The interleaved genome

Jane Mellor, Ronja Woloszczuk and Françoise Howe.

Department of Biochemistry, South Parks Road, Oxford, OX1 3QU, UK

Eukaryotic genomes are pervasively transcribed but until recently this non-coding transcription was considered to be simply noise. Non-coding transcription units overlap with genes and genes overlap other genes, meaning genomes are extensively interleaved. Experimental interventions reveal high degrees of interdependency between these transcription units, which have been co-opted as gene regulatory mechanisms. The precise outcome depends on the relative orientation of the transcription units and whether two overlapping transcription events are contemporaneous or not, but generally involves chromatin-based changes. Thus transcription itself regulates transcription initiation or repression at many regions of the genome.

Eukaryotic genomes are pervasively transcribed resulting in an interleaved genome

Transcription is the process leading to the production of RNA transcripts. **RNA-seq** (see Glossary) captures the transcriptional output of the genome, known as the transcriptome, which represents the stable transcripts in a particular cell type. The transcriptome shows a great deal of heterogeneity [1, 2] and contains many uncharacterised non-coding transcripts [3], such as the **stable unannotated transcripts** (SUTs) [4] in *Saccharomyces cerevisiae* and the **long non-coding RNA** (lncRNAs) in other genomes [5]. Some of these transcripts are antisense to conventional RNA polymerase II **transcription units** (TUs) and are known as **natural antisense transcripts** (NATs) [6-10].

New methods, such **nascent elongating transcript sequencing**, NET-seq [11-13] and **precision run-on sequencing**, PRO-seq [14, 15], provide a quantitative readout of RNA polymerase II (RNAPII) activity across the genome at base-pair resolution (Box 1). These techniques reveal a large number of transcription events leading to the production of unstable transcripts, such as the **cryptic unstable transcripts** (CUTs) and **Xrn1-sensitive unannotated transcripts** (XUTs) in *S. cerevisiae* [9, 16], that are not typically represented in transcriptome annotations. These transcription events may originate in **intergenic regions** [12, 13, 17] and as divergent transcription events from enhancers [18] and promoters [4, 12-14, 16]. In addition, these methods uncover a great deal of overlapping transcription. This can be antisense to conventional RNA polymerase II TUs; it can span intergenic regions between

tandem genes, giving rise to **bi-cistronic or tri-cistronic transcripts** [1, 19]; span the **3' untranslated region** of convergently arranged genes, and cross over promoters and into TUs regardless of orientation [12, 13, 19] (Figure 1). As the depth of sequencing improves, more of these transcription events are uncovered.

The effects of this **pervasive transcription** can spread to influence genes or non-coding TUs in the vicinity by altering **DNA supercoiling** or determining the structure and composition of the chromatin, which in turn influence whether transcription can initiate. Thus, these pervasive transcription events harbour a wealth of additional gene regulatory potential, and present two new concepts in gene regulation.

The first is the concept of the interleaved genome, in which there is a high degree of interdependency between transcription units, which is co-opted as a gene regulatory mechanism [19]. The second concept is that transcription itself regulates transcription initiation or repression at many regions of the genome [19]. The precise outcome depends on the relative orientation of the TUs and whether transcription events are contemporaneous. Most data are obtained from genetically identical but phenotypically heterogeneous populations of cells or tissues. This heterogeneity arises in part because of the inherent cycles within cells that influence transcriptional output, such as the cell division cycle [20] and the metabolic cycle [19, 21], and is an important consideration in our understanding of the consequences of an interleaved genome. In this review, the functional relationships between TUs will be explored, focusing on how interspersion and orientation of different types of TUs has been exploited as a regulatory feature of canonical gene expression in eukaryotes.

How transcription units depend on each other.

An interleaved genome is one in which there is high degree of interdependency between transcription units as a result of overlapping transcription. These transcription units may be arranged in convergent, divergent or tandem orientations with respect to one another.

Overlapping transcription: convergent overlapping transcription over the 3' region of TUs

In most organisms, convergent overlapping transcription over the 3' region of TUs is widespread. For example, taking the 7272 coding and non-coding TUs of *S. cerevisiae* into consideration, 2258 convergent TUs show overlapping transcription while 854 do not. The median overlap is larger when one of the convergent transcription units is non-coding (426bp) compared to both being protein coding genes (92bp) [19]. As the number of non-

coding TUs and the extent of transcription downstream of the **transcript end site (**TES) of genes continues to be revised in mammalian cells, *Drosophila melanogaster* and yeast [12, 13, 15], it is likely that the current predictions of the number of convergent overlapping TUs is vastly underestimated.

Convergent overlapping antisense transcription over the gene body

In mammalian cells, *D. melanogaster* and yeast, many genes show overlapping convergent transcription over the **gene body** (Figure 1). These antisense transcription events may initiate divergently from downstream promoters [4, 13-15, 22-24], from the 3' ends of genes independently of a conventional promoter in the vicinity [6, 25] or from enhancers located within introns [26] (Figure 1). Moreover, antisense transcription, and thus the production of NATs, is conserved in budding yeast species [27, 28], and is regulated by meiosis [29-32], the cell cycle [20], the metabolic cycle [19] and environmental stress [6, 19, 33-36]. The production of NATs is attenuated by a variety of different mechanisms, including removing TATA binding protein (TBP) from the antisense promoter [37]; forming **R-loops** [38]; early transcription termination [39-42] including the action of the nuclear exosome [43-45]; formation of repressive chromatin environments [46, 47]; and the formation of mutually exclusive **gene loops** [6, 19, 23, 48]. Despite these mechanisms to limit antisense transcription, detectable levels are found throughout genes in mammals, flies and *S. cerevisiae*.

In mammalian cells, there are short regions of convergent transcription initiating downstream of the **+1 nucleosome** proximal to 25% of active promoters, particularly at lower-expressed genes [11, 12, 49]. Prominent **DNase I hypersensitivity sites** flank the convergent TUs, indicating that antisense transcription initiates at a defined promoter located 150±50bp downstream of the sense TSS, likely reflecting the +1 nucleosome positioned between the sense and antisense TSSs. Subtle shifts in the positioning or composition of the +1 nucleosome, as a result this promoter-proximal convergent transcription, could influence sense transcription positively or negatively or influence the fate of the transcripts [50-52]. Antisense transcription is found in the vicinity of sense promoters at approximately 75% of genes in *S. cerevisiae* [53]. In yeast, levels of sense and antisense transcription do not correlate genome-wide [6, 53], suggesting that antisense transcription influences sense transcription in distinct ways at different genes, promoting [54, 55] or repressing sense

transcription [19, 32, 56], or having no effect at the population level [53], depending on the context [33].

Transcriptional interference at tandemly arranged TUs

In yeast, flies and mammalian cells, there is extensive transcription both upstream and downstream of conventional genes on the same strand (Figure 1). Much of the upstream transcription over the promoter and into the TU is from non-coding TUs that have the capacity to interfere with the transcription of the downstream gene by **transcriptional interference** (TI) [57-63]. Similarly, upstream genes can mediate TI of downstream genes when in tandem [19, 64]. Using **transcript isoform sequencing** (TIF-seq) [1], high resolution mapping of transcripts at the 2747 tandem gene pairs in the *S. cerevisiae* genome reveals that 27% (743) express overlapping isoforms, with long isoforms having the potential to regulate the downstream gene [19, 65]. Furthermore, 95% of genes in HeLa S3 cells, 88% of genes in HEK293T cells and many genes in flies and *S. cerevisiae* show transcription downstream from the TES [11-13, 15], suggesting the possibility for more extensive TI. The TIF-seq dataset for *S. cerevisiae* also reveals that 6.7% of genes produce bi-cistronic or tri-cistronic transcripts, although this number is likely to be a vast underestimate due to the low circularization rate of long transcripts in this assay [1]. Like the long isoform transcripts, transcription of the multi-cistronic transcripts also has the potential to repress genes downstream, particularly those in tandem arrays [19].

In summary, overlapping transcription results from transcription upstream, downstream and antisense to any given gene and the consequences depend on the environment of that gene and its relative orientation with respect to neighbouring genes.

Consequences of overlapping transcription

It is clear that genomes are extensively interleaved. The experimental evidence supporting the high degree of interdependency between transcription units as a result of overlapping transcription is reviewed in this section and covers the role of DNA supercoiling, double-stranded siRNA-mediated formation of heterochromatin as a result of overlapping transcription, the mechanisms that limit transcriptional interference, and the evidence for a plastic interleaved genome in which novel TUs are formed as a result of simply reducing the expression of a gene in interleaved gene cluster.

Coupling transcription at a distance by torsional stress

Torsional stress, a result of twisting DNA, is generated by virtually all processes that act on DNA and produces supercoils that propagate over several kilobases [66-68] (Figure 2A). Supercoiling acts locally such that transcription of genes within 3kb of each other does not occur independently, but is coupled, leading to mutual effects without any interaction of transcription machineries. This may underlie the widespread co-regulation of neighbouring genes, particularly divergent genes, whose expression is enhanced by torsional stress [66]. Divergent RNA transcription generates negative and positive supercoiling behind and in front of RNA polymerase II, respectively, which alters nucleosome occupancy [68] and gene activation [67, 69]. Negatively supercoiled DNA is known to favour transient separation of the two DNA strands [70]. For eukaryotic divergent promoters this may allow transcription factor-independent initiation of a neighbour [71] or simply increase transcription frequency [72], and could explain the high frequency of divergent non-coding transcription from promoters in eukaryotic genomes [4, 12-14, 16, 73]. In HeLa S3 and HEK293T cells, more than 80% of genes show divergent antisense transcription upstream of canonical promoters (average length 500nt [45], transcription start sites (TSSs) 250 ±50bp apart) [12]. Moreover, because transcription of both lncRNAs and antisense transcripts result in the accumulation of supercoils, torsion is a potential candidate for how these transcription events affect promoters at a distance or in close proximity [66, 74]. Resolution of accumulated supercoils by **topoisomerases** (Topo I, II) in human cells is significantly slower than the polymerase elongation rate through chromatin, meaning that the effects of accumulated supercoils can be dissociated in time from the act of transcription that produces them [75]. Another consequence of negative supercoiling-induced transient DNA strand separation is the formation of R-loops [70]: structures resulting from annealing of RNA to its genomic template either during transcription, or via a stable RNA *in trans*, to generate an DNA:RNA hybrid and a displaced region of single-stranded DNA [76, 77] (Figure 2B). R-loops are associated with the initiation of antisense transcripts to generate dsRNA [25].

Local heterochromatin formation by overlapping transcripts

Changes to the chromatin are commonly associated with overlapping transcription and can be mediated by transcripts, the act of transcription, or both. The formation of **heterochromatin** is associated with the double-stranded RNA (dsRNA) products of overlapping transcription (Figure 2) [78]. In mammalian cells, the formation of heterochromatin acts as a safety mechanism because endogenous dsRNA triggers an interferon response and cell death [79]. **Dicer** limits the accumulation of dsRNA by producing **small interfering RNAs** (siRNAs)

that function in the nucleus in a complex with an **argonaute** protein to mediate the conversion of chromatin from an active to a repressed heterochromatic state [80]. Generation of siRNAs from dsRNA and targeted formation of heterochromatin also occurs in *Schizosaccharomyces pombe*, *Caenorhabditis elegans* and *D. melanogaster* [81, 82]. Transient formation of heterochromatin leading to termination of transcription at loci with convergent overlapping transcripts may also contribute to genome stability by preventing collisions between RNA and DNA polymerases [83, 84]. Differentiated human cells contain distinct argonaute complexes enriched for siRNAs from genes with NATs, suggesting that NATs and their complementary transcripts can form dsRNA [85]. In the plant *Arabidopsis thaliana*, the *COOLAIR* antisense transcripts are required for cold-dependent silencing of Flowering Locus C during **vernalisation**, mediating the replacement of histone H3 lysine 36 methylation (H3K36me) with H3K27me3, and transient polycomb-mediated heterochromatic silencing at the nucleation site within the gene body [86].

Although originally a mechanism to protect cells, to discriminate self- from non-self-RNAs and to stabilise repetitive genomic regions, in many organisms dsRNA-mediated local formation of heterochromatin via H3K9 or H3K27 methylation creates local roadblocks for RNAPII elongation and has been exploited as a regulatory feature, controlling different aspects of gene expression [78]. In *S. pombe*, genes that encode the RNAi machinery itself are themselves convergently arranged and subject to siRNA-mediated co-repression of expression [80, 87]. Remarkably, down-regulated expression is transient and cell cycle-dependent. In G1, overlapping transcription occurs, whereas in G2, efficient transcript termination dominates, preventing overlapping transcription. **Cohesin** is recruited via transient heterochromatin formation and mediates the cell cycle-dependent switch to the proximal TES, and efficient transcription termination, allowing expression of the RNAi machinery (Figure 2C). This has important consequences for centromere function in *S. pombe*, because overlapping transcription and an siRNA:argonaute complex mediate formation of heterochromatin at the *S. pombe* centromere, which is required for chromosome segregation during mitosis [87].

Local formation of heterochromatin to reinforce transcription termination also occurs at the 3’ ends of mammalian genes that do not overlap with a convergent TU. Here, the dsRNA is formed as a result of the production of short overlapping antisense transcripts [25]. The production of the short antisense transcripts requires an R-loop, formed during sense

transcript 3' end formation [25, 88-90]. How the R-loop activates antisense transcription is not understood but antisense transcription arising from terminator regions in the absence of a canonical downstream promoter has been described in *S. cerevisiae* [6]. These events are associated with TBP and TFIIB, suggesting they are the result of *bona fide* initiation. Transcription of repetitive DNA tends to produce R-loops, leading to genomic instability. Subsequent convergent antisense transcription facilitates dsRNA-mediated heterochromatin formation, which stabilises the repetitive DNA while repressing gene expression observed in conditions such as fragile X syndrome and amyotrophic lateral sclerosis (ALS) [90-94]. Thus R-loops, coupled to local overlapping antisense transcription and heterochromatin formation, can be used to aid transcription termination at the end of genes. When R-loops are formed elsewhere in the gene, particularly in the vicinity of the promoter, the consequential repression of gene expression is often associated with disease.

Chromatin-mediated TI at tandem TUs

S. cerevisae lacks the machinery for RNAi-mediated heterochromatin formation and repression of transcription, allowing an insight into the consequences of the act of overlapping transcription itself, which also influences chromatin structure. Canonical TUs show a stereotypical distribution of post-translational modifications to the nucleosomal histone proteins over the transcribed region. Trimethylation of lysine 4 on histone H3 (H3K4me3) together with H3 lysine acetylation (H3Kac) is found at the start of a TU [95], while H3K36me3 accumulates over the transcribed region, stabilising the chromatin, controlling the kinetics of transcription elongation, promoting histone deacetylation and limiting transcription initiation [78, 96] (Figure 3A). Extending transcription of canonical TUs beyond the normal TES is common and leads to deposition of elongation-specific post-translational modifications that repress transcription initiation from downstream promoters (Figure 3B) [19, 59, 60]. Some regulatory non-coding TUs that mediate TI in tandem are also marked in this way [31]. However, this is not a universally conserved mechanism, as some non-coding TUs mediate TI by simply increasing nucleosome density at the promoter independent of H3K36me3 [59] or by controlling binding of transcription factors to the promoter [62, 97]. As the production of these upstream TUs is often regulated, they in turn regulate transcription of the gene subject to TI.

Insulating genes from TI

Given that transcription beyond the TES is common in yeast and mammals [12, 13, 65], it is not surprising that mechanisms exist to shield genes in tandem arrays from TI. In *S.cerevisiae*, one mechanism involves using protein complexes to block transcription, such as the TFIIIB component of the RNAPIII transcription machinery or other DNA-bound transcription factors (TFs) such as Reb1 [42, 98, 99]. *SUT467* is separated from *ATG31* by a tRNA gene. If TFIIIB complex binding to the tRNA gene is weakened, transcriptional interference of *ATG31* occurs by transcription from *SUT467*, reducing *ATG31* expression leading to defects in autophagy and loss of fitness [98]. The Brf1 component of TFIIIB shows growth-regulated expression, suggesting that in quiescent cells, there will be an increase in TI events normally protected by TFIIIB. Thus, protection of neighbouring genes from TI may be additional and important functions for the TFIIIB complex and TFs.

A second mechanism, again in *S.cerevisiae*, involves transcription of overlapping antisense TUs, specifically those upstream and divergent from a downstream promoter in a tandem array, which might explain their near ubiquitous nature (Figure 3C). This effectively creates a divergent promoter, which forms a boundary domain in the higher order folding of the chromosome [100], so limiting transcription and thus TI. At the *HMS2:BAT2* tandem array, transcription of the *HMS2* antisense TU, *SUT650*, insulates the *BAT2* promoter from interference by the *HMS2* long isoform and bi-cistronic transcription events [19]. This requires an allele of *SUA7*, encoding TFIIB but specifically lacking the ability to form gene loops [101, 102], which are RNAPII-dependent higher order structures in chromatin between the beginnings and ends of genes or loci, probably similar to chromatin globules [100], and known to be associated with both coding and non-coding TUs [6]. This suggests that part of the process of insulating *BAT2* from *HMS2* will involve the formation of a boundary between these genes and that this requires the transcription of *SUT650*. At a given locus, it is envisaged that only one gene loop can form, involving either the sense or antisense TU, explaining how **transcriptional insulation** occurs. A genome-wide analysis identified many other tandem gene pairs in *S. cerevisiae* that share features with the *HMS2:BAT2* pair, including transcriptional interference of the downstream gene and insulation by antisense transcription [19]. Switching of these tandem clusters between the sense and antisense states is coordinated with the inherent cycles of growth and quiescence within the cell, supporting a regulatory function [19, 103]. In addition, switching between sense- or antisense-dominant states also leads to the coordinated up- and down-regulation of transcription at neighbouring genes (Figure 3D) [19].

Antisense transcription is not simply the reverse of sense transcription

It is often assumed that antisense transcription will deposit histone modifications in a reverse pattern to sense, thus repressing expression from the sense promoter [104]. However, this is not always the case. In mammalian cells, an antisense-mediated R-loop prevents chromatin compaction at the vimentin promoter, activating transcription and maintaining cellular identity [52], and in yeasts antisense-mediated chromatin remodelling also activates some promoters [53-55, 105] whereas others are repressed [31, 53, 56]. Genome-wide analysis in *S. cerevisiae* reveals that sense and antisense transcription are different, with antisense transcription associated with major functional changes to the chromatin over the promoter and gene body that have context-dependent functions [33, 53]. The chromatin becomes more dynamic; over the gene body there is increased histone exchange (the incorporation of newly synthesised histones in the chromatin), increased H3Kac and the loss of histone modifications associated with stable nucleosomes, such as H3K79me3 and H3K36me3 [53] (Figure 4). At the promoter, nucleosome occupancy increases, although turnover rates are high, with a concomitant reduction in the size of the **nucleosome depleted region** (NDR) [53, 106]. Antisense-mediated closure of the NDR and its subsequent opening by sense transcription creates a bi-stable two-state chromatin state, which is proposed to be a feature of genes that show bursts of transcription and noise, resulting in variability in the levels of transcripts in individual cells within populations [107, 108] (Figure 4). Interestingly, the antisense-associated gene body chromatin features are also associated with genes that show noise [108].

Cell-to-cell variability in gene expression mediated by overlapping transcription

Transcription factors (TFs) compete with nucleosomes for occupancy at the promoter [109]. As overlapping transcription resets nucleosome positioning at the promoter, it can indirectly influence transcription factor binding and thus the levels and mode of sense transcription [53, 59] (Figure 4). One consequence of antisense transcription is variability in the number of cells in a population that are able to respond to a TF [19, 31, 33, 110]. This is exemplified by *PHO84* in *S.cerevisiae*, which has an antisense-mediated nucleosome-occupied promoter and very high levels of sense transcription. This was something of a paradox, because high nucleosome occupancy generally reflects repression of transcription [56, 110, 111]. Single molecule RNA FISH revealed that sense transcripts are restricted to just 20% of the cells in the population [110]. In the remaining 80% of cells, the chromatin presumably adopts the nucleosome-occupied antisense state and many of these cells contain a single antisense transcript. Low levels of convergent antisense transcription, such as at *PHO84*, are often

written-off as irrelevant noise, but for a mechanism where a modulation of state involves altering the chromatin structure, only one round of transcription is required. Moreover, once changed, this state may be maintained though cell divisions. Support for chromatin-mediated events leading to modulation of sense transcription comes from a number of overlapping non-coding transcription-related events in *S. cerevisiae* [7]. For many of these situations, the presence of sense and regulatory non-coding transcripts is strongly anti-correlated; they are rarely co-expressed in single cells. This suggests the cells show bi-modal switches, moving from one state to another. The proportion of cells in either state is regulated by the number of cells undergoing non-coding transcription [7, 19, 31]. Thus antisense transcription might not change overall levels of transcription of a gene in the population, but rather change the number of cells in the population in the sense state.

A plastic interleaved genome in S. cerevisiae

An interleaved genome with overlapping transcription units requires that promoters and transcription termination signals in the sense or antisense orientations are by-passed. Gene loops may play a key role in linking a particular promoter:terminator combination to form a transient functional TU while excluding other promoters or terminators by read-through TI or chromatin structures (Figure 5). Thus, any transcription event, however short, would have a defined start and end point before initiation occurs and other productive and non-productive (non-coding or bi-cistronic) TUs would not be able to form unless a switch occurs. This explains how an antisense TU such as *SUT650* could insulate the downstream *BAT2* promoter: the *HMS2* promoter would be unable to participate in the formation of a functional TU. Only one TU will exist for at a given locus in one cell in a population at any one time, but in a population of cells all the potential TUs that can be expressed under a given set of conditions will be evident in the transcriptome.

Finally, engineering interleaved overlapping TUs reveals the plasticity underlying the formation of alternative TUs. The sense output of *HMS2* is moderated by antisense transcription, demonstrated by insertion of a transcription terminator to separate the sense and antisense TUs (Figure 5). The antisense TU acts by restricting the number of the cells in the population expressing transcripts from the *HMS2* promoter [19]. When the *KanMX* TU is embedded in tandem within *HMS2*, the promoter driving *KanMX* forms divergent TUs in both the sense and antisense orientations with proximal terminators, limiting sense transcription from the *HMS2* promoter, just as observed with the native gene. However,

weakening the *KanMX* promoter by mutation results in a redefinition of the TUs and a new antisense TU that links the terminator and promoter region of *HMS2*, transcribing through the inserted tandem gene and its terminator [19]. Thus, distinct overlapping antisense TUs balance the transcriptional output from the *HMS2* promoter, and importantly, the number of cells in which the *HMS2* promoter is active. Moreover, the function of terminators is conditional, consistent with a model in which a gene loop defines the beginning and end of a TU rather that a DNA sequence dictating initiation or termination *per se*. These observations illustrate the inherent plasticity of an interleaved genome and show how the genetic modification of one TU can lead to the appearance of a novel TU.

Two main mechanisms by which overlapping transcription mediates its effects have been discussed: at a distance by torsional stress and by altering the chromatin as a direct result of transcription or indirectly via the action of the dsRNA products of overlapping transcription. Whichever mechanism is used, both influence local gene expression either positively, for example aiding transcription termination or promoting transcription initiation, or negatively by down-regulating a gene. In many situations, the outcome depends on where local overlapping transcription occurs within a TU or more generally, the context of a TU, for example the orientation and strength of the neighbouring TU. The near-ubiquitous divergent transcription from a promoter is likely to help insulate that promoter from transcriptional interference from upstream TUs, especially important given that transcription generally extends way beyond the TES.

Figure Legends

Figure 1. **Examples of overlapping transcription**. **(A)** A 24kb region of *S. cerevisiae* chromosome II showing 11 protein coding genes and 2 tRNA genes (blue boxes for all annotated transcripts) and transcription (NET-seq) on the Watson (W; top) or Crick (C; bottom) strands. **(B-D)** Examples of overlapping transcription (mNET-seq) in HeLa cells at three regions of the human chromosome 1 containing genes in tandem (**B**), convergent genes (**C**) or divergent genes (**D**). Lines with arrow heads indicated the extent and direction of

transcription and are coloured according to type: blue, transcripts over coding regions; pink, antisense transcripts; cyan, intergenic transcription producing unstable transcripts (e.g.cryptic unstable transcripts; CUTs or Xrn1-sensitive unannotated transcripts; XUTs). The type of overlapping transcription is indicated by the coloured boxes below the gene map as indicted in the key. Antisense +150 is antisense transcription initiating 150±50bp downstream of the sense promoter. The gene maps are not shown at the same scale. See key in for details. Schematics show NET-seq data from [19] and [11]

Figure 2. **Convergent overlapping transcription leads to local heterochromatin formation**. **(A)** Transcription alters torsional stress (supercoiling), which is resolved by topoisomerases (Topo I, II) (adapted from [68]). Negative torsion behind RNA polymerase II can lead to melting of the two DNA strands and can promote divergent transcription or the formation of an R-loop (see B). **(B)** In mammals, a co-transcriptional (1) R-loop (2) is associated with overlapping antisense transcription (3) and the formation of dsRNA (4). dsRNA is a substrate for dicer, which yields siRNA (5) that is delivered to a silencing complex that includes an argonaute protein (AGO). The silencing complex then converts the local chromatin structure to heterochromatin (e.g., H3K9me3; pink nucleosome) (6) followed by recruitment of cohesin leading to efficient termination of transcription (7). **(C)** In *S.pombe*, convergent overlapping transcription leads to dsRNA (1), which is diced and sliced into siRNA (2), associated with the RITS complex (3) and used to form local heterochromatin (pink nucleosomes) (4). This recruits cohesin (5), which in turn promotes transcription termination (6), reducing overlapping transcription allowing the production of pre-mRNAs. Transcripts (pink) subject to dsRNA formation are unstable and degraded. Thus, heterochromatin formation is required for efficient gene expression at some convergent genes.

Figure 3. **Interference and insulation by overlapping transcription in tandem gene arrays**. **(A)** Typical chromatin organization for two genes, B and C, when there is no transcriptional interference TI and no antisense transcription. Ovals are nucleosomes, solid-filled ovals are positioned nucleosomes; translucent ovals are fuzzy, poorly positioned, nucleosomes; pink ovals have H3Kac found around the TSS of active genes; K4me3 is found at nucleosome +1 downstream of the TSS of active genes; blue ovals have H3K36me3 found in the gene body of active genes; orange and green ovals are not marked with these modifications and are found in genes or intergenic regions, respectively. Promoters are green lines, terminators are red lines, NDR is nucleosome depleted region, TES is transcript end

site, TSS is transcription start site. The level of selected post-translational modifications to histone H3 are indicated by the different font sizes. The number of the nucleosomes from -1 to +8 is shown for each gene. Arrows indicate direction of transcription. **(B)** Changes to chromatin as a result of TI by transcription beyond the TES of gene B into gene C. **(C)** Changes to chromatin at gene B as a result of antisense transcription. Note the nucleosome occupied promoter for gene B, histone turnover and histone acetylation throughout the antisense transcribed chromatin. **(D)** Schematic showing a cluster of four genes (A-D) whose expression is regulated by growth (A and B) or quiescence (C and D) and the role of the TU antisense to gene B in switching between one state and the other. At the *HMS2* locus the genes in order A-D would be *RPS4A, HMS2, BAT2* and *YJR149W*. *SUT650* is antisense to *HMS2*. Note the expression of the four genes is controlled simply by the growth–specific TFs that bind to the divergent promoter between genes A and B. Blocked lines indicate repression of a TU. Protein coding transcripts (blue), non-coding transcripts (pink), transcription factors (TFs), red or dark green ovals.

Figure 4 Nucleosomes compete with TFs for promoter occupancy and influence modes of sense transcription (A) (i) Open chromatin at a Poisson promoter, not subject to antisense transcription, (ii) which initiates transcription at a constant rate and (iii) individual cells in the population show little variance in initiation of transcription with respect to the mean level. Nucleosomes -1 and +1 are well positioned with a clear NDR, facilitating dynamic TF binding and regular transcription initiation events. **(B)** (i-iii) Closed but dynamic chromatin at the promoter of a gene subject only to antisense transcription. The nucleosome-occupied promoter effectively competes the binding of low affinity TFs (reflecting changed environmental conditions) and there are very low levels of sense transcription. **(C)** (i) Mixed open and closed chromatin at a two-state promoter subject to both sense and antisense transcription. TFs compete with dynamic nucleosomes for occupancy of the promoter. When nucleosome-occupied there is no transcription initiation but transcription initiation occurs when the nucleosome is displaced by TF binding, leading to alternative on and off states. The uniquely marked chromatin over the body of the gene (dynamic, highly acetylated on H3 and poorly methylated at H3K36) is proposed to facilitate transcription elongation and re-initiation of transcription, leading a transcription burst (ii). Assuming no change in the rate of turnover of transcripts, this will lead to variability in the rate at which transcripts accumulate in the cytoplasm in different cells, giving rise to noise, unless transcript turnover rates are altered to accommodate this variance from the mean levels of transcription initiation for the

two state promoter (iii). Thus antisense transcription may be one of the missing factors involved in switching the two-state promoters from on to off.

Figure 5. **The plastic interleaved genome in *S. cerevisiae*. (A)** Schematics showing (i) the sense (blue) and antisense (pink) TUs formed at the *HMS2* gene or (ii-iv) at *HMS2* when interleaved in tandem with three versions of a TU expressing *KanMX* (green), thus separating the *HMS2* sense and antisense TUs. P is promoter and T is terminator, colour-coded according to the nature of the TU. (ii) Separating the *HMS2* and *KanMX* TUs with a transcription terminator (T, red) results in three different TUs and no antisense TU over the *HMS2* promoter region; (iii) removing the terminator between the *HMS2* and *KanMX* promoters results in four different TUs including at new antisense TU over the *HMS2* promoter. This construct has two promoters in tandem and two terminators in tandem; (iv) disabling the *KanMX* promoter results in a new antisense TU (pink) spanning the *HMS2* terminator to promoter and reading though the *KanMX* terminator (T, red). There are two other TU on this construct and very low levels of transcripts from the disabled divergent *KanMX* promoter. **(B)** (i-iv) Schematics illustrating the distribution of sense and antisense transcripts detected in seven yeast cells using RNA FISH with sense and antisense-specific probes at the position shown in (**A**i). The colour of a yeast cell indicates whether it contains sense (S; blue) or antisense (AS) transcripts and the source of that transcript (pink, the *HMS2* antisense promoter; green, the *KanMX* divergent promoter). White cells contain neither transcript. The number of cells with each type of transcript is shown below the box. The number of cells expressing the *HMS2* sense TU is moderated by an antisense TU over the same region. **(C)** (i-iv) Schematics illustrating alternative gene loops for each of the 2, 3 or 4 TU for the constructs shown in (**A**i-iv) anchored at their bases by RNAPII (black oval). Only one loop exists at one locus at one time but the locus alternates between different loops over time. The promoter and terminator involved in the gene loop are shown together with those that are not transcriptionally engaged at the same locus, colour coded as in **(A)**. The arrow indicates the direction of transcription (sense: arrow L to R, or antisense; arrow R to L).

Box 1 Figure 1. A comparison of different methods for assessment nascent transcription in different organisms. Screenshot from the integrated genome viewer (IGV) [112] showing nascent transcription (NET-seq) in a region around *CUT300* in *S.cerevisiae* (**A**), PRO-seq for a region around *CG3224* in *D.melanogaster* S2 cells (**B**), the region around the *SIK1* gene in HeLa cells comparing GRO-seq and mNET-seq for the crick strand only (**C**), and mNET-seq

around the *SIK1* genes on both strands of DNA (**D**). C is the Crick strand and W is the Watson strand of DNA.

Glossary Box

+1 nucleosome: The nucleosome downstream from the TSS often enriched for PTMs such as H3K4me3 and H3Kac.

3' untranslated region (3' UTR): 3'untranslated region is located between the translation stop codon and the transcript end site or polyadenylation site.

Argonaute: component of silencing complexes containing the ssRNA guide strand of siRNA. One type functions in the nucleus to mediate the establishment of heterochromatin.

Bi- and tri-cistronic transcripts: transcripts that start at the TSS of one gene and are transcribed continuously until the TES of the next gene downstream (bi-cistronic) or the gene beyond (tri-cistronic). They mediate transcriptional interference of the downstream genes in tandem arrays by altering the chromatin environment over their promoters. These transcripts are generally low abundance but can be detected using Northern blots or TIF-seq.

Cohesin: a large ring-shaped protein complex that functions to keep sister chromatin together until anaphase in mitosis, usually at heterochromatin regions of the genome. When the heterochromatic regions form within transcription units, RNA polymerase II slows down and is more likely to terminate transcription.

Cryptic unstable transcript (CUT): a class of transcript degraded by the nuclear exosome and defined as being stabilized in the absence of the nuclear component of the exosome Rrp6 (yeast and flies) or its equivalent in mammalian cells. Note that some CUTs are fully transcribed and subject to degradation, while at others Rrp6 plays a role in early transcription termination leading to an increase in both transcription and transcript stability in an *rrp6* mutant strain.

Dicer: an endonuclease that cleaves dsRNA to yield siRNAs.

DNase 1 hypersensitivity sites: found at promoters and enhancers in chromatin, reflecting the binding of proteins to the DNA.

DNA supercoiling: refers to the over- or under-winding of a DNA strand.

Gene body: The region between the transcription start site (TSS) and the transcript end site (TES).

Gene loop: a protein-mediated interaction between the 5' and 3' region of a gene detectable using capturing chromosome conformation (3C) technology.

Gene silencing: usually refers to a small interfering RNA (siRNA)-based mechanism. Gene silencing may occur in the cytoplasm, involving argonaute complexes known as RISC (RNA induced silencing complex) that lead to silencing of transcripts, preventing their translation. Gene silencing also occurs in the nucleus where RNAi-associated argonaute complexes known as RITS (RNA induced transcription silencing *in cis*) lead to local or regional formation of heterochromatin involving single nucleosomes or larger blocks of nucleosomes that can be associated with modulation (often repression) of transcription.

Heterochromatin: different states of chromatin generally associated with the gene silencing. Heterochromatin can be associated with histone post-translational modifications such as H3K9me3 and H3K27me3.

Intergenic region: a stretch of DNA sequences located between genes.

Long non-coding RNA (lncRNA): an RNA molecule, usually >200nt in length, that is predicted not to encode protein.

Natural antisense transcript (NAT): Cis-natural antisense transcripts (cis-NATs) are transcribed from the same genomic locus but from the opposite DNA strand and can therefore form a perfect RNA:RNA hybrid if both transcription events are contemporaneous.

Nascent elongating transcript sequencing (**NET-seq):** nascent elongating transcript sequencing reveals the density of RNA polymerase across the genome with single nucleotide resolution by capturing nascent RNA transcripts directly from live cells through their association with the elongating ternary complex (DNA:RNA:RNAP).

Nucleosome depleted region (NRD): promoters generally show regions depleted for nucleosomes at which transcription factors and RNAPII bind.

Pervasive transcription: transcription that is spread widely throughout the genome.

Precision run-on sequencing (**PRO-seq):** a derivative of global run-on sequencing which assesses the incorporation of BrUTP into nascent elongating transcripts. The labelled RNA is

purified using anti-BrUTP antibodies and sequenced after converting the products to cDNA. PRO-seq allows a single labelled nucleotide to be incorporated by run-on, thus giving strand-specific and nucleotide resolution data.

Small interfering RNA (siRNA): double stranded RNA molecules 20 to 25 bp long.

RNA-seq: also known as Whole Transcriptome Shotgun Sequencing, RNA-seq reveals a snapshot of the RNA present in a cell at a given moment of time, the transcriptome, using next-generation sequencing.

R-loop: a three-stranded nucleic acid structure, consisting of a DNA:RNA duplex and a displaced ssDNA found in cytosine-rich genomic regions during transcription.

Stable unannotated transcript (SUT): SUTs are generally non-coding transcripts and may be found antisense to genes or in intergenic regions in *S. cerevisiae*. Many SUTs are also XUTs, being substrates for cytoplasmic degradation by Xrn1. SUTs may represent the non-coding transcripts that are retained in the nucleus, or sequestered in P-bodies, and thus protected from the action of Xrn1.

Transcript isoform sequencing (TIF-seq): a method for mapping the 5' and 3' ends of individual transcript isoforms. Transcript isoforms are versions of a transcript that have altered usage of the TSS or the TES, changing the beginning or end of the transcript, or use alternative splice sites, often changing the coding capacity of the transcript.

Transcript end site (TES): the site at which the transcript ends and the site of addition of the polyadenylate (polyA) tail. The TES is not the same as the transcription termination site (TTS).

Transcriptional Interference (TI): the direct negative impact of one transcriptional activity on a second transcriptional activity in *cis*.

Transcriptional Insulation: the formation of TUs or higher order structures in the chromatin that prevent transcriptional interference.

Topoisomerases: enzymes that regulate the over-winding or under-winding of DNA by cutting the DNA to unwind the DNA.

Transcript start site (TSS): the site at which transcription initiates and the site of addition of the ^{7me}G cap.

Transcription Unit (TU): a region of transcribed DNA beginning at the transcription start site (TSS) and ending at the transcription termination site (TTS). Note the TTS is not the same as the transcript end site (TES), which is the site of transcript cleavage and polyadenylation. Transcription units often overlap.

Xrn1-sensitive unannotated transcript (XUT): ≈ 66% of XUTs are antisense to protein coding regions in *S. cerevisiae* but are also found in intergenic regions. XUTs are fully processed, capped and polyadenylated, before being exported to the cytoplasm for degradation.

Vernalisation: the promotion of flowering by exposure of young plants to cold temperatures.

Box 1 Methods for assessing nascent transcription.

There are currently two methods for assessing nascent transcription at single-nucleotide resolution, native elongating transcript sequencing (NET-seq) [11, 13] and precision run-on sequencing (PRO-seq) [15]. PRO-seq was developed from the established genomic run-on sequencing (GRO-seq) method [14] and uses biotin-labeled ribonucleotide triphosphate analogues (biotin-NTP) in nuclear run-on reactions in isolated nuclei in vitro. This allows efficient affinity purification of nascent RNAs for high throughput sequencing from their 3' ends. Usually only one of one of the four biotin-labelled bases is supplied to the reaction, restricting the incorporation of a single or a few identical bases by RNAPII, resulting in sequence reads that have the same 3' end base within each library. The incorporation of the first biotin-base inhibits further transcript elongation, ensuring base-pair resolution. This technique produces good read counts but has the disadvantage of requiring the isolation of nuclei. This is straightforward for cells in culture but is more difficult for yeasts, worms and plants due to their tough cell wall or cuticle. NET-seq was originally developed in yeast [13] and had now been applied to mammalian cells (mNET-seq) [11]. This technique does not involve a cross-linking step and relies on the immunoprecipitation of RNA polymerase II with its associated nascent transcript from DNase1 or micrococcal nuclease digested chromatin. The 3'OH of a nascent transcript reflects where the active site of RNAPII lies along the genome and which strand is being transcribed. It is this feature of the nascent transcript that is exploited in NET-seq. The output reports the position of the 3'OH across the genome, thus providing base pair resolution, but does not distinguish RNAPII that is elongating, stalled or backtracked. This is a disadvantage compared to PRO-seq in which the

incorporation of the biotin-NTP requires transcript elongation. Figure 1 compares screenshots of nascent transcription profile, obtained using NET-seq, PRO-seq, GRO-seq or mNET-seq, at genes clusters in yeast, flies and mammals. The different methods all report extensive transcription on both strands of the DNA. A comparison of GRO-seq and mNET-seq around the mammalian *SIK1* gene reveal very similar gross profiles, suggesting the different techniques are capturing similar features of nascent transcription. There are other methods for assessing nascent transcription. These include: nascent RNA-seq in which nascent transcripts, with their 3'OH protected by RNAPII, are isolated from chromatin without an immunoprecipitation step, and are then sequenced from their 3'OH [12]; *in vivo* pulse labelling of RNA synthesis with bromouridine (BrU) [113]or thiouridine (4sU or 6tU) [114]; or RNAPII Chromatin immunoprecipitation (ChIP). In contrast to nascent transcript sequencing methods, ChIP is not strand-specific, cannot distinguish active from inactive molecules of RNAPII, has lower resolution (50-150bp) and uses a cross-linking step to freeze the elongation complex.

References

1 Pelechano, V.*, et al.* (2013) Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* 497, 127-131

2 de Klerk, E. and t Hoen, P.A. (2015) Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. *Trends Genet* 31, 128-139

3 Holoch, D. and Moazed, D. (2015) RNA-mediated epigenetic regulation of gene expression. *Nat Rev Genet* 16, 71-84

4 Xu, Z.*, et al.* (2009) Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033-1037

5 Yang, L.*, et al.* (2014) Long noncoding RNAs: fresh perspectives into the RNA world. *Trends Biochem Sci* 39, 35-43

6 Murray, S.C.*, et al.* (2012) A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription. *Nucleic acids research* 40, 2432-2444

7 Castelnuovo, M. and Stutz, F. (2015) Role of chromatin, environmental changes and single cell heterogeneity in non-coding transcription and gene regulation. *Curr Opin Cell Biol* 34, 16-22

8 Khorkova, O.*, et al.* (2014) Natural antisense transcripts. *Hum Mol Genet* 23, R54-63

9 van Dijk, E.L*., et al.* (2011) XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature* 475, 114-117
10 Pelechano, V. and Steinmetz, L.M. (2013) Gene regulation by antisense transcription. *Nat Rev Genet* 14, 880-893
11 Nojima, T*., et al.* (2015) Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* 161, 526-540
12 Mayer, A*., et al.* (2015) Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* 161, 541-554
13 Churchman, L.S. and Weissman, J.S. (2011) Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368-373
14 Core, L.J*., et al.* (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845-1848
15 Kwak, H*., et al.* (2013) Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* 339, 950-953
16 Neil, H*., et al.* (2009) Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457, 1038-1042
17 Gardini, A. and Shiekhattar, R. (2015) The many faces of long noncoding RNAs. *FEBS J* 282, 1647-1657
18 Lam, M.T*., et al.* (2014) Enhancer RNAs and regulated transcriptional programs. *Trends Biochem Sci* 39, 170-182
19 Nguyen, T*., et al.* (2014) Transcription mediated insulation and interference direct gene cluster expression switches. *eLife* 3, e03635
20 Granovskaia, M.V*., et al.* (2010) High-resolution transcription atlas of the mitotic cell cycle in budding yeast. *Genome Biol* 11, R24
21 Tu, B.P. and McKnight, S.L. (2006) Metabolic cycles as an underlying basis of biological oscillations. *Nat Rev Mol Cell Biol* 7, 696-701
22 Scruggs, B.S*., et al.* (2015) Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Molecular cell* 58, 1101-1112
23 Tan-Wong, S.M*., et al.* (2012) Gene loops enhance transcriptional directionality. *Science* 338, 671-675
24 Preker, P*., et al.* (2008) RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 322, 1851-1854
25 Skourti-Stathaki, K*., et al.* (2014) R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* 516, 436-439
26 Meng, F.L*., et al.* (2014) Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* 159, 1538-1548
27 Goodman, A.J*., et al.* (2013) Pervasive antisense transcription is evolutionarily conserved in budding yeast. *Mol Biol Evol* 30, 409-421
28 Yassour, M*., et al.* (2010) Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. *Genome Biol* 11, R87
29 Chen, H.M. and Neiman, A.M. (2011) A conserved regulatory role for antisense RNA in meiotic gene expression in yeast. *Curr Opin Microbiol* 14, 655-659
30 Gelfand, B*., et al.* (2011) Regulated antisense transcription controls expression of cell-type-specific genes in yeast. *Mol Cell Biol* 31, 1701-1709
31 van Werven, F.J*., et al.* (2012) Transcription of two long noncoding RNAs mediates mating-type control of gametogenesis in budding yeast. *Cell* 150, 1170-1181

32 Hongay, C.F.*, et al.* (2006) Antisense transcription controls cell fate in Saccharomyces cerevisiae. *Cell* 127, 735-745
33 Xu, Z.*, et al.* (2011) Antisense expression increases gene expression variability and locus interdependency. *Mol Syst Biol* 7, 468
34 Conley, A.B. and Jordan, I.K. (2012) Epigenetic regulation of human cis-natural antisense transcripts. *Nucleic acids research* 40, 1438-1445
35 Bitton, D.A.*, et al.* (2011) Programmed fluctuations in sense/antisense transcript ratios drive sexual differentiation in S. pombe. *Mol Syst Biol* 7, 559
36 Nadal-Ribelles, M.*, et al.* (2014) Control of Cdc28 CDK1 by a stress-induced lncRNA. *Molecular cell* 53, 549-561
37 Koster, M.J. and Timmers, H.T. (2015) Regulation of anti-sense transcription by Mot1p and NC2 via removal of TATA-binding protein (TBP) from the 3'-end of genes. *Nucleic acids research* 43, 143-152
38 Sun, Q.*, et al.* (2013) R-loop stabilization represses antisense transcription at the Arabidopsis FLC locus. *Science* 340, 619-621
39 Castelnuovo, M.*, et al.* (2014) Role of histone modifications and early termination in pervasive transcription and antisense-mediated gene silencing in yeast. *Nucleic acids research* 42, 4348-4362
40 Schulz, D.*, et al.* (2013) Transcriptome surveillance by selective termination of noncoding RNA synthesis. *Cell* 155, 1075-1087
41 Almada, A.E.*, et al.* (2013) Promoter directionality is controlled by U1 snRNP and polyadenylation signals. *Nature* 499, 360-363
42 Porrua, O. and Libri, D. (2015) Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol* 16, 190-202
43 Pefanis, E.*, et al.* (2015) RNA exosome-regulated long non-coding RNA transcription controls super-enhancer activity. *Cell* 161, 774-789
44 Lubas, M.*, et al.* (2015) The human nuclear exosome targeting complex is loaded onto newly synthesized RNA to direct early ribonucleolysis. *Cell Rep* 10, 178-192
45 Pefanis, E.*, et al.* (2014) Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* 514, 389-393
46 Marquardt, S.*, et al.* (2014) A chromatin-based mechanism for limiting divergent noncoding transcription. *Cell* 157, 1712-1723
47 Zofall, M.*, et al.* (2009) Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs. *Nature* 461, 419-422
48 Whitehouse, I.*, et al.* (2007) Chromatin remodelling at promoters suppresses antisense transcription. *Nature* 450, 1031-1035
49 Jonkers, I.*, et al.* (2014) Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. *eLife* 3, e02407
50 Morillon, A.*, et al.* (2005) Dynamic lysine methylation on histone H3 defines the regulatory phase of gene transcription. *Molecular cell* 18, 723-734
51 Morillon, A.*, et al.* (2003) Isw1 chromatin remodeling ATPase coordinates transcription elongation and termination by RNA polymerase II. *Cell* 115, 425-435
52 Boque-Sastre, R.*, et al.* (2015) Head-to-head antisense transcription and R-loop formation promotes transcriptional activation. *Proceedings of the National Academy of Sciences of the United States of America* 112, 5785-5790
53 Murray, S.C.*, et al.* (2015) Sense and antisense transcription are associated with distinct chromatin architectures across genes. *Nucleic acids research*

54 Uhler, J.P., *et al.* (2007) A role for noncoding transcription in activation of the yeast PHO5 gene. *Proceedings of the National Academy of Sciences of the United States of America* 104, 8011-8016
55 Crisucci, E.M. and Arndt, K.M. (2012) Paf1 restricts Gcn4 occupancy and antisense transcription at the ARG1 promoter. *Mol Cell Biol* 32, 1150-1163
56 Camblong, J., *et al.* (2007) Antisense RNA stabilization induces transcriptional gene silencing via histone deacetylation in S. cerevisiae. *Cell* 131, 706-717
57 Afek, A. and Lukatsky, D.B. (2013) Genome-wide organization of eukaryotic preinitiation complex is influenced by nonconsensus protein-DNA binding. *Biophys J* 104, 1107-1115
58 Martens, J.A., *et al.* (2004) Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature* 429, 571-574
59 Hainer, S.J., *et al.* (2011) Intergenic transcription causes repression by directing nucleosome assembly. *Genes Dev* 25, 29-40
60 Colin, J., *et al.* (2011) Cryptic transcription and early termination in the control of gene expression. *Genet Res Int* 2011, 653494
61 Toesca, I., *et al.* (2011) Cryptic transcription mediates repression of subtelomeric metal homeostasis genes. *PLoS Genet* 7, e1002163
62 Martianov, I., *et al.* (2007) Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* 445, 666-670
63 Proudfoot, N.J. (1986) Transcriptional interference and termination between duplicated alpha-globin gene constructs suggests a novel mechanism for gene regulation. *Nature* 322, 562-565
64 Greger, I.H., *et al.* (2000) Balancing transcriptional interference and initiation on the GAL7 promoter of Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America* 97, 8415-8420
65 Pelechano, V., *et al.* (2014) Genome-wide identification of transcript start and end sites by transcript isoform sequencing. *Nat Protoc* 9, 1740-1759
66 Meyer, S. and Beslon, G. (2014) Torsion-mediated interaction between adjacent genes. *PLoS Comput Biol* 10, e1003785
67 Naughton, C., *et al.* (2013) Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. *Nat Struct Mol Biol* 20, 387-395
68 Teves, S.S. and Henikoff, S. (2014) Transcription-generated torsional stress destabilizes nucleosomes. *Nat Struct Mol Biol* 21, 88-94
69 Roca, J. (2011) Transcriptional inhibition by DNA torsional stress. *Transcription* 2, 82-85
70 Roy, D., *et al.* (2010) Competition between the RNA transcript and the nontemplate DNA strand during R-loop formation in vitro: a nick can serve as a strong R-loop initiation site. *Mol Cell Biol* 30, 146-159
71 Leblanc, B.P., *et al.* (2000) An initiation element in the yeast CUP1 promoter is recognized by RNA polymerase II in the absence of TATA box-binding protein if the DNA is negatively supercoiled. *Proceedings of the National Academy of Sciences of the United States of America* 97, 10745-10750
72 Naughton, C., *et al.* (2013) Divergent RNA transcription: a role in promoter unwinding? *Transcription* 4, 162-166
73 Seila, A.C., *et al.* (2008) Divergent transcription from active promoters. *Science* 322, 1849-1851
74 Lavelle, C. (2014) Pack, unpack, bend, twist, pull, push: the physical side of gene expression. *Curr Opin Genet Dev* 25, 74-84

75 Kouzine, F.*, et al.* (2008) The functional response of upstream DNA to dynamic supercoiling in vivo. *Nat Struct Mol Biol* 15, 146-154
76 Yu, K.*, et al.* (2003) R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nat Immunol* 4, 442-451
77 Costantino, L. and Koshland, D. (2015) The Yin and Yang of R-loop biology. *Curr Opin Cell Biol* 34, 39-45
78 Naftelberg, S.*, et al.* (2015) Regulation of alternative splicing through coupling with transcription and chromatin structure. *Annu Rev Biochem* 84, 165-198
79 White, E.*, et al.* (2014) Human nuclear Dicer restricts the deleterious accumulation of endogenous double-stranded RNA. *Nat Struct Mol Biol* 21, 552-559
80 Gullerova, M. and Proudfoot, N.J. (2012) Convergent transcription induces transcriptional gene silencing in fission yeast and mammalian cells. *Nat Struct Mol Biol* 19, 1193-1201
81 Castel, S.E. and Martienssen, R.A. (2013) RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. *Nat Rev Genet* 14, 100-112
82 Guerin, T.M.*, et al.* (2014) Transgenerational functions of small RNA pathways in controlling gene expression in C. elegans. *Epigenetics* 9, 37-44
83 Castel, S.E.*, et al.* (2014) Dicer promotes transcription termination at sites of replication stress to maintain genome stability. *Cell* 159, 572-583
84 Zaratiegui, M.*, et al.* (2011) RNAi promotes heterochromatic silencing through replication-coupled release of RNA Pol II. *Nature* 479, 135-138
85 Werner, A.*, et al.* (2014) Contribution of natural antisense transcription to an endogenous siRNA signature in human cells. *BMC Genomics* 15, 19
86 Csorba, T.*, et al.* (2014) Antisense COOLAIR mediates the coordinated switching of chromatin states at FLC during vernalization. *Proceedings of the National Academy of Sciences of the United States of America* 111, 16160-16165
87 Gullerova, M. and Proudfoot, N.J. (2008) Cohesin complex promotes transcriptional termination between convergent genes in S. pombe. *Cell* 132, 983-995
88 El Hage, A.*, et al.* (2014) Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet* 10, e1004716
89 Wimberly, H.*, et al.* (2013) R-loops and nicks initiate DNA breakage and genome instability in non-growing Escherichia coli. *Nat Commun* 4, 2115
90 Loomis, E.W.*, et al.* (2014) Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. *PLoS Genet* 10, e1004294
91 Wang, J.*, et al.* (2015) Emerging role of RNA*DNA hybrids in C9orf72-linked neurodegeneration. *Cell Cycle* 14, 526-532
92 Salvi, J.S. and Mekhail, K. (2015) R-loops highlight the nucleus in ALS. *Nucleus* 6, 23-29
93 Butler, J.S. and Napierala, M. (2015) Friedreich's ataxia - a case of aberrant transcription termination? *Transcription*, 1-4
94 Lin, W.Y.*, et al.* (2014) Convergent transcription through microsatellite repeat tracts induces cell death. *Mol Biol Rep* 41, 5627-5634
95 Zhou, V.W.*, et al.* (2011) Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* 12, 7-18
96 Venkatesh, S. and Workman, J.L. (2013) Set2 mediated H3 lysine 36 methylation: regulation of transcription elongation and implications in organismal development. *Wiley Interdiscip Rev Dev Biol* 2, 685-700

97 Bumgarner, S.L., *et al.* (2009) Toggle involving cis-interfering noncoding RNAs controls variegated gene expression in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 106, 18321-18326
98 Korde, A., *et al.* (2014) Intergenic transcriptional interference is blocked by RNA polymerase III transcription factor TFIIIB in Saccharomyces cerevisiae. *Genetics* 196, 427-438
99 Colin, J., *et al.* (2014) Roadblock termination by reb1p restricts cryptic and readthrough transcription. *Molecular cell* 56, 667-680
100 Hsieh, T.H., *et al.* (2015) Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* 162, 108-119
101 O'Sullivan, J.M., *et al.* (2004) Gene loops juxtapose promoters and terminators in yeast. *Nat Genet* 36, 1014-1018
102 Singh, B.N. and Hampsey, M. (2007) A transcription-independent role for TFIIB in gene looping. *Molecular cell* 27, 806-816
103 Cai, L. and Tu, B.P. (2012) Driving the cell cycle through metabolism. *Annu Rev Cell Dev Biol* 28, 59-87
104 Xue, Z., *et al.* (2014) Transcriptional interference by antisense RNA is required for circadian clock function. *Nature* 514, 650-653
105 Hirota, K., *et al.* (2008) Stepwise chromatin remodelling by a cascade of transcription initiation of non-coding RNAs. *Nature* 456, 130-134
106 Dai, Z. and Dai, X. (2012) Antisense transcription is coupled to nucleosome occupancy in sense promoters. *Bioinformatics* 28, 2719-2723
107 Zenklusen, D., *et al.* (2008) Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* 15, 1263-1271
108 Weinberger, L., *et al.* (2012) Expression noise and acetylation profiles distinguish HDAC functions. *Molecular cell* 47, 193-202
109 Li, M., *et al.* (2015) Dynamic regulation of transcription factors by nucleosome remodeling. *eLife* 4
110 Castelnuovo, M., *et al.* (2013) Bimodal expression of PHO84 is modulated by early termination of antisense transcription. *Nat Struct Mol Biol* 20, 851-858
111 Zhang, Z. and Pugh, B.F. (2011) High-resolution genome-wide mapping of the primary structure of chromatin. *Cell* 144, 175-186
112 Thorvaldsdottir, H., *et al.* (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* 14, 178-192
113 Paulsen, M.T., *et al.* (2014) Use of Bru-Seq and BruChase-Seq for genome-wide assessment of the synthesis and stability of RNA. *Methods* 67, 45-54
114 Sun, M., *et al.* (2012) Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res* 22, 1350-1359