally near promoters and are proportional to Pol II occupancy and gene activity (Barski et al., 2007; Bernstein et al., 2005; Rahl et al., 2010; Zhou et al., 2011). H3K4me3 also interacts indirectly with the U2 snRNP and promotes co-transcriptional splicing (Sims et al., 2007), showing the influence of this activating mark on splicing. Importantly, both activating marks potentiate recruitment of the general transcription factor (GTF), TFIID, to promoter-proximal gene regions and antagonize transcriptional silencing by chromatin regulators (Klymenko and Müller, 2004; Vermeulen et al., 2007). Therefore, we sought to address the possibility that splicing may feedback to transcription via chromatin signatures near promoters.

Links between exon-intron organization and gene activity have not been examined globally. Characteristic patterns of H3K4me3 and H3K9ac near promoters have been revealed by previous analyses, in which ChIP-seq reads were aligned to TSSs. These marks peak slightly downstream of TSSs and decrease gradually over the first 1kb (Barski et al., 2007; Rahl et al., 2010). The majority of mammalian first exons are shorter than 1kb (Figure S1A), raising the possibility that first exon features may contribute to these promoter-proximal signatures. Here we investigate the role of first exons in transcription. We determine the distribution of activating chromatin marks as well as GTFs and Pol II with respect to exon-intron boundaries and use experimental models to further show that both splicing and first exon length determine chromatin signatures and transcriptional properties consistent with an enhancer-like activity.

RESULTS AND DISCUSSION

H3K4me3 and H3K9ac Are Concentrated at the Ends of First Exons

To address the possibility that activating histone marks localize to exon-intron landmarks, ChIP-seq data sets available through the human ENCODE project were analyzed (ENCODE Project

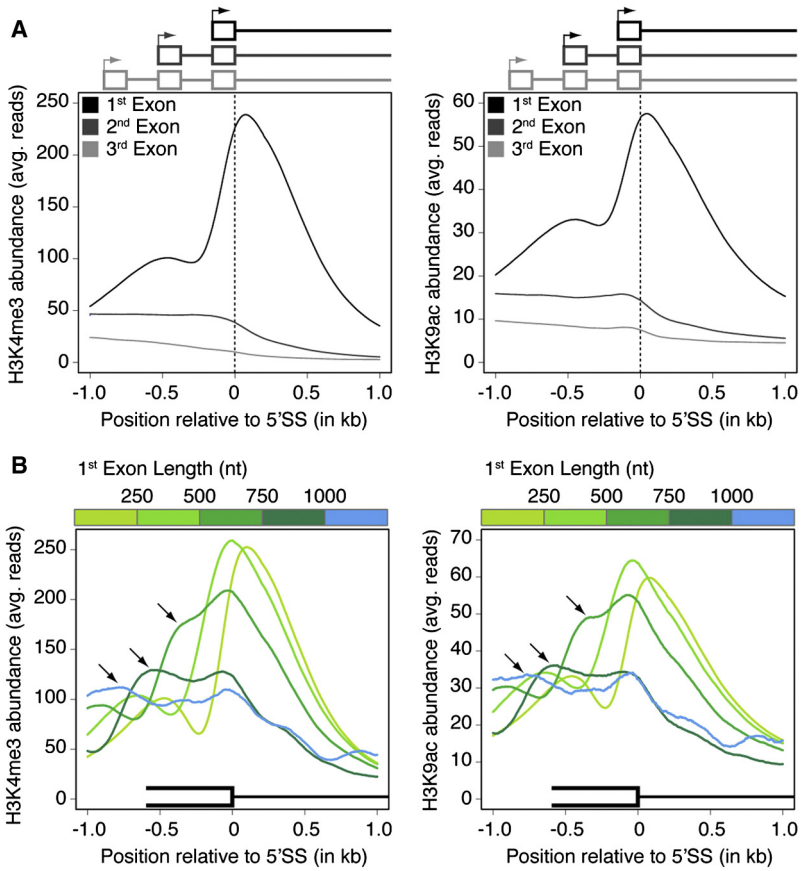

Figure 1. H3K4me3 and H3K9ac Are Enriched at the First 5′SS

Distributions of H3K4me3 (left panels) and H3K9ac (right panels) were determined by analysis of genome-wide ChIP-seq experiments in human K562 cells (ENCODE-SYDH-Histone).

(A) Average distribution of the histone marks aligned to the 5′ splice site (5′SS) of first, second and third exons. Both signals peak at first 5′SSs, and no significant enrichment is observed at downstream 5′SSs. Average distributions were calculated for all annotated intron-containing genes (Table S1).

(B) Average distributions aligned to first 5′SSs of genes grouped by first exon length (see Figure S1A and Table S1), as indicated by the color code above the traces. With increasing first exon length, two distinct peaks of H3K4me3 and H3K9ac become apparent, reflecting separate promoter (arrows) and 5′SS distributions.

See also Figures S1 and S2 and Table S1.

Consortium, 2011). H3K4me3 and H3K9ac distributions were aligned to the first three exon-intron boundaries—represented by annotated 5′ splice sites (5′SSs)—of all human intron-containing genes. Strikingly, alignment of both marks to first 5′SSs yields a tight distribution, whereas no enrichment was observed at second and third 5′SSs (Figure 1A). In contrast, alignment to the TSS produces a skewed peak, centered downstream of the TSS (Figure S1B), as observed previously (Barski et al., 2007; Rahl et al., 2010). Neither mark is present at 5′SSs further downstream, even when located close to the TSS (Figures 1A and S1C). Moreover, specific concentration at the first 5′SS is reproducible among mammalian cell lines and tissues (Figures S1D–S1F). Finally, other histone modifications assayed do not show the characteristic pattern observed for the activating marks (Figures S2A–S2D). We conclude that H3K4me3 and H3K9ac marks are specifically concentrated at the ends of first exons.

If activating histone marks are present at first exon-intron boundaries, are they also located near promoters? Because first exon lengths vary among the genes in the population, distinctions between promoters and 5′SSs are obscured by global analysis. Therefore, we grouped genes according to first exon length (Figure S1A) and specifically queried H3K4me3 and H3K9ac concentrations at first 5′SSs. Remarkably, as promoters and 5′SSs move apart with increasing first exon length, two distinct peaks become visible (Figure 1B). One peak is promoter-proximal, and one peak occurs at the first 5′SS. In the shortest first exon group (0–250 nt), these peaks are shifted slightly downstream of the 5′SS, because the nucleosome-depleted region extends beyond the location of the 5′SS (Figure S2E). Therefore, H3K4me3 and H3K9ac signals are concentrated at both 5′SSs and promoters; when first exons are longer than 500 nt, two prominent peaks are present.

It is currently believed that promoter-proximal H3K4me3 and H3K9ac marks are present on positioned nucleosomes; an appealing scenario would be that nucleosomes are specifically positioned over first 5′SSs. However, in contrast to this model, the first positioned nucleosome is present at a constant distance from the TSS, irrespective of first exon length (Figure S2F). Furthermore, positioned nucleosomes are not detected at first 5′SSs (Figure S2F), showing that nucleosome positioning per se does not play a role in the concentration of histone marks at exon-intron boundaries. Interestingly, up to four positioned nucleosomes downstream of the TSS can be marked by H3K4me3 (Figure S2G). We conclude that concentration of activating histone marks at the ends of first exons is independent of nucleosome positioning.

H3K4me3 Levels and Position Depend on Splicing

To test whether localization of activating marks at first 5′SSs represents a functional link between splicing and transcription, we established transgenic cell lines harboring model genes that differ in intron number and organization. To maintain an endogenous chromatin context and preserve regulation, a bacterial artificial chromosome (BAC) containing the mouse *FOS* gene with ~100 kb surrounding genomic sequence was obtained. Recombineering was used to generate intronless as well as single intron versions containing either intron1 or intron2 at endogenous positions (Figure 2A). Stable HeLa cell lines were generated for each transgene, and the mouse model genes were distinguished from endogenous human *FOS* with species-specific primers. In all of the following experiments, H3K4me3 served

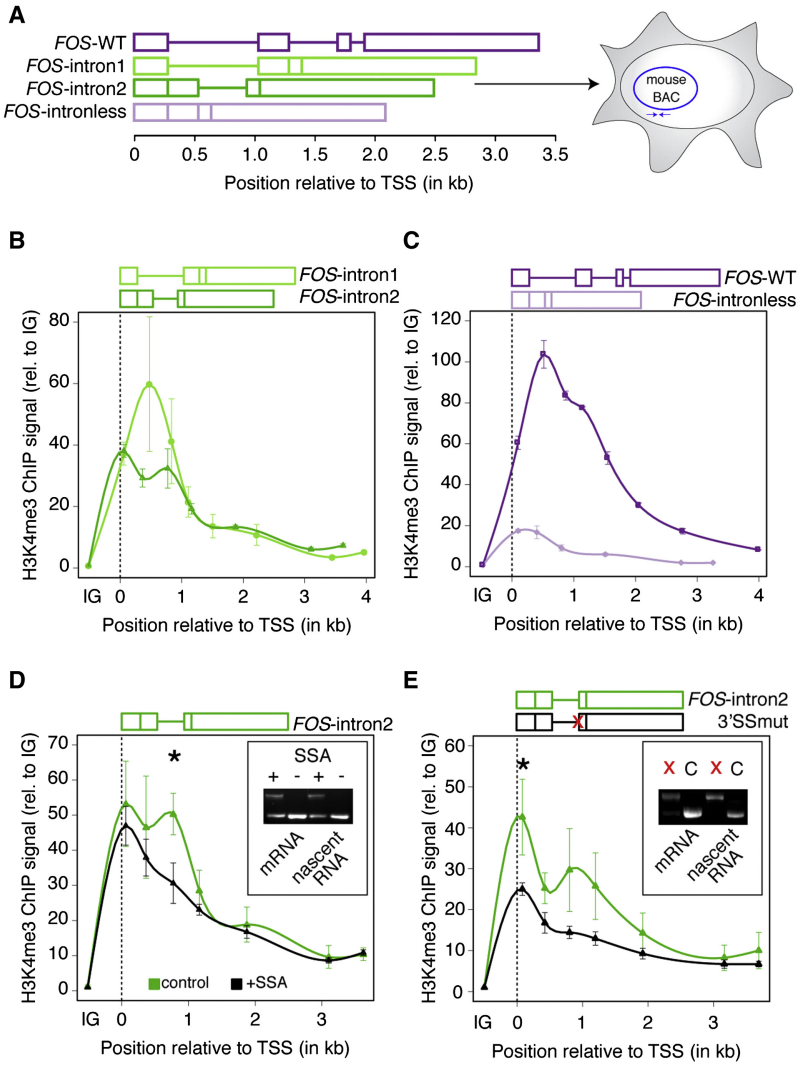

Figure 2. H3K4me3 Is Dependent on Splicing

(A) *FOS* model genes with varying exon-intron organization were created by BAC recombineering. Two single intron model genes (*FOS*-intron1 and *FOS*-intron2, named with respect to the remaining intron) and an intronless version were constructed by deleting endogenous introns. Each BAC was stably transfected into HeLa cells; mouse model genes were distinguished from endogenous human *FOS* with species-specific primers.

(B) H3K4me3 ChIP profiles along *FOS*-intron1 (light green) and *FOS*-intron2 (dark green) reveals distinct peak positions. A single H3K4me3 peak is present near the 5′SS of *FOS*-intron1; two distinct peaks near the promoter and the 5′SS are present in *FOS*-intron2. ChIP enrichment was calculated relative to input and normalized to an intergenic region (IG). Mean ± SEM are shown, n = 3–4.

(C) H3K4me3 ChIP profiles along *FOS*-WT (dark mauve) and *FOS*-intronless (light mauve). *FOS*-WT exhibits an ~5-fold increase in H3K4me3 levels and a shift in peak position toward the first 5′SS. Mean ± SEM are shown, n = 3–4.

(D) Inhibition of splicing with SSA leads to removal of the intron-associated H3K4me3 peak on *FOS*-intron 2 (black trace). Mean ± SEM are shown, n = 3. This effect is significant (*$p < 0.05$), although splicing inhibition was incomplete as judged by RT-PCR of total and nascent RNA (inset).

(E) Mutation of the *FOS*-intron2 3′SS (AG to TG) severely reduced splicing (inset, C = *FOS*-intron2, X = 3′SSmut) as well as the H3K4me3 peaks near the promoter and the 5′SS (black trace). *$p < 0.05$. Mean ± SEM are shown, n = 3–4.

See also Figure S3 and Table S1.

as a proxy for both activating marks, because H3K4me3 and H3K9ac distributions and functions are highly related (Figures 1, S1B–S1D, and S3A) (Vermeulen et al., 2007).

To determine whether first exon length specifies the position of H3K4me3 in promoter-proximal gene regions, ChIP was performed on the *FOS* model genes. *FOS*-intron1 contains only intron1 at its endogenous position with 280 nt spacing between 5′SS and TSS (Figure 2A); on this gene, H3K4me3 is concentrated in one peak positioned over the intron (Figure 2B). In contrast, fusion of exons 1 and 2 results in a longer first exon (503 nt) in *FOS*-intron2; here, two H3K4me3 peaks—one promoter-proximal and one near the 5′SS—become detectable (Figure 2B). The separability of these two peaks agrees with the genome-wide data, showing that first 5′SS distance from the promoter determines H3K4me3 profile.

The dependence of H3K4me3 profile on the position of the first 5′SS suggests that splicing itself may affect H3K4me3. We tested this in three ways. First, H3K4me3 profiles were determined on *FOS*-WT and *FOS*-intronless genes; removal of endogenous introns led to a 5-fold reduction of overall H3K4me3 signal with a single small peak shifted toward the promoter, independent of changes in nucleosome occupancy (Figures 2C and S3B). We conclude that intron content and position contributes to H3K4me3 levels as well as profile.

Second, splicing was inhibited by the small molecule spliceostatin A (SSA) (Kaida et al., 2007), which decreased splicing of *FOS*-intron2 mRNA by ~50%. SSA treatment led to the loss of the second H3K4me3 peak positioned over the intron (Figure 2D). Similarly, inhibition of both transcription and splicing by DRB abolished the second H3K4me3 peak (Figure S3C). Third, mutation of intronic sequences at the 3′ splice site blocks splicing more effectively and results in an overall loss of H3K4me3 signal, both close to the promoter and over intron2 (Figure 2E). None of the perturbations altered nucleosome distribution significantly (Figure S3D). Taken together, these data indicate that promoter-proximal splicing events result in marking of the first exon-intron boundary by H3K4me3.

Transcriptional Output Is Splicing-Dependent

The relationship between activating chromatin signatures, exon-intron organization, and splicing established above prompts the question of whether transcriptional output is directly affected by splicing. Taking advantage of our *FOS* model genes, we assayed transcriptional differences by RT-qPCR, using a primer

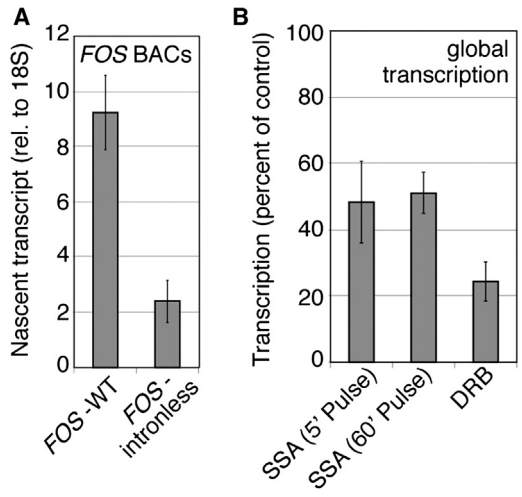

Figure 3. Transcriptional Output Is Splicing-Dependent

(A) The effect of introns on *FOS* transcription was measured by RT-qPCR on nascent RNA of *FOS*-WT and *FOS*-intronless. Normalization to 18S rRNA accounts for variation in cell number. Mean ± SEM are shown, n = 3.

(B) Metabolic labeling shows the effect of splicing inhibition on global transcription. Cells were labeled with P^{32}-orthophosphate for 5 min or 1 hr in the presence or absence of SSA. Direct inhibition of Pol II transcription with DRB is shown for comparison. Incorporation upon treatment is shown as a fraction of control. Mean ± SEM are shown, n > = 3.

See also Figure S4 and Table S1.

downstream of the polyA cleavage site to select nascent RNA. Figure 3A shows that *FOS*-intronless exhibits 4-fold reduced transcription levels compared to intron-containing *FOS*-WT. In a second approach, potential feedback from splicing to transcription was assessed globally, by inhibiting splicing with SSA and measuring total transcription by metabolic labeling. DRB treatment indicates the expected maximum effect caused by direct inhibition of transcription to be ~4-fold (Figure 3B). Global transcription was reduced ~2-fold by SSA (Figure 3B), in agreement with decreased transcription of intronless *FOS* and our analysis of published tiling-microarray data (Figure S4). To test whether this effect was due to RNA degradation, the radioactive pulse was shortened to 5 min, below the time needed to transcribe the average human gene. The relative change in transcription upon SSA treatment was not dependent on the length of the pulse, indicating an effect on transcription and not on RNA stability (Figure 3B).

Short First Exons Promote Transcriptional Accuracy

How does splicing exert a positive influence on transcription? Above, we have shown that promoter-proximal profiles of activating histone marks depend on first exon length and splicing. Because the general transcription factor (GTF) TFIID binds directly to H3K4me3 (Vermeulen et al., 2007), we hypothesized that exon-intron organization influences recruitment profiles of GTFs. If so, genes with short first exons are expected to exhibit focused recruitment of GTFs near TSSs. Conversely, genes with long first exons are expected to display broader GTF profiles with peaks downstream of TSSs. To test this, ChIP-seq traces of GTFs were aligned to TSSs of genes grouped according to first exon length. TFIID, TFIIB, and TFIIF peak at annotated TSSs for all gene groups (Figures 4A and S5A). However, genes with long first exons display decreased peak localization at TSSs and additional downstream peaks at 5′SSs (Figures 4A and S5B). These data suggest that the fidelity of Pol II recruitment to annotated TSSs may be compromised when first exons are long. Indeed, promoter-proximal Pol II profiles show a similar dependence on first exon length, and Pol II is localized to 5′SSs downstream of TSSs (Figures 4B and S5C). Finally, genes with long first exons display lower transcript levels than genes with short first exons (Figure S5D). We conclude that first exon length determines the profiles and degree to which GTFs and Pol II concentrate at annotated promoters, such that short exons—by virtue of their proximity to TSSs—provide the greatest positive feedback to transcription.

Does transcription actually initiate downstream of annotated TSSs when first exons are long? To test this, the distribution of EST 5′ ends was analyzed for the different gene groups. In all groups, EST 5′ ends peak around annotated TSSs (Figures 4C and S5E). Remarkably, genes with long first exons exhibit extensive TSS usage within the first exon and at the 5′SS, confirming promoter-distal initiation (Figures 4C and S5C). In contrast, short exons appear to enforce usage of canonical TSSs, in agreement with more robust localization of activating histone marks, GTFs and Pol II.

Antisense transcription is common and may reflect the availability of both DNA strands in promoter regions (Wei et al., 2011). Indeed, nucleosomes are relatively depleted along the length of all first exons (Figure S2F). Thus, initiation at promoter-distal positions described above might lead to higher levels of antisense transcription in genes with long first exons. Therefore, we asked whether exon-intron organization impacts the orientation of transcription initiation. To address this question, the fraction of sense versus antisense transcripts was analyzed for the gene groups. A significant increase in the proportion of antisense transcription was observed for genes with long first exons (Figure 4D). This indicates that spatial separation of promoter and 5′SS impacts directionality. We propose that short exons enforce sense transcription, by reducing downstream initiation.

First 5′ Splice Sites Are Position-Dependent Transcriptional Enhancers

Four lines of evidence provided here indicate that exon-intron organization contributes significantly to the transcriptional properties of genes: (1) activating histone marks, GTFs, and Pol II align to promoters and first 5′SSs genome-wide, with 5′SSs determining H3K4me3 and H3K9ac profiles near promoters, (2) short first exons are associated with robust concentration of GTFs and Pol II at promoters, enforcing uniform transcription initiation at TSSs and promoting higher levels of gene expression, (3) intron-associated H3K4me3 profiles as well as global transcriptional output are splicing-dependent, and (4) sense transcription is favored at genes with short first exons. Moreover, the present findings provide an explanation for recent reports showing that H3K36me3 marks present in downstream regions of active genes are splicing-dependent (de Almeida et al., 2011; Kim et al., 2011). H3K36me3 modification is generated

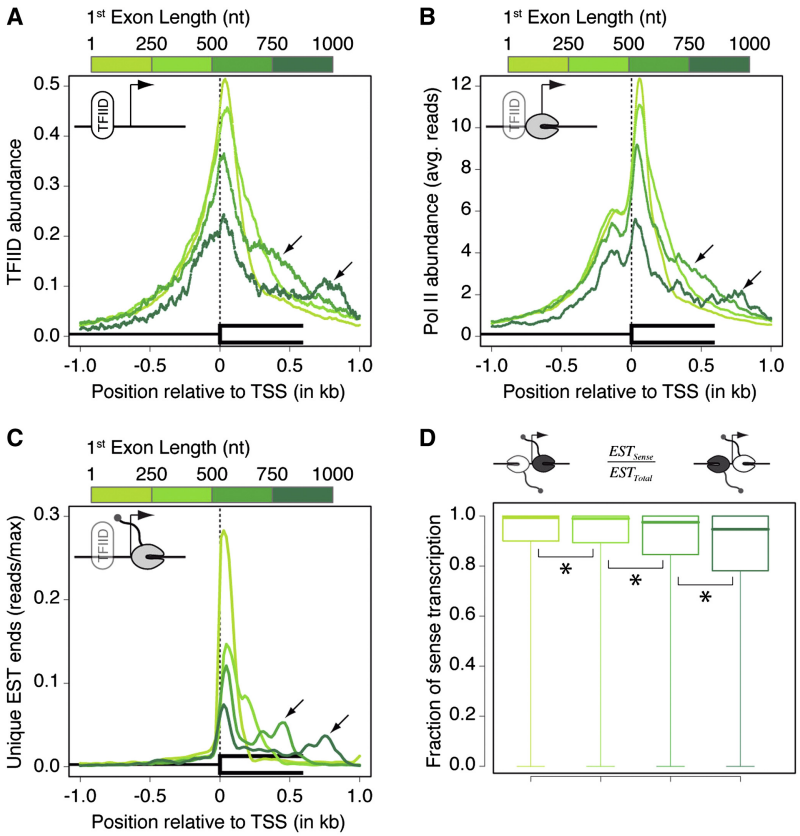

Figure 4. First Exon Length Enforces Transcription Quality

(A and B) Average distributions of TFIID (ENCODE-HAIB-TFBS) (A), and Pol II (ENCODE-SYDH-TFBS) (B) around TSSs were determined for genes grouped according to first exon length (see Figure S1A and color-code). Peak levels at annotated TSSs decrease with increasing first exon lengths, whereas levels within first exons increase (marked by arrows).

(C) Distribution of initiation events (unique EST 5′ ends) (Fujita et al., 2011) around annotated gene starts was analyzed for genes grouped according to first exon length. Average traces show that initiation events are more dispersed when genes have long first exons, qualitatively following the recruitment of Pol II (B).

(D) The fraction of sense transcription relative to total transcription was deduced for first exon length groups described above. Mann-Whitney test between each neighboring pair shows that antisense transcription is significantly increased with increasing first exon lengths ($p < 0.001$). Boxes represent values within three quartiles of the data, whereas whiskers indicate the range of the data set. Black horizontal lines in the respective boxes show median values.

See also Figure S5 and Table S1.

by the Setd2 methyltransferase, recruited by elongating Pol II; therefore, splicing-dependency of H3K36me3 is likely a consequence of reduced H3K4me3 levels near promoters, leading to decreased transcriptional firing.

These observations suggest that gene architecture has evolved to incorporate the role of splicing in transcription regulation, in a manner analogous to gene-specific enhancers and promoters. Specifically, our data indicate that short first exons provide positive feedback to promoters and minimize splicing-dependent downstream initiation. Accordingly, genes with long first exons are less transcriptionally active on average. One half of all human genes have first exon lengths in the range of 128–350 nt (Figure S1A). Remarkably, these genes show a single peak of activating histone marks, which is perfectly positioned at first 5′SSs (Figures 1B and S2E). Conversely 25% of genes are shorter or longer than this optimal first exon length window, suggesting that these genes forego this splicing-dependent enhancer feature, either to enable other forms of regulation (e.g., by transcription factors), to permit low expression levels, or to promote alternative TSS usage.

How persistent this mode of splicing-dependent transcription regulation might be is an interesting question, because transcription and splicing are largely shut down during mitosis, upon heat shock, and in gametes and early embryos. Interestingly, transcriptionally inert sperm exhibit a similar distribution of H3K4me3 at first 5′SSs (Figure S1F), suggesting that prior transcription and splicing activity determines germ cell chromatin signatures during gametogenesis (Hammoud et al., 2009). This indicates that the transcription and splicing history of genes could impact chromatin signatures and thereby subsequent phases of gene activity over longer periods. How previous rounds of transcription and splicing contribute to the establishment of chromatin states in early development, including the intriguing interplay between activating and repressive marks seen among "bivalent" genes, remains to be investigated (Eissenberg and Shilatifard, 2010; Vastenhouw and Schier, 2012).

EXPERIMENTAL PROCEDURES

Cell Culture and Treatments

HeLa Kyoto cells were cultured in high glucose (4.5 g/l) DMEM GlutaMax (Invitrogen) supplemented with 100 U/ml penicillin, 100 μg/ml streptomycin (penicillin/streptomycin, PAA) and 10% (v/v) fetal bovine serum (FBS, Invitrogen) at 37°C and 5% CO_2. When culturing BAC transgenic cell lines, selection was carried out with 200 μg/ml hygromycin B (Roche). *FOS* model genes were induced by 2h serum starvation in DMEM GlutaMax without FBS followed by 15 min addition of 5 μM calcimycin (A-23187 free acid, Invitrogen). For splicing inhibition, cells were treated with 100 ng/ml spliceostatin A (SSA, kindly provided by Minoru Yoshida, RIKEN Advanced Science Institute, Japan) (Kaida et al., 2007) or control-treated with 0.1% (v/v) methanol for 3 hr. Transcription was inhibited by addition of 100 μM DRB for 6 hr. For control treatment, an equal volume of the solvent DMSO was used.

BAC Recombineering

The BAC clone RP24-208N11 harboring the mouse *FOS* locus was obtained from the BACPAC Resources Center. For selection in eukaryotic cells, a hygromycin resistance cassette (kindly provided by Francis Stewart) was introduced 10 kb downstream of the *FOS* gene using Red/ET recombineering (Zhang et al., 1998). Model genes based on the mouse *FOS* gene were constructed by Red/ET recombineering using the Gene Bridges Counter Selection Kit.

Antibodies

The following commercial antibodies were used in this study: α-H3K4me3 (Millipore 07-473), α-H3 (Abcam ab1791), and α-IgG (Sigma I-5256).

Chromatin Immunoprecipitation

Chromatin Immunoprecipitation and qPCR procedures were modified from Listerman et al. (2006). Magnetic Dynabeads Protein G (Invitrogen) were used for immunoprecipitation. A list of primers used for qPCR and a more detailed protocol is provided in the Extended Experimental Procedures.

RNA-Extraction and RT-qPCR

Cells were grown to 90% confluency on a 6-well plate (~10^6 cells), RNA isolated using TRIzol extraction (Invitrogen), precipitated, resuspended in nuclease free H_2O and DNase-treated with DNAfree (Ambion). Reverse transcription was performed using Superscript III (Invitrogen) with 0.5–5 μg RNA per reaction. For reverse transcription of mRNA, a gene specific primer in the terminal exon was used. Nascent RNA was selected, using a reverse primer located downstream of the annotated poly(A) cleavage site (Carrillo Oesterreich et al., 2010; Pandya-Jones and Black, 2009). The cDNA was analyzed by conventional or quantitative PCR. A list of primers and a detailed description of the qPCR analysis is provided in the Extended Experimental Procedures.

Global Transcriptional Activity

To measure transcriptional activity, radioactive orthophosphate (32 P, 200 μCi) was added to HeLa Kyoto cells grown to 70% confluency on 6-well plates in Phosphate free medium (Invitrogen). After further incubation for 1 hr or 5 min cells were lysed and total RNA extracted using the TRIzol reagent (Invitrogen). RNA was precipitated, resuspended in H_2O and incorporation of radioactive nucleotides into the RNA measured by scintillation counting. Values gained from treated cells were normalized to untreated controls.

Genome-wide Analysis

For detailed information on data sets and methods employed for genome-wide analysis please refer to the Extended Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2012.05.019.

LICENSING INFORMATION

WEB RESOURCES

The URLs for data presented herein are as follows:

BACPAC Resources Center, http://bacpac.chori.org/
ENCODE-Broad-Histone, http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHistone/
ENCODE-Caltech-RNASeq, http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCaltechRnaSeq/
ENCODE-HAIB-TFBS. http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeHaibTfbs/
ENCODE-LICR-Histone. http://hgdownload.cse.ucsc.edu/goldenPath/mm9/encodeDCC/wgEncodeLicrHistone/
ENCODE-Stanf-Nucleosome. http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeSydhNsome/
ENCODE-SYDH-Histone. http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeSydhHistone/
ENCODE-SYDH-TFBS. http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeSydhTfbs/
ENCODE-UW-Histone. http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwHistone/
FastQC, http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/
Primer3Plus, http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi
The R Project for Statistical Computing, http://www.r-project.org/
RefSeq, hg19, http://www.ncbi.nlm.nih.gov/RefSeq/
University of California, Santa Cruz Encyclopedia of DNA Elements (ENCODE), http://genome.ucsc.edu/ENCODE/

ACKNOWLEDGMENTS

We thank Stephan Grill, Michael Hiller, Jonathon Howard, Francis Stewart, Vineeth Surendranath, Tim Mercer, and Pavel Tomancak for helpful discussions and comments on the manuscript. Expert technical assistance by Claudia Semprich is gratefully acknowledged. We are grateful to Minoru Yoshida for his generous gift of SSA. This research was funded by the Max Planck Society and the Deutsche Forschungsgemeinschaft (NE909/3-1 to K.M.N.).

Received: March 13, 2012
Revised: April 17, 2012
Accepted: May 22, 2012
Published online: July 19, 2012

REFERENCES

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169–181.

Brinster, R.L., Allen, J.M., Behringer, R.R., Gelinas, R.E., and Palmiter, R.D. (1988). Introns increase transcriptional efficiency in transgenic mice. Proc. Natl. Acad. Sci. USA *85*, 836–840.

Carrillo Oesterreich, F., Bieberstein, N., and Neugebauer, K.M. (2011). Pause locally, splice globally. Trends Cell Biol. *21*, 328–335.

Carrillo Oesterreich, F., Preibisch, S., and Neugebauer, K.M. (2010). Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. Mol. Cell *40*, 571–581.

Damgaard, C.K., Kahns, S., Lykke-Andersen, S., Nielsen, A.L., Jensen, T.H., and Kjems, J. (2008). A 5′ splice site enhances the recruitment of basal transcription initiation factors in vivo. Mol. Cell *29*, 271–278.

Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P., and Reed, R. (2007). SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. Mol. Cell *26*, 867–881.

de Almeida, S.F., Grosso, A.R., Koch, F., Fenouil, R., Carvalho, S., Andrade, J., Levezinho, H., Gut, M., Eick, D., Gut, I., et al. (2011). Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. Nat. Struct. Mol. Biol. *18*, 977–983.

Eissenberg, J.C., and Shilatifard, A. (2010). Histone H3 lysine 4 (H3K4) methylation in development and differentiation. Dev. Biol. *339*, 240–249.

ENCODE Project Consortium. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. *9*, e1001046.

Fong, Y.W., and Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. Nature *414*, 929–933.

Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A., et al. (2011). The UCSC Genome Browser database: update 2011. Nucleic Acids Res. *39* (Database issue), D876–D882.

Furger, A., O'Sullivan, J.M., Binnie, A., Lee, B.A., and Proudfoot, N.J. (2002). Promoter proximal splice sites enhance transcription. Genes Dev. *16*, 2792–2799.

Hammoud, S.S., Nix, D.A., Zhang, H., Purwar, J., Carrell, D.T., and Cairns, B.R. (2009). Distinctive chromatin in human sperm packages genes for embryo development. Nature *460*, 473–478.

Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., Watanabe, H., Kitahara, T., Yoshida, T., Nakajima, H., et al. (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nat. Chem. Biol. *3*, 576–583.

Kim, S., Kim, H., Fong, N., Erickson, B., and Bentley, D.L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. Proc. Natl. Acad. Sci. USA *108*, 13564–13569.

Klymenko, T., and Müller, J. (2004). The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. EMBO Rep. *5*, 373–377.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. Nat. Struct. Mol. Biol. *15*, 819–826.

Listerman, I., Sapra, A.K., and Neugebauer, K.M. (2006). Cotranscriptional coupling of splicing factor recruitment and precursor messenger RNA splicing in mammalian cells. Nat. Struct. Mol. Biol. *13*, 815–822.

Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. Cell *144*, 16–26.

Pandya-Jones, A., and Black, D.L. (2009). Co-transcriptional splicing of constitutive and alternative exons. RNA *15*, 1896–1908.

Parra, G., Bradnam, K., Rose, A.B., and Korf, I. (2011). Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants. Nucleic Acids Res. *39*, 5328–5337.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432–445.

Rose, A.B. (2004). The effect of intron location on intron-mediated enhancement of gene expression in Arabidopsis. Plant J. *40*, 744–751.

Sims, R.J., 3rd, Millhouse, S., Chen, C.-F., Lewis, B.A., Erdjument-Bromage, H., Tempst, P., Manley, J.L., and Reinberg, D. (2007). Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. Mol. Cell *28*, 665–676.

Vastenhouw, N.L., and Schier, A.F. (2012). Bivalent histone modifications in early embryogenesis. Curr. Opin. Cell Biol. *24*, 374–386.

Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W.M.P., van Schaik, F.M.A., Varier, R.A., Baltissen, M.P.A., Stunnenberg, H.G., Mann, M., and Timmers, H.T.M. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. Cell *131*, 58–69.

Wei, W., Pelechano, V., Järvelin, A.I., and Steinmetz, L.M. (2011). Functional consequences of bidirectional promoters. Trends Genet. *27*, 267–276.

Zhang, Y., Buchholz, F., Muyrers, J.P., and Stewart, A.F. (1998). A new logic for DNA engineering using recombination in Escherichia coli. Nat. Genet. *20*, 123–128.

Zhou, V.W., Goren, A., and Bernstein, B.E. (2011). Charting histone modifications and the functional organization of mammalian genomes. Nat. Rev. Genet. *12*, 7–18.