nature structural & molecular biology

Chromatin organization marks exon-intron structure

Schraga Schwartz[1], Eran Meshorer[2] & Gil Ast[1]

An increasing body of evidence indicates that transcription and splicing are coupled, and it is accepted that chromatin organization regulates transcription. Little is known about the cross-talk between chromatin structure and exon-intron architecture. By analysis of genome-wide nucleosome-positioning data sets from humans, flies and worms, we found that exons show increased nucleosome-occupancy levels with respect to introns, a finding that we link to differential GC content and nucleosome-disfavoring elements between exons and introns. Analysis of genome-wide chromatin immunoprecipitation data in humans and mice revealed four specific post-translational histone modifications enriched in exons. Our findings indicate that previously described enrichment of H3K36me3 modifications in exons reflects a more fundamental phenomenon, namely increased nucleosome occupancy along exons. Our results suggest an RNA polymerase II–mediated cross-talk between chromatin structure and exon-intron architecture, implying that exon selection may be modulated by chromatin structure.

Mammalian genomes are packaged together with histone proteins in the form of chromatin. The intertwined DNA harbors genes, most of which are made up of short stretches of exonic sequences interrupted by long noncoding introns. This organization imposes two distinct code sets: the splicing code and the chromatin code (nucleosome occupancy). The splicing code, which comprises a set of four signals at the exon-intron junctions and a vast array of splicing regulatory elements (SREs), directs the spliceosomal machinery to the exon-intron boundaries, allowing precise identification of exons[1–4]. Despite decades of research, the factors allowing differentiation of exons from long flanking introns are far from understood, especially for 'higher' eukaryotes. Although exon lengths in these organisms seem to be under strong evolutionary pressure to remain within a constant range of ~140 nucleotides (nt)[5], introns have expanded to several thousands of nucleotides in length, and their length does not seem to be under evolutionary selection.

DNA sequence modulates how and where the DNA is packaged around nucleosomes, a concept embodied in the idea of a 'chromatin code': this information is to a great extent encoded directly within the genome sequence[6,7]. Nucleosome occupancy is modulated by means of specific modifications of histone tails, including acetylation, methylation, phosphorylation and ubiquitination[8–10]. By regulating chromatin structure and DNA accessibility, these modifications influence and modulate gene expression levels in different developmental stages, tissue types and disease states[9,11].

The splicing code and the chromatin codes are traditionally understood as acting at two different levels: the chromatin at the DNA level and the splicing code at the level of RNA. However, an increasing body of evidence suggests that processes occurring at these two levels are coupled. Splicing occurs co-transcriptionally, and introns are removed while the nascent transcript is still tethered to the DNA by RNA polymerase II (RNAPII)[12–14]. RNAPII is also associated with many splicing factors via its C-terminal domain (CTD)[12,15,16], and the transcription rate of RNAPII affects splicing[17–19]. In parallel, during transcription, chromatin is altered by nucleosome displacement and histone tail modifications. Several chromatin remodelers are associated with RNAPII, and some of these factors are also involved in the regulation of pre-mRNA splicing[20,21]. Moreover, recently published reports demonstrate that two different splicing regulators from the family of SR proteins dynamically interact with chromatin[22,23]. Finally, recruitment of splicing factors is mediated by H3K4 trimethylation[24], and the U1 small nuclear RNA is associated with chromatin[25]. These findings raised the possibility that chromatin structure, histones and the chemical modifications to which they are subjected help direct and modulate splicing.

In this study, we set out to explore the potential interplay between chromatin structure and exon-intron architecture. On the basis of experimental data in humans, *Drosophila melanogaster* and *Caenorhabditis elegans*, and computational predictions among other metazoans, we found that exons are differentially marked from introns both in terms of nucleosome occupancy and in terms of specific histone modifications. Moreover, our results indicate that the basis for this different chromatin landscape lies within the DNA sequence itself. This implies that splicing signals, which were previously thought to act only at the RNA level, may also be meaningful at the DNA level and may mediate the observed differences in the chromatin landscapes of exons and introns.

RESULTS

We hypothesized that chromatin organization might differentially mark exon and intron units. Two observations supported this hypothesis: first, DNA fragments wrapped around histone octamers are

[1]Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel-Aviv University, Ramat Aviv, Israel. [2]Department of Genetics, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem, Israel. Correspondence should be addressed to G.A. (gilast@post.tau.ac.il).

Received 6 May; accepted 21 July; published online 16 August 2009; doi:10.1038/nsmb.1659

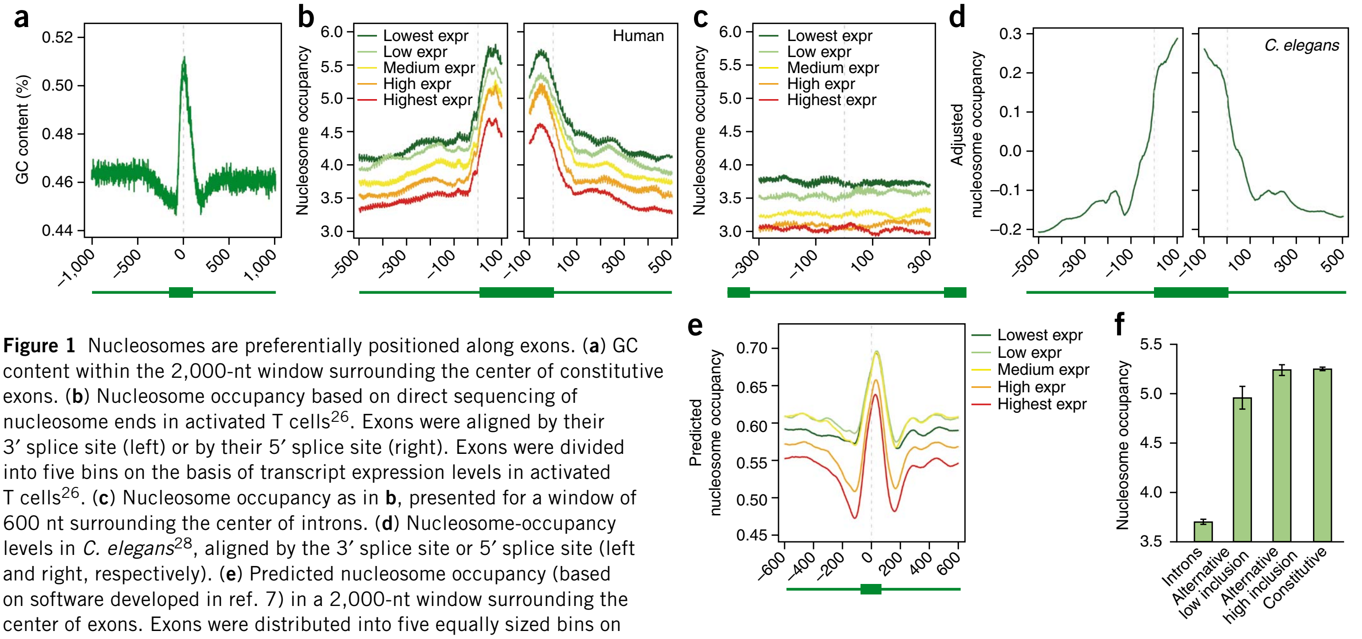

Figure 1 Nucleosomes are preferentially positioned along exons. (**a**) GC content within the 2,000-nt window surrounding the center of constitutive exons. (**b**) Nucleosome occupancy based on direct sequencing of nucleosome ends in activated T cells[26]. Exons were aligned by their 3′ splice site (left) or by their 5′ splice site (right). Exons were divided into five bins on the basis of transcript expression levels in activated T cells[26]. (**c**) Nucleosome occupancy as in **b**, presented for a window of 600 nt surrounding the center of introns. (**d**) Nucleosome-occupancy levels in *C. elegans*[28], aligned by the 3′ splice site or 5′ splice site (left and right, respectively). (**e**) Predicted nucleosome occupancy (based on software developed in ref. 7) in a 2,000-nt window surrounding the center of exons. Exons were distributed into five equally sized bins on the basis of expression levels as in **b**. (**f**) Mean nucleosome occupancy in introns, alternatively spliced exons included in less than 50% of transcripts, alternatively spliced exons included in at least 50% of transcripts and constitutively spliced exons. Error bars represent the s.e.m.

147 nt in length, which is approximately the length of an average internal exon[5]. Second, when we aligned a data set of constitutively spliced internal exons by their centers and examined the frequency of guanines and cytosines in each of the 2,000 nt surrounding this center (the GC content), we observed that the GC content in exons was significantly higher than in the flanking introns (*t*-test, $P \approx 0$; **Fig. 1a**). As nucleosome positioning is directed to a great extent by the DNA sequence[7], we suspected that differences in GC content could differentially modulate nucleosome assembly in exons and introns.

Nucleosomes preferentially bind to exons rather than introns

To assess whether nucleosomes are differentially distributed across exons and introns, we used the table of RefSeq genes to generate data sets of 4,570 human alternatively spliced internal exons, 69,580 constitutively spliced internal exons and 37,996 introns. We next obtained a data set of nucleosome positioning within the human genome derived through Solexa high-throughput sequencing of DNA fragments attached to nucleosomes in activated T cells, following micrococcal nuclease (MNase) digestion[26]. We divided all exons into five equally sized bins on the basis of their expression levels, which we derived from ref. 26, and examined the mean nucleosome occupancy levels around the 3′ splice site and 5′ splice site. We observed a distinct peak of nucleosome occupancy within exons (**Fig. 1b**) but not within introns (**Fig. 1c**). This phenomenon was also observed across different individual loci (**Supplementary Fig. 1**), although not all loci showed such behavior. **Figure 1b** also revealed an inverse correlation between gene expression levels and nucleosome occupancy, consistent with the observations that nucleosomes are depleted in actively transcribed regions[27].

To verify these findings, we performed the following analyses. (i) We repeated our analysis with a high-resolution nucleosome position map from a mixed-stage, mixed-tissue population of *C. elegans* cells, based on SOLiD parallel sequencing[28], which we mapped against a data set of 89,343 exons from *C. elegans* (**Fig. 1d**). (ii) We repeated this analysis with data from resting T cells[26] (**Supplementary Fig. 2a**). (iii) To address the possibility that coding restraints were obfuscating our observations, we repeated the analysis with noncoding exons (**Supplementary Fig. 2b**). (iv) To address possible sequencing bias resulting from GC content[29,30], we analyzed a Solexa high-throughput sequence of sheared DNA from Jurkat cells[31] (**Supplementary Fig. 2c**). (v) We analyzed data of MNase-digested chromatin from *Drosophila melanogaster* hybridized to tiling arrays[32] (**Supplementary Fig. 3a**). (vi) Last, we applied the nucleosome-prediction software developed by the Segal laboratory[7], providing nucleosome-occupancy predictions based solely on DNA sequences, to the regions surrounding the centers of the human exons and introns (**Fig. 1e**). These analyses confirmed our observations that there is increased nucleosome occupancy along exons, ruled out the possibility that GC content or protein-coding regions were confounding or biasing our observations and demonstrated that this phenomenon is coded directly within the genomic sequence (see also **Supplementary Methods**).

Our conclusions were further supported by the observation that nucleosome-occupancy levels in humans correlate with inclusion levels. We observed modest but significant increases (Kruskal-Wallis rank sum test, $P < 2.2 \times 10^{-16}$) in nucleosome occupancy from introns, to alternatively spliced exons included in less than 50% of transcripts (based on EST data; see Online Methods), to alternatively spliced exons included in at least 50% of transcripts, to constitutively spliced exons (**Fig. 1f**).

The splicing code and the chromatin code overlap

In a recent study, all 1,024 possible pentamers were scored on the basis of their empirically observed tendency to be covered by a nucleosome[33]. From this pentamer scoring table we extracted 248 pentamers that disfavored nucleosome binding. We then determined the distribution of these 248 pentamers within the 600-nt regions that flanked the exons (**Fig. 2a**). We found that pentamers that disfavored binding of the nucleosome were depleted within exons with respect to introns. This was not unexpected, as many nucleosome-disfavoring pentamers are AT-rich, and introns are AT-rich in comparison to exons (**Fig. 1a**)[33]. More notably, we

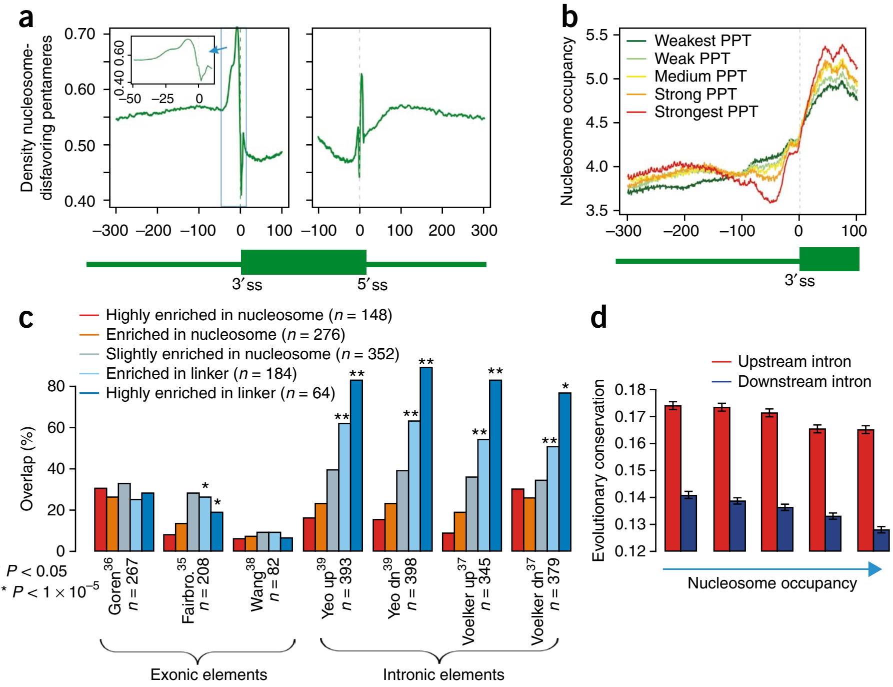

Figure 2 Overlap between chromatin code and splicing code. (**a**) Distribution of nucleosome-disfavoring sequences identified (ref. 33) within the 600-nt region surrounding human constitutively spliced exons aligned at the 3′ splice site (3′ ss, left) or at the 5′ splice site (5′ ss, right). The ordinate depicts the fraction of exons in which a given position is overlapped by a nucleosome-disfavoring sequence. (**b**) Nucleosome-occupancy levels in activated T cells within 300 nt upstream and 100 nt downstream of the 3′ ss. Introns were divided into five bins on the basis of the strength of their PPT. (**c**) Overlap between nucleosome-favoring or nucleosome-disfavoring sequences and between different groups of splicing regulatory sequences. The 1,024 possible pentamers were divided into five bins on the basis of their nucleosome-favoring or nucleosome-disfavoring score[33]. For each group of splicing regulatory sequences (labeled by the first author of the relevant publication and referencing that publication), the fraction of sequences overlapping sequences in each nucleosome-favoring or nucleosome-disfavoring bin was calculated. Levels of significance are indicated by one or two asterisks, indicating hypergeometric *P*-values of $P < 0.05$ or $P < 1 \times 10^{-5}$, respectively. Significant values of this test for a given nucleosome-favoring or nucleosome-disfavoring bin indicate that the overlap between the nucleosome-favoring or nucleosome-disfavoring pentamers in that bin and a given set of SREs is significantly greater than expected by chance. 'up' and 'dn' refer to the data sets of k-mers found to be enriched upstream or downstream of exons, respectively. (**d**) Mean conservation levels, based on phastCons scores for 18 placental organisms, within the 50 intronic nucleotides preceding and following exons. Exons were divided into five groups on the basis of the mean nucleosome-occupancy levels in activated T cells within the respective intronic regions.

observed a peak in nucleosome-disfavoring elements at both ends of the exons, consistent with the two valleys flanking the central peak in **Figure 1a**,**e**. The peak at the 5′ splice site is narrow and represents the specific nucleotide composition of the 5′ splice site. We concentrated on the broader peak observed within the ~30 nt upstream of the 3′ splice site (**Fig. 2a**, inset; see **Supplementary Fig. 3b** for the region surrounding the 5′ splice site), in the region harboring the polypyrimidine tract (PPT), one of the core splicing signals at the 3′ end of introns. To assess the role of the PPT in modulating nucleosome occupancy, we aligned all exons by their 3′ splice site and divided them into five bins on the basis of the strength of the PPT, scored as described previously[34]. Analysis of nucleosome-occupancy levels within these groups revealed that stronger PPTs are linked with decreased nucleosome occupancy within the intronic regions immediately preceding exons, but with increased nucleosome occupancy within exons (**Fig. 2b**). Thus, whereas at the RNA level the PPT functions in mRNA splicing, at the DNA level it serves to discriminate between exons and introns in terms of nucleosome occupancy.

To gain an increased understanding of the relationship between the splicing code and the chromatin code, we analyzed the extent to which splicing regulatory elements overlap with nucleosome favoring and disfavoring pentamers. The set of 1,024 pentamers[33] was divided into five bins on the basis of nucleosome disfavoring/favoring ratios. We next determined the extent and significance of overlap between the sequences in each of these bins with seven data sets of intronic and exonic splicing regulatory elements (ISRs and ESRs)[35–39] (see Online Methods). We found that ISRs, both upstream and downstream of exons, tended to be significantly and highly enriched in nucleosome-disfavoring sequences (**Fig. 2c**). This was not the case for ESRs. This indicates that one role of ISRs, which were originally identified on the basis of their overabundance and high conservation within intronic regions adjacent to exons, may be to control the exon-intron nucleosome-occupancy gradient. Supporting this hypothesis, we found an inverse correlation between nucleosome occupancy and conservation in the 50 nt within introns that immediately precede and follow an exon (**Fig. 2d**), possibly indicative of evolutionary pressure to maintain nucleosome-free regions at both ends of the exons.

Post-translational modifications enriched along exons

The fact that exons tended to be occupied by nucleosomes raised the possibility that specific modifications of histones may mark exons as well. To assess this, we analyzed genome-wide ChIP-seq data sets containing data on sequences bound by histones with 38 modifications in human activated T cells[40,41]. We used site identification from short sequence reads (SISSRs)[42] to identify genomic regions enriched with each modification (**Supplementary Table 1**). We assessed the prevalence of every enriched region across a 2,000-nt window surrounding the center of exons, after dividing the exons into five equally sized bins on the basis of the expression levels of transcripts in activated T cells[40]. We performed identical analyses on the 2,000 nt surrounding the center of introns and on a data set of 11,473 promoters with sequences aligned by the transcription start site (TSS). Four post-translational histone modifications presented peaks within exons: the most prominent peak was observed for trimethylation of H3K36 (H3K36me3), and somewhat less prominent peaks were observed for monomethylations of H3K79, H4K20 and H2BK5 (**Fig. 3a**). Peaks for each of these four modifications increased in amplitude with increasing gene expression levels (**Fig. 3a**). The most salient peak was found for H3K36me3 (**Fig. 3a**, above right), consistent with recently reported results performed independently of ours[43]. Additional analyses of this modification in the context of alternative and constitutive exons, and when comparing between different tissues and organisms, were also congruent with ref. 43 (**Supplementary Results** and **Supplementary**

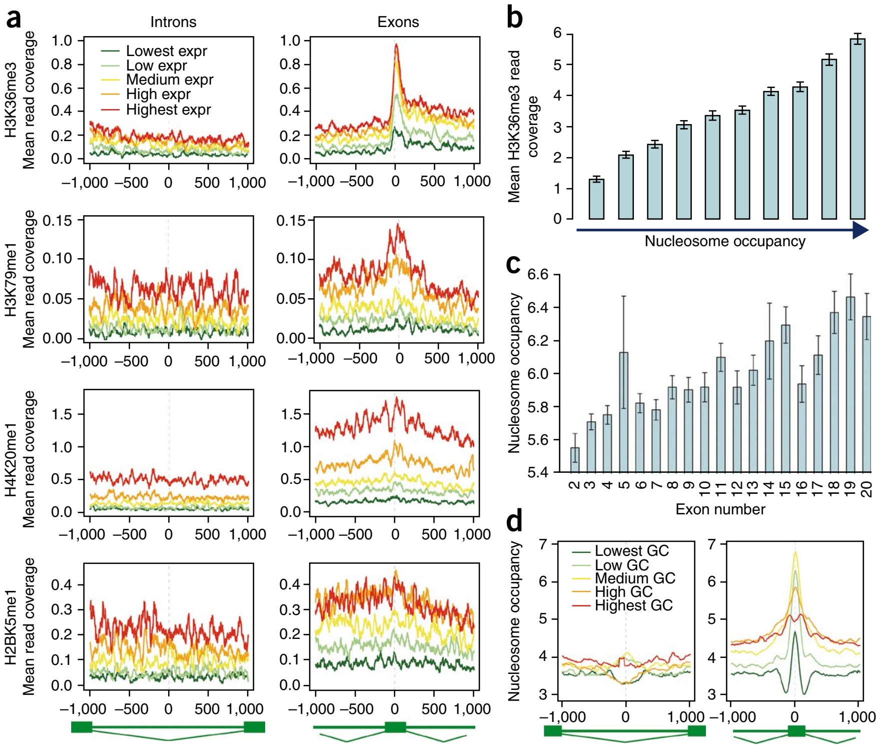

Figure 3 Post-translational histone modifications occurring along exons and analysis of factors correlating with nucleosome occupancy. (**a**) Profiles for binding of H3K36me3, H3K79me1, H2BK5me1 and H4K20me1, across 2,000-nt windows surrounding the center of introns (left) and constitutively spliced exons (right). The sequences were divided into five equally sized bins on the basis of transcript expression levels (derived as in ref. 40). (**b**) Correlation between H3K36me3 levels and nucleosome occupancy within constitutively spliced exons. All exons were divided into ten bins of gradually increasing nucleosome occupancy (based on ref. 26), and mean H3K36me3 levels were calculated for each bin. Error bars represent the s.e.m. (**c**) Nucleosome occupancy as a function of location within resting T cells. Nucleosome occupancies were calculated as the mean nucleosome occupancies within exons[26]. Error bars represent the s.e.m. (**d**) Nucleosome occupancy in activated T cells in a 2,000-nt window surrounding the center of introns and constitutively spliced exons. Exons and introns were distributed into five equally sized groups on the basis of GC content within a 400-nt window surrounding the center of the exon or intron.

Figs. 4 and **5**). Thus, our results are consistent with this report[43], but our analyses also highlight two important points. First, we demonstrate that there are additional modifications that show similar patterns. Second, our findings suggest that H3K36me3 modifications of exons are the consequence of a more fundamental phenomenon, namely the increased nucleosome occupancy along exons. Consistent with this hypothesis, we observed a clear correlation between nucleosome occupancy levels along exons and H3K36me3 modification levels (**Fig. 3b**). This correlation was also apparent when we examined individual loci (**Supplementary Fig. 1**).

Increased levels of nucleosome coverage in 3′ exons

Previous studies have reported that binding of H3K36me3 increases toward the 3′ end of genes[40,44,45]. Consistently, our data sets indicate that H3K36me3 is found more often in downstream exons than in more 5′ exons (**Supplementary Fig. 6a**), whereas other modifications show the opposite trend and peak at the 5′ end of genes (**Supplementary Fig. 6b–h**). This led us to examine whether the same is true for nucleosome-occupancy levels. Indeed, this was the case in resting T cells (**Fig. 3c**), as well as in activated T cells and in the *C. elegans* data set (data not shown). This phenomenon may reflect the fact that transcription events initiate from the 5′ end of genes but may not be completed[46], resulting in nucleosome depletion at the 5′ end of genes.

Effect of GC composition

In light of the different GC composition of exons and introns (**Fig. 1a**), we set out to determine how GC content affects nucleosome occupancy. Exons and introns were distributed into five bins of gradually increasing GC content, based on a 400-nt window surrounding the center of the exons and introns. We found that nucleosome occupancy within exons is highly correlated with GC content, with the highest occupancy levels in cases of intermediate GC content of 41–57% (**Fig. 3d**, right). Within introns, no such relationship was observed (**Fig. 3d**, left), confirming that a sequencing bias for GC-rich regions[29,30] does not account for the nucleosome-occupancy peak observed within exons. Levels of histone modifications also changed as a function of GC content (**Supplementary Fig. 7a–d**), and GC content was a more informative measure than expression for predicting whether a modified histone binds (**Supplementary Fig. 7e**). Thus, the differential GC composition between exons and introns seems to be one of the driving forces behind the differential levels of nucleosome occupancy and histone modifications between exons and introns.

Levels of binding of RNAPII is higher in exons than introns

The factor thought to be responsible for cross-talk between the chromatin structure of DNA and the exon-intron architecture of RNA is RNAPII, which is linked to both transcription and splicing. We noted that binding levels of RNAPII are increased in exons, compared to introns, across all levels of expression (**Supplementary Fig. 8a**), consistent with previous findings[47]. This may suggest that nucleosomes bound within exons and introns could serve as ‘speed bumps’ that slow the rate of RNAPII, thereby improving selection of exons[13,17,48]. To further examine the interplay between expression and nucleosome occupancy, we examined the dynamics of exonic nucleosome occupancy levels in activated and inactivated T cells as a function of changes in gene expression levels between these two conditions. We found that decreased nucleosome occupancy in a given condition correlated with increased expression and vice versa (**Supplementary Fig. 8b**). This demonstrates that nucleosome occupancy levels are dynamically altered and that changes in these levels are linked with changes in expression levels.

Exons harbor nucleosomes throughout metazoan evolution

Finally, we sought to determine whether preferential nucleosome occupancy of exons is conserved throughout evolution. We assembled data sets of internal exons from seven organisms across the metazoan

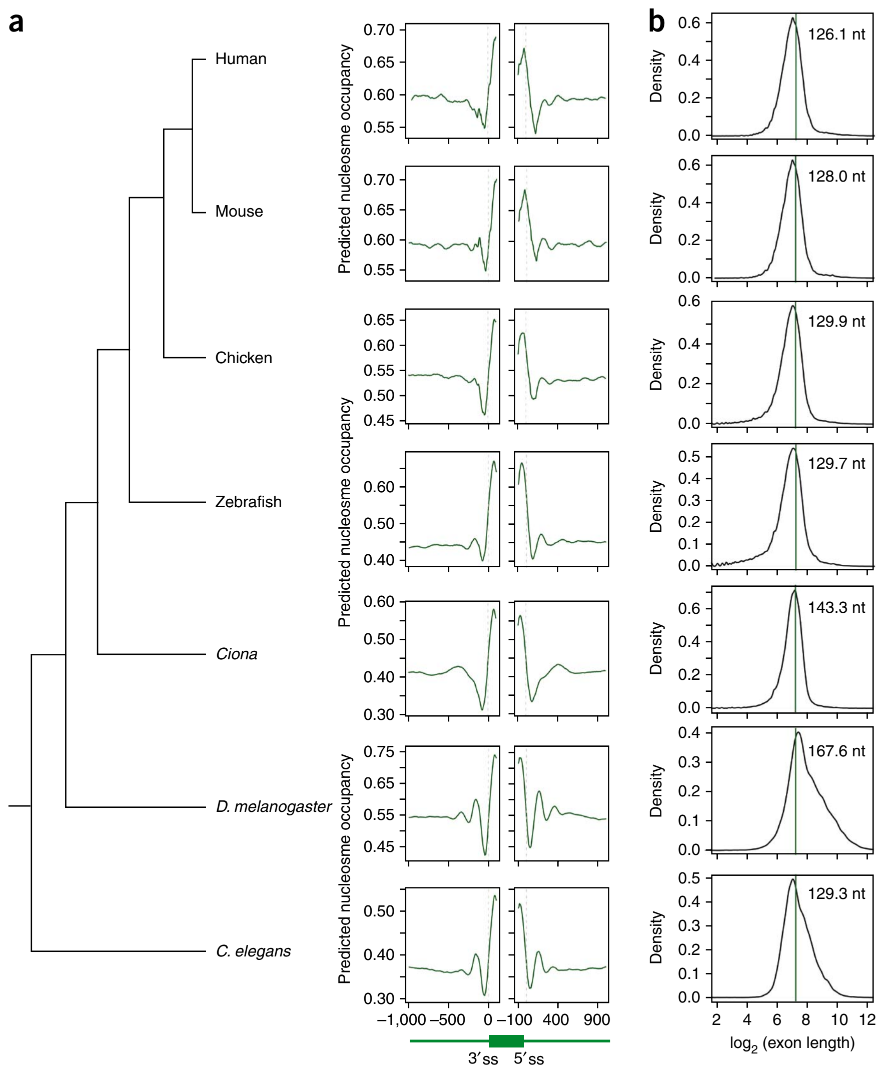

Figure 4 Positioning of nucleosomes along exons is a conserved phenomenon throughout the metazoan evolution. (**a**) Nucleosome occupancy along 10,000 randomly sampled internal exons from each organism was predicted (based on ref. 7). Results are shown for a 2,000-nt window surrounding the 3′ splice site (3′ ss, left) and the 5′ splice site (5′ ss, right). (**b**) Distributions of exon lengths among the different organisms. Density estimates of the distributions of log2-transformed exon lengths were calculated using the 'density()' function in R (http://www.r-project.org/). A vertical line indicating the position of 147 nt was inserted to allow visual comparison between the length of DNA wrapped around a mononucleosome and the peak of exon length distributions. The precise peak of the exon length distribution is indicated in the upper-right corner of each chart.

tree. For each organism, we used the Segal laboratory's software[7] to predict the nucleosome occupancies across a 2,000-nt window centered around the 3′ splice site and the 5′ splice site of 10,000 randomly sampled internal exons. Among all analyzed organisms, exons harbored a well-positioned nucleosome (**Fig. 4a**). In parallel, we examined the length distributions of the internal exons among the different organisms. We found that the length distributions of exons among all analyzed metazoans peak between 125 nt and 165 nt (**Fig. 4b**), in marked correspondence with the 147 nt of DNA that is wrapped around a mononucleosome.

DISCUSSION

There is a long-standing question about the ability of the splicing machinery to select short exons within vast intronic oceans[1–4]. Our results suggest that marking of exons by nucleosomes may have a role in defining the exon-intron architecture of a gene. Thus, the tendency to be occupied by mononucleosomes may be one of the forces that acts on exons to keep their length within their observed range. This does not rule out an additional evolutionary force believed to act on metazoan exons, namely, the requirement that spliceosomal proteins that bind both ends of the exons must physically interact with each other, in a process termed 'exon recognition'[1,49,50].

Nucleosome positioning at the DNA level may affect exon recognition at the RNA level through at least two mechanisms. The first possibility is that the nucleosomes function as 'speed bumps' to slow the rate of RNAPII elongation. A reduction in transcription rate has been shown to increase inclusion of alternatively spliced exons[17]. A second possibility is that the preferential positioning of nucleosomes along exons marks the exons with specifically modified histones that subsequently interact with the splicing machinery to enhance recognition of exons. The fact that both trimethylated H3K36 and, to a lesser extent, monomethylated H4K20 were enriched within exons is notable because these two histone modifications were previously reported as marks of transcription elongation[44,45,51,52]. Transcription elongation is tightly coupled to splicing[13,18,48]. The H3K36me3 modification is mediated by Setd2 (ref. 53), which is recruited to the phosphorylated CTD of the elongating RNAPII[45]; the CTD is also associated with different splicing factors[18]. Thus, H3K36me3-modified nucleosomes, which preferentially bind within exons, may serve as a scaffold for recruiting different splicing factors[13]. A nonmutually exclusive possibility is that nucleosomes confer protection to the exonic sequences coiled around them[54]. Finally, it cannot be ruled out that the link between nucleosome occupancy and exon-intron architecture is indirect, and mediated by sequence compositions of exons and introns optimized to allow, for example, binding of SR proteins. However, in light of emerging data showing the temporal and spatial link between RNA processing and chromatin structure, we consider such a purely circumstantial link unlikely.

Previous findings that DNA length is synchronized between successive splice sites[55] and that splice sites have a tendency to reside within a few base pairs from the nucleosome dyad axis[54] were suggested to reflect chromatin structure. On the basis of much broader experimental data sets from various organisms, our analysis now demonstrates that exon-intron architecture is reflected in the chromatin structure of genes. Moreover, we show that exons rather than introns tend to contain nucleosomes and that splice sites (especially the 3′ splice site) contain nucleosome-disfavoring sequences and may thus shift nucleosome occupancy to exons. Our results, in conjunction with those of Tilgner *et al.*[56], add an important layer to our gradually accumulating understanding of the coupling between chromatin structure, transcription and splicing.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank D. Schones (Laboratory of Molecular Immunology, US National Heart, Lung and Blood Institute) and Y. Field (Weizmann Institute of Science) for sharing data sets, E. Segal (Weizmann Institute of Science) and M. Kupiec (Tel Aviv University) for critical review of the manuscript, and B. Zornberg for inspiration. S.S. is a fellow of the Edmond J. Safra bioinformatic program at Tel Aviv University. G.A. is supported by a grant from the Israel Science Foundation (ISF 61/09), Joint Germany-Israeli Research Program (ca-139), Deutsche-Israel Project (DIP MI-1317), and European Alternative Splicing Network (EURASNET). E.M. is supported by the Israel Science Foundation (ISF 215/07), the European Union (IRG-206872) and an Alon fellowship. This work was performed in partial fulfillment of the requirements for a Ph.D. degree of S.S., Sackler Faculty of Medicine, Tel Aviv University.

AUTHOR CONTRIBUTION

S.S., E.M. and G.A. conceived the analyses; S.S. designed and conducted all analyses in the manuscript; S.S. prepared the manuscript, aided by E.M. and G.A.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Berget, S.M. Exon recognition in vertebrate splicing. *J. Biol. Chem.* **270**, 2411–2414 (1995).
2. Graveley, B.R. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* **17**, 100–107 (2001).
3. Smith, C.W. & Valcarcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* **25**, 381–388 (2000).
4. Black, D.L. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**, 291–336 (2003).
5. Wang, Z. & Burge, C.B. Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**, 802–813 (2008).
6. Segal, E. *et al.* A genomic code for nucleosome positioning. *Nature* **442**, 772–778 (2006).
7. Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362–366 (2008).
8. Jenuwein, T. & Allis, C.D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
9. Bernstein, B.E., Meissner, A. & Lander, E.S. The mammalian epigenome. *Cell* **128**, 669–681 (2007).
10. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
11. Jones, P.A. & Baylin, S.B. The epigenomics of cancer. *Cell* **128**, 683–692 (2007).
12. de Almeida, S.F. & Carmo-Fonseca, M. The CTD role in cotranscriptional RNA processing and surveillance. *FEBS Lett.* **582**, 1971–1976 (2008).
13. Allemand, E., Batsche, E. & Muchardt, C. Splicing, transcription, and chromatin: a *ménage à trois*. *Curr. Opin. Genet. Dev.* **18**, 145–151 (2008).
14. Moore, M.J. & Proudfoot, N.J. Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136**, 688–700 (2009).
15. Howe, K.J. RNA polymerase II conducts a symphony of pre-mRNA processing activities. *Biochim. Biophys. Acta* **1577**, 308–324 (2002).
16. Goldstrohm, A.C., Albrecht, T.R., Sune, C., Bedford, M.T. & Garcia-Blanco, M.A. The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol. Cell. Biol.* **21**, 7617–7628 (2001).
17. de la Mata, M. *et al.* A slow RNA polymerase II affects alternative splicing *in vivo*. *Mol. Cell* **12**, 525–532 (2003).
18. Kornblihtt, A.R. Chromatin, transcript elongation and alternative splicing. *Nat. Struct. Mol. Biol.* **13**, 5–7 (2006).
19. Schor, I.E., Rascovan, N., Pelisch, F., Allo, M. & Kornblihtt, A.R. Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc. Natl. Acad. Sci. USA* **106**, 4325–4330 (2009).
20. Nogues, G., Kadener, S., Cramer, P., Bentley, D. & Kornblihtt, A.R. Transcriptional activators differ in their abilities to control alternative splicing. *J. Biol. Chem.* **277**, 43110–43114 (2002).
21. Batsché, E., Yaniv, M. & Muchardt, C. The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat. Struct. Mol. Biol.* **13**, 22–29 (2006).
22. Kress, T.L., Krogan, N.J. & Guthrie, C. A single SR-like protein, Npl3, promotes pre-mRNA splicing in budding yeast. *Mol. Cell* **32**, 727–734 (2008).
23. Loomis, R.J. *et al.* Chromatin binding of SRp20 and ASF/SF2 and dissociation from mitotic chromosomes is modulated by histone H3 serine 10 phosphorylation. *Mol. Cell* **33**, 450–461 (2009).
24. Sims, R.J. III *et al.* Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol. Cell* **28**, 665–676 (2007).
25. Jobert, L. *et al.* Human U1 snRNA forms a new chromatin-associated snRNP with TAF15. *EMBO Rep.* **10**, 494–500 (2009).
26. Schones, D.E. *et al.* Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887–898 (2008).
27. Lee, C.K., Shibata, Y., Rao, B., Strahl, B.D. & Lieb, J.D. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* **36**, 900–905 (2004).
28. Valouev, A. *et al.* A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res.* **18**, 1051–1063 (2008).
29. Dohm, J.C., Lottaz, C., Borodina, T. & Himmelbauer, H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.* **36**, e105 (2008).
30. Hillier, L.W. *et al.* Whole-genome sequencing and variant discovery in *C. elegans*. *Nat. Methods* **5**, 183–188 (2008).
31. Valouev, A. *et al.* Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat. Methods* **5**, 829–834 (2008).
32. Mavrich, T.N. *et al.* Nucleosome organization in the *Drosophila* genome. *Nature* **453**, 358–362 (2008).
33. Field, Y. *et al.* Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLOS Comput. Biol.* **4**, e1000216 (2008).
34. Schwartz, S.H. *et al.* Large-scale comparative analysis of splicing signals and their corresponding splicing factors in eukaryotes. *Genome Res.* **18**, 88–103 (2008).
35. Fairbrother, W.G., Yeh, R.F., Sharp, P.A. & Burge, C.B. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**, 1007–1013 (2002).
36. Goren, A. *et al.* Comparative analysis identifies exonic splicing regulatory sequences—the complex definition of enhancers and silencers. *Mol. Cell* **22**, 769–781 (2006).
37. Voelker, R.B. & Berglund, J.A. A comprehensive computational characterization of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing. *Genome Res.* **17**, 1023–1033 (2007).
38. Wang, Z. *et al.* Systematic identification and analysis of exonic splicing silencers. *Cell* **119**, 831–845 (2004).
39. Yeo, G.W., Van Nostrand, E.L. & Liang, T.Y. Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. *PLoS Genet.* **3**, e85 (2007).
40. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
41. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903 (2008).
42. Jothi, R., Cuddapah, S., Barski, A., Cui, K. & Zhao, K. Genome-wide identification of *in vivo* protein-DNA binding sites from ChIP-Seq data. *Nucleic Acids Res.* **36**, 5221–5231 (2008).
43. Kolasinska-Zwierz, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* **41**, 376–381 (2009).
44. Bell, O. *et al.* Localized H3K36 methylation states define histone H4K16 acetylation during transcriptional elongation in *Drosophila*. *EMBO J.* **26**, 4974–4984 (2007).
45. Du, H.N., Fingerman, I.M. & Briggs, S.D. Histone H3 K36 methylation is mediated by a *trans*-histone methylation pathway involving an interaction between Set2 and histone H4. *Genes Dev.* **22**, 2786–2798 (2008).
46. Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R. & Young, R.A. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88 (2007).
47. Brodsky, A.S. *et al.* Genomic mapping of RNA polymerase II reveals sites of co-transcriptional regulation in human cells. *Genome Biol.* **6**, R64 (2005).
48. Kornblihtt, A.R. Coupling transcription and alternative splicing. *Adv. Exp. Med. Biol.* **623**, 175–189 (2007).
49. Ram, O. & Ast, G. SR proteins: a foot on the exon before the transition from intron to exon definition. *Trends Genet.* **23**, 5–7 (2007).
50. Robberson, B.L., Cote, G.J. & Berget, S.M. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10**, 84–94 (1990).
51. Talasz, H., Lindner, H.H., Sarg, B. & Helliger, W. Histone H4-lysine 20 monomethylation is increased in promoter and coding regions of active genes and correlates with hyperacetylation. *J. Biol. Chem.* **280**, 38814–38822 (2005).
52. Vakoc, C.R., Sachdeva, M.M., Wang, H. & Blobel, G.A. Profile of histone lysine methylation across transcribed mammalian chromatin. *Mol. Cell. Biol.* **26**, 9185–9195 (2006).
53. Edmunds, J.W., Mahadevan, L.C. & Clayton, A.L. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* **27**, 406–420 (2008).
54. Kogan, S. & Trifonov, E.N. Gene splice sites correlate with nucleosome positions. *Gene* **352**, 57–62 (2005).
55. Beckmann, J.S. & Trifonov, E.N. Splice junctions follow a 205-base ladder. *Proc. Natl. Acad. Sci. USA* **88**, 2380–2383 (1991).
56. Tilgner, H. *et al.* Nucleosome positioning as a determinant of exon recognition. *Nat. Struct. Mol. Biol.* advance online publication, doi:10.1038/nsmb.1658 (16 August 2009).

ONLINE METHODS

Data set of exons, introns and promoters in human and mouse. We downloaded coordinates of human (hg18) and mouse (mm9) exons based on the Refseq track and coordinates of spliced EST alignments from the UCSC genome browser (http://genome.ucsc.edu/). We established the inclusion levels of exons on the basis of alignment of ESTs to the exons and flanking introns. We extracted transcription start site annotations using the RefSeq track. All introns flanking the constitutively spliced exons were used to form the intronic data set. Exons longer than 300 nt and introns shorter than 2,000 nt were discarded to avoid contamination of intronic regions with exonic ones in the different profile plots.

Calculation of inclusion levels. Inclusion levels of exons were derived on the basis of EST data, which was downloaded from the Spliced EST table using the UCSC table browser. For an EST to support exon inclusion, we demanded that at least 50 nt of the exon and either of its two splicing signals form part of the alignment. Alignment gaps of less than 8 nt were ignored, as in the UCSC visualization defaults. An EST was defined as supporting exon skipping if there was no alignment between the EST and the exon and if the EST supported the joining of the two flanking exons. We defined all exons whose inclusion was supported by at least 15 ESTs and lacking any ESTs supporting exon skipping as constitutively spliced exons, whereas all exons with at least two ESTs supporting inclusion, two ESTs supporting skipping and an inclusion level greater than 0.05 and lower than 0.95 were defined as alternatively spliced. Inclusion levels were calculated as $EST_{inc}/(EST_{skip}+EST_{inc})$, where EST_{skip} and EST_{inc} represent the number of ESTs supporting skipping and inclusion of the exon, respectively.

Nucleosome occupancy. We obtained nucleosome-occupancy levels by direct sequencing of nucleosome ends in human resting and activated T cells using nucleosome score profiles generated by ref. 26 and downloaded from the Dynamic Regulation of Nucleosome Positioning in the Human Genome site (http://dir.nhlbi.nih.gov/papers/lmi/epigenomes/hgtcellnucleosomes.aspx). We calculated these scores by assigning each 10-nt sliding window the sum of the reads mapping to the sense strand 80 nt upstream to the window and to the antisense strands 80 nt downstream of the window. Gene expression levels for resting and activated T cells were also obtained from ref. 26. For each exon, mean nucleosome-occupancy levels for the exon, and for the 50 intronic nucleotides immediately preceding and following it, were calculated. We obtained nucleosome-occupancy levels in *C. elegans* via the 'adjusted nucleosome coverage' scores generated in ref. 28. These scores are calculated as $[(1+n)/N] / [(1+c)/C]$, where n and c are the numbers of putative 147-bp cores covering each base pair from nucleosome and control data, and N and C are the total number of nucleosome and control reads obtained by SOLiD sequencing[28]. The control reads represent reads that were obtained following light digestions of MNase and that were size selected between 400–900 nt[28]. To predict nucleosome occupancy of exons and introns, we applied the software developed by the Segal laboratory[7] to a region of 5,000 nt surrounding the center of exons, introns and transcription starts sites, with default parameters.

ChIP-seq data sets. ChIP-seq reads pertaining to 38 histone modifications in human activated T cells, as well as ChIP-seq data for the histone variant H2A.Z, for RNAPII and for CTCF were obtained from refs. 40–41. ChIP-seq reads for H3K36me3 in mouse embryonic stem cells, neural progenitors and mouse embryonic fibroblasts were obtained from ref. 57. We also obtained gene expression data from these publications. Enriched genomic regions were identified using SISSRs version 1.2 (ref. 42) with default parameters, a false detection rate of 0.001 and binding sites reported as a 40-nt window centered on the inferred binding point. Exons, introns and promoters were defined as bound by a given modified protein if any 40-nt modification window overlapped a 400-nt window surrounding the center of the exon, intron or the transcription start site, respectively. The GC content used in the statistics of **Figures 1** and **3d** and **Supplementary Figure 7** was also calculated based on this 400-nt window. We disregarded modifications that were present in fewer than 700 constitutive exons (**Supplementary Table 1**), because the signal-to-noise ratio within these data sets was low.

Overlap between ESRs and ISRs and nucleosome-favoring and nucleosome-disfavoring elements. We defined disfavoring/favoring ratios above 1 as enriched in linker and those above 1.5 as highly enriched in linker. Conversely, we defined favoring/disfavoring ratios above 1 as slightly enriched in nucleosomes, above 1.5 as enriched in nucleosomes and above 2 as highly enriched in nucleosomes. Data sets of splicing regulatory sequences were obtained from refs. 35–39. Because these data sets contain sequences ranging from 4 nt to 8 nt, we first discarded all sequences shorter than 5 nt and then broke down each k-mer with $k > 5$ into the $k - 4$ consecutive pentamers within it. Finally, we assessed the percentage of pentamers in each ESR/ISR group shared by each nucleosome-favoring (or nucleosome-disfavoring) pentamer. The significance of this overlap was determined by hypergeometric tests.

Conservation. We measured conservation for specific positions within the genome on the basis of phastCons scores for 18 placental organisms, which were downloaded from UCSC (http://hgdownload.cse.ucsc.edu/goldenPath/hg18/phastCons28way/placental/). For each exon, mean conservation levels for the 50 intronic nucleotides immediately preceding and immediately following the exon, as well as for the entire exon, were calculated.

Conservation of nucleosome occupancy along exons throughout metazoan evolution. We downloaded gene structure tables for the various organisms from the UCSC table browser (http://genome.ucsc.edu/cgi-bin/hgTables), as specified in **Supplementary Table 2**. The tables were parsed to extract introns and exons. Redundant exons were filtered out on the basis of their genomic coordinates. First and last exons were discarded. The number of exons extracted for each organism, as well as the mean lengths of the exons and introns within each organism, are specified in **Supplementary Table 2**. For the analysis presented in **Figure 6**, we randomly sampled 10,000 exons from each organism and extracted the 5,000-nt regions surrounding their 5′ splice site and 3′ splice site to allow prediction of nucleosome occupancy by the Segal software[7], as described above.

57. Mikkelsen, T.S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).

doi:10.1038/nsmb.1659