Transcriptional interference networks coordinate the expression of functionally related genes clustered in the same genomic loci

Zsolt Boldogköi*

Department of Medical Biology, Faculty of Medicine, University of Szeged, Szeged, Hungary

The regulation of gene expression is essential for normal functioning of biological systems in every form of life. Gene expression is primarily controlled at the level of transcription, especially at the phase of initiation. Non-coding RNAs are one of the major players at every level of genetic regulation, including the control of chromatin organization, transcription, various post-transcriptional processes, and translation. In this study, the Transcriptional Interference Network (TIN) hypothesis was put forward in an attempt to explain the global expression of antisense RNAs and the overall occurrence of tandem gene clusters in the genomes of various biological systems ranging from viruses to mammalian cells. The TIN hypothesis suggests the existence of a novel layer of genetic regulation, based on the interactions between the transcriptional machineries of neighboring genes at their overlapping regions, which are assumed to play a fundamental role in coordinating gene expression within a cluster of functionally linked genes. It is claimed that the transcriptional overlaps between adjacent genes are much more widespread in genomes than is thought today. The Waterfall model of the TIN hypothesis postulates a unidirectional effect of upstream genes on the transcription of downstream genes within a cluster of tandemly arrayed genes, while the Seesaw model proposes a mutual interdependence of gene expression between the oppositely oriented genes. The TIN represents an auto-regulatory system with an exquisitely timed and highly synchronized cascade of gene expression in functionally linked genes located in close physical proximity to each other. In this study, we focused on herpesviruses. The reason for this lies in the compressed nature of viral genes, which allows a tight regulation and an easier investigation of the transcriptional interactions between genes. However, I believe that the same or similar principles can be applied to cellular organisms too.

Keywords: transcriptional interference, antisense RNA, polycistronic RNAs, tandem genes, genomic organization, herpesvirus, pseudorabies virus, *Hox* genes

INTRODUCTION

THE RATIONALE BEHIND THE OPERATION OF TINs

The TIN hypothesis assumes that the process of transcription itself plays a regulatory role in the coordination of gene expression within a cluster of functionally linked genes. The mechanism of this coordination is based on the confrontation of the transcriptional machineries at the various overlaps formed by neighboring genes of a genetic modules (GM). The rationale behind the operation of the TINs lies in providing a straightforward genetic algorithm to coordinate the ON/OFF transcription pattern of functionally linked gene clusters, which would otherwise be possible only via sophisticated mechanisms, including continuous monitoring of the actual state of gene expression and modification of the transcription profiles of the genes accordingly. In a cluster of tandem genes, transcription of an upstream gene causes the inhibition of the downstream genes. This mechanism represents a unidirectional interaction between these genes (Waterfall model). On the contrary, the various convergent overlaps allow a two-way interplay between adjacent genes, resulting in a mutually exclusive expression of the interacting partners (Seesaw model). Besides the collision-based models, another way of interference between genes occurs through the competition between the transcriptional machineries for determining the direction of the transcription from bidirectional promoters. The collision of the transcriptional apparatuses results in a non-linear effect in gene expression. The outcome of the collisions is dependent on two important factors, which are the activity of the promoter and the efficiency of the read-through. Both processes are differentially regulated and can change in time. The outcome of the interaction between two convergent genes is controlled by the frequency of transcriptional

---

Abbreviations: AST, antisense transcript; CUT, cryptic unstable transcript; E, early; E/L, early/late; GM, genetic module; HSV, herpes simplex virus; IE, immediate-early; IRES, internal ribosome entry site; L, late; LAT, latency-associated transcript; LLT, long latency transcript; Lnc-RNA, long non-coding RNA; miRNA, micro RNA; NAT, natural antisense transcript; ORF, open reading frame; Pi, post-infection; PRV, pseudorabies virus; RNAP, RNA polymerase II; TI, transcriptional interference; TIN, transcriptional interference network; Ul, unique long; Us, unique short; UTR, untranslated region.

read-through; the genes with higher rate will win. Genes with initially higher activity and/or higher transcriptional read-through efficiency can significantly or entirely suppress the expression of genes with lower initial activity (the "Winner takes all" principle), for which the reason is at least threefold. First, the initiation step of transcription is a time-consuming process, which can therefore be effectively blocked via the rapidly advancing RNA polymerase II (RNAP) from a gene with high transcriptional activity. Second, the clash of the RNAP molecules occurs more frequently at the loci (half-space) of genes with lower activity in the case of convergent read-through overlaps. The reason for the unequal effect of the collisions is that the dislocation of the RNAP molecule during the transcription of a coding region results in the generation of functionless transcripts, but does not significantly influence translation in the read-through part of the transcripts. Furthermore, clashing of RNAPs outside the gene locus enables the production of a large population of normally terminated mRNA molecules (non-read-through transcripts). According to the Seesaw model, transcriptional collisions produce a self-regulatory automatism whereby the higher-rate transcription mechanically makes gene expression topple over from one kinetic (sub)class to another, subsequently maintaining the switched state for a given period of time. In other words, the expression of a gene is not only dependent on the presence or absence of transcription factors and the epigenetic state of the given locus, but also on the activity status of adjacent genes and the control of poly(A) recognition by factors such as ICP27 in herpesviruses (McLauchlan et al., 1992; McGregor et al., 1996). The TINs provide an auto-regulatory mechanism that controls the stepwise turning on and off of genes in a controlled manner. Furthermore, the operation of the TINs results in inverse synchronization, fine tuning, and economizing of gene expressions of adjacent genes.

GENETIC REGULATION

The recent view held that transcription is controlled by the combinatorial action of transcription factors and other regulatory proteins through binding to the enhancer and promoter sequences of the genes. This notion had its roots in the biochemical perspective that proteins comprise the structural and core catalytic framework of the cell. RNAs were thought to play subsidiary roles compared to the two truly important molecules: DNA, which carries the genetic information, and proteins, which execute nearly every function in the cell. However, this picture has been dramatically changing in recent years with the realization that RNAs have far more important functions in cells than was considered before (Weinberg et al., 2009). Alpha-herpesviruses have relatively large double-stranded DNA genomes containing more than 70 genes. These viruses have a very compact genomic structure with few introns and very short intergenic regions; therefore, they serve as ideal models to investigate eukaryotic transcription regulation. An additional beneficial feature of these viruses is that their replication cycle is rapid, requiring less than 24 h to produce infectious progeny. The program of herpesvirus gene expression is primarily controlled at the level of transcription. The lytic cycle of the herpesviruses can be characterized by the tightly coordinated expression of viral proteins belonging to different temporal groups, which are traditionally termed immediate-early (IE), early (E), early/late (E/L), and late (L) kinetic classes. The IE proteins are transcription factors that control the transcription of other viral genes. Most of the E proteins are required for the synthesis of viral DNA, while the L proteins mainly serve structural functions making up the capsid, tegument, and envelope, as well as promoting virion maturation (Roizman, 1996; Boldogköi and Nógrádi, 2003; Pomeranz et al., 2005). Besides the lytic pathway, herpesviruses can also establish a life-long latent infection in various cell types. The *Hox* genes are a set of transcription factor genes that are instrumental in regulating the timing and route of animal body formation during development. They belong to the homeotic gene family, which is characterized by a DNA sequence of approximately 180 base pairs called the homeobox encoding the homeodomain, a DNA-binding segment of proteins. *Hox* proteins act like genetic switches, turning off and on other genes with a mechanism whose key parameter is the gene dosage. The *Hox* genes themselves are regulated by more posteriorly acting *Hox* proteins and several other types of transcription factors. Many *Hox* genes have remarkably conserved functions. For example, the *Pax6* gene is involved in eye development both in insects and vertebrates, and is even expressed in the photoreceptor cell of the *Cnidarian* species. Vertebrate *Hox* genes are organized into four clusters (A, B, C, and D), while invertebrates contain only a single set of *Hox* genes. One of the most striking properties of the *Hox* genes is that they are activated in a collinear fashion; that is, the relative order of their expression along the anterioposterior axis correlates with the position of the genes in the gene cluster.

GENOME ORGANIZATION

The genes of bacteria are organized into groups of transcriptionally linked genes called operons, whereas eukaryotic genes have been traditionally thought of as having non-operon-like structures. In fact, the genes of complex eukaryotes are non-randomly distributed in the DNA; they form small "islands" on the chromosome, which are surrounded by large "deserts" of non-coding DNA. It has been demonstrated that genes residing at the same genomic locus tend to be co-expressed (Lercher et al., 2002). In an attempt to explain this phenomenon, several hypotheses have been put forward, mainly on the basis of the assumption of selection for the co-regulation of functionally related genes (Lee and Sonnhammer, 2003; Peri et al., 2003; Hurst et al., 2004), but neutral components of the co-expression are also emphasized (Michalak, 2008). The co-localized genes have evolved either from a shared ancestor by gene duplication or recruited from distinct loci of the same chromosome or from different chromosomes. It has been estimated that around 38% of the human genes and 49% of the *Caenorhabditis elegans* genes arose from gene duplication (Rubin et al., 2000; Li et al., 2001). The actual proportions may be even higher, since ancient duplications cannot be recognized through a homology search on genome databases. Tandemly arrayed genes (TAGs) are a result of tandem duplications on the chromosomes (they can be separated by spacer genes, which are not homologous to the members of TAGs). A survey of 11 vertebrate genomes revealed that TAGs account for some 14% of all genes in these genomes (Pan and Zhang, 2008). It has been demonstrated that most of the tandem arrays (60–83% in the

various species) are composed of two members, and that the majority of duplicated genes (72–94%) exhibit a parallel transcription orientation (Yi et al., 2007). By using *in silico* analysis, Al-Shahrour et al. (2010) assembled a detailed functional cartography for the genomes of eight eukaryotic model species (human, chimpanzee, mouse, rat, chicken, zebra fish, fruit fly, *C. elegans*, and *Arabidopsis thaliana*), and revealed that a large fraction of the DNA is arranged in neighborhoods of functionally related genes. Intriguingly, they further found that these functional gene clusters were generated more frequently through the reorganization of the genome than through gene duplication. Genes of a cluster may form interaction networks, in which the proteins interact with each other directly by forming multimeric proteins, or serving as upstream and downstream components in the signaling pathways (Hurst et al., 2004; Teichmann and Veitia, 2004), or through the mutual regulation of their expressions, forming gene networks of transcription factor genes (e.g., *Hox* genes). Alternatively, co-localized genes can feature in the same biochemical pathway without their protein products participating in direct interactions, e.g., enzymes taking part in a common metabolic pathway, β-globins carrying oxygen in the muscle, or in fetal or adult blood, or tRNAs and MHC proteins involved in the translation and immune defense, respectively. In either case, the clustering of genes appears to confer a selective advantage for the formation and maintenance of gene clusters. Besides the coordinated gene expression, the selective pressure may also favor the long-term linkage (co-inheritance) of clustered genes. For example, the existence of large regions of linkage disequilibrium has been reported among inbred mouse lines, which were shown to be correlated with the biological function (Petkov et al., 2005). Additionally, it has been reported that clusters of homologous genes in *C. elegans* tend to comprise species-specific gene families that play roles in detoxification and immunity (Thomas, 2006). In the genome of pseudorabies virus (PRV), the functionally linked genes appear to be clustered: the genes *rr1* and *rr2* are involved in the synthesis of viral DNA (Kaliman et al., 1994), the genes *ul18–21* in capsid maturation (see references in Tombácz et al., 2009), etc. (**Figure 1**). However, the adaptive origin and maintenance of genetic networks in complex eukaryotes has been called into question in several publications (e.g., Lynch, 2007a). It is claimed that the "strong" effect of natural selection and a "weak" effect of genetic drift in unicellular organisms and viruses could have led to the emergence of functional gene clusters for the coordinate regulation of metabolic genes in prokaryotes and unicellular eukaryotes (Lynch and Conery, 2003; DeLong et al., 2010), and of replication and host recognition genes in large DNA viruses (e.g., Shackelton and Holmes, 2004). However, the evidence appears to suggest that the effect of genetic drift in multicellular eukaryotes is more considerable than that of natural selection, which compromises the formation and maintenance of functional gene clusters and their interactions in regulatory networks (Lynch, 2007a,b). Further data are needed to clarify this issue.

In sum, the tandem arrangement of functionally related genes on the DNA appears to represent a common organization principle in both cellular organisms and viruses with large genome sizes. Mammalian *Hox* genes are all arrayed in a tandem orientation,

while the arrangement of some *Hox* genes in arthropods (insects and crustaceans) has been reversed during evolution (**Figure 2**). In various arthropod species, genomic reorganization events have generated antiparallel gene pairs within the *Hox* gene cluster (**Figure 2**). Herpesviral genes exhibit a characteristic modular organization, which includes two convergently oriented gene clusters, each containing parallelly overlapping genes (**Figure 1**). Questions arise as to whether this modular genomic organization is common in cellular organisms too and if so, whether it has any functional significance or represents a mere curiosity without any apparent function.

NON-CODING RNAs
Genome-wide transcription of non-coding RNAs
Current technological advances have allowed the investigation of overall expression of human and mouse genomes with an exceptional resolution. These analyses have revealed that the major parts of these genomes are transcribed and that protein-coding sequences account for only a small proportion of the total transcriptional output (Bertone et al., 2004; Cheng et al., 2005; Birney et al., 2007; Kapranov et al., 2007). There is still ongoing debate as to whether this pervasive transcription represents anomalies for the most part or if the non-coding RNAs (ncRNAs) have functions that have not yet been identified (Mattick, 2004; Struhl, 2007; Ebisuya et al., 2008; Kapranov and St. Laurent, 2012). The transcriptional noise hypothesis implies that cells can benefit from allowing some random transcription to occur rather than suppressing non-specific transcription, and this appears to be supported by the fact that individual ncRNAs are transcribed at much lower levels than individual mRNAs (Louro et al., 2009). In the past couple of years, multiple types of these "dark matter RNAs" have been discovered with diverse functionality. This unforeseen level of transcriptome complexity has led to the realization that ncRNAs comprise a formerly hidden layer of genetic programming in multicellular organisms. NcRNAs appear to form complex regulatory circuits, which control gene expression at multiple levels (Mattick, 2010).
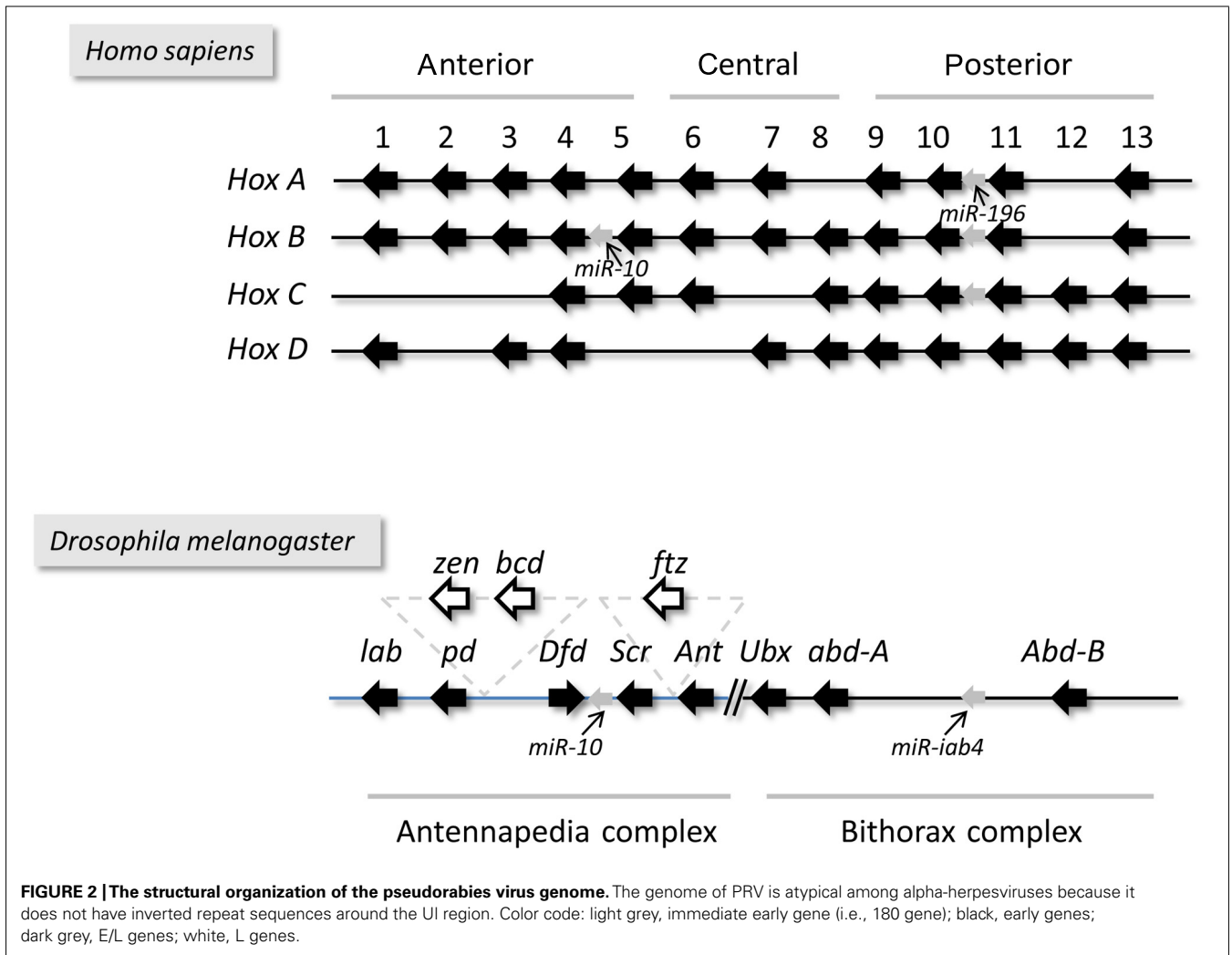
Natural antisense transcripts
Natural antisense transcripts (NATs) represent a group of ncRNAs that have complementarity to other transcripts. Many loci of chromosomes contain transcription units on both DNA strands. Often, one transcript encodes a protein, whereas the RNA from the complementary DNA strand is non-coding. These latter RNA molecules are termed *cis*-NATs. In contrast, *trans*-NATs are encoded outside the genomic loci that specify the target mRNAs. *Cis*-NATs are entirely complementary to their mRNA partners, whereas *trans*-NATs generally display imperfect complementarity to their target mRNAs (Li et al., 2006). NATs can occur in promoters, enhancers, introns, exons, intergenic sequences, and untranslated regions of the genome (Li and Ramchandran, 2010). NATs have been shown to exert control on their target genes at various levels, including transcription, RNA processing, splicing, stability and cellular transport of RNAs, and translation. The most prominent representatives of *trans*-NATs are the microRNAs (miRNAs) and the endogenous small interfering RNAs (siRNAs), although these transcripts can be encoded
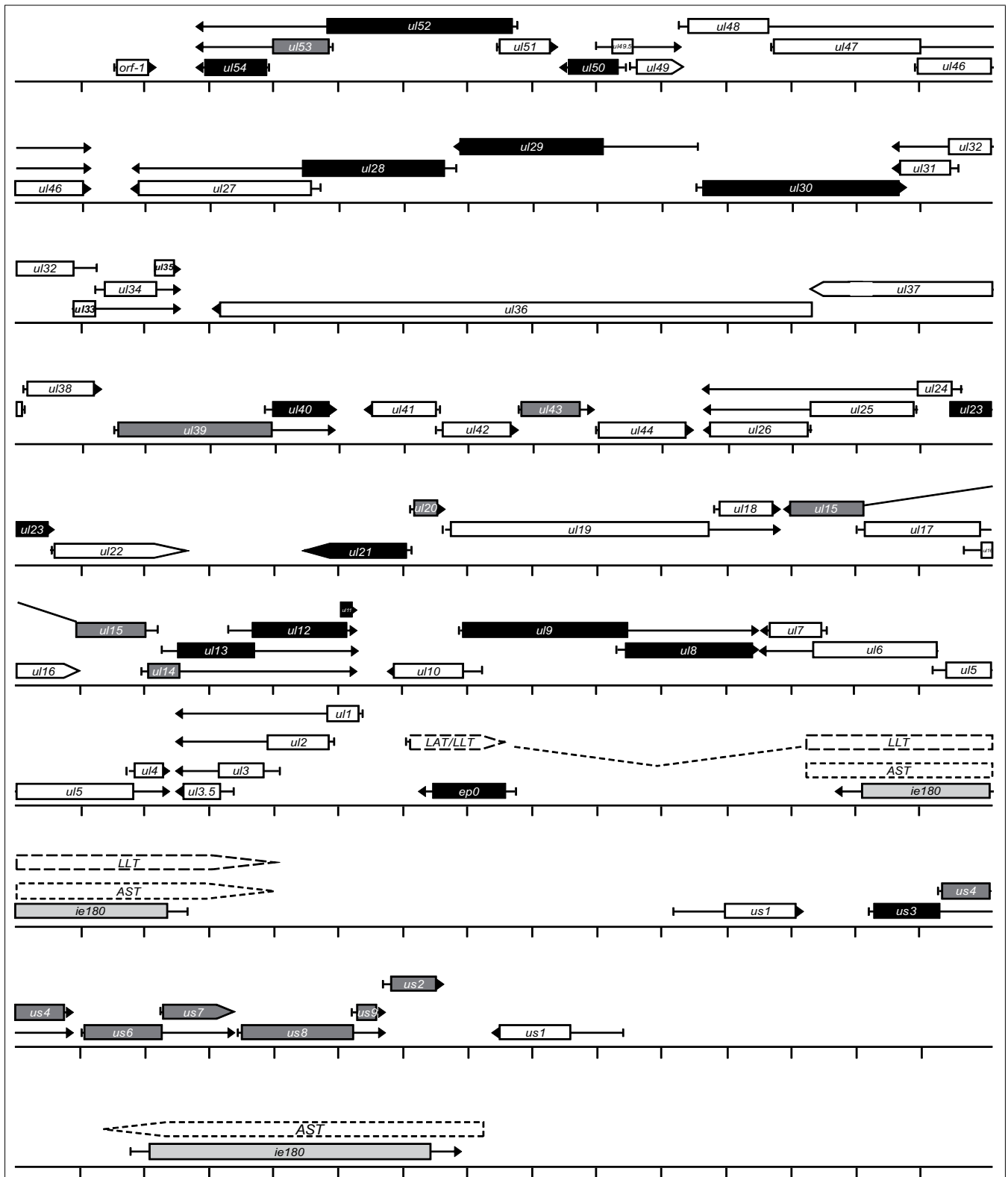
FIGURE 1 | The structural organization of the *Hox* clusters. *Drosophila* has a single set of *Hox* genes termed the Homeotic complex, which is made up of two clusters, the *Antennapedia* and the *Bithorax* complexes.

Vertebrates have four sets of *Hox* clusters generated by two rounds of polyploidization events that occurred prior to the divergence of jawless and jawed vertebrates.

FIGURE 2 | The structural organization of the pseudorabies virus genome. The genome of PRV is atypical among alpha-herpesviruses because it does not have inverted repeat sequences around the Ul region. Color code: light grey, immediate early gene (i.e., 180 gene); black, early genes; dark grey, E/L genes; white, L genes.

in *cis*, too. MiRNAs comprise short non-coding transcripts, which are post-transcriptional regulators that bind to complementary sequences on target mRNAs and modulate gene activity by interfering with mRNA stability or translation (Carthew, 2006; Bartel, 2009). Small RNAs can repress transcription in two alternative ways: either through chromatin modification (DNA methylation or histone remodeling) or, alternatively, through targeting the transcription start sites, thereby blocking procession and/or recruiting of RNAP (Janowski et al., 2005; Napoli et al., 2009). The human genome encodes over 1,000 different miRNAs, which are supposed to target at least 60% of the mammalian genes (Friedman et al., 2009).

Long non-coding transcripts

Recent attention has focused on miRNAs. However, the most abundant and least annotated class of ncRNAs is the long non-coding RNAs (lnc-RNAs; Mattick and Makunin, 2006), which are defined as transcripts exceeding 100 nucleotides without containing functional open reading frames (ORFs). The mouse transcriptome was recently calculated to contain approximately 180,000 transcripts; however, only 20,000 of them encode proteins

(Carninci et al., 2005). It has been shown that a large proportion of the human DNA also encodes lnc-RNAs (Wilusz et al., 2009). Many protein-encoding genes specify lnc-RNAs transcribed from the sense (plus) DNA strands as templates, which are called antisense lnc-RNAs. Long overlapping transcripts can include two protein-encoding mRNAs, an mRNA/lnc-RNA pair or two lnc-RNAs. Transcriptome analysis of antisense lnc-RNAs and their mRNA partners revealed frequent concordant regulation of expression (Katayama et al., 2005). Furthermore, it has been shown that sense-antisense pairs often exhibit reciprocal expression patterns (Lehner et al., 2002), indicating interplay between these two RNA molecules. The antisense lnc-RNAs have been reported to suppress the expression of their mRNA counterparts in cultured human cells (Izant and Weintraub, 1985). While several reasonable hypotheses have been put forward in an attempt to explain how the antisense lnc-RNAs regulate mRNA expression, there is little experimental evidence supporting these explanations (Stolt and Zillig, 1993; Rinn et al., 2007; Roeszler et al., 2012). Data generated over the past few years have provided insights into these mechanisms, which involve, among others, remodeling of the chromatin structure at target loci by antisense lnc-RNAs

(Morris and Vogt, 2010) and transcriptional interference (TI) between the transcriptional machineries at the overlapping DNA region (Martens et al., 2004).

Transcription from bidirectional promoters

Current sequencing and annotation methods have revealed that divergent organization of coding genes is widespread in various genomes (Trinklein et al., 2004; Hermsen et al., 2008). It has also been pointed out that eukaryotic promoters are intrinsically bidirectional (Wei et al., 2011). The function of bidirectional promoters has long been held to provide joint coordination of expression of two divergent genes, whose protein products are typically involved in the same biochemical function (Meier and Straus, 1993). However, it has recently been demonstrated that bidirectional promoters, are involved in the generation of ncRNAs called cryptic unstable transcripts (CUTs; Neil et al., 2009). CUTs have been described in various eukaryotic species, ranging from yeasts (Xu et al., 2009) to mammals (Core et al., 2008; Seila et al., 2008). There have been speculations on the potential function of CUTs. It is possible that the process of CUT transcription results in the nucleosomes being repelled from the promoters, thereby allowing either prolonged access of the transcription factors to the *cis*-regulatory elements (Mavrich et al., 2008), or the linking of the promoter region with the 3′-end of the gene (Tan-Wong et al., 2009). Alternatively, the CUTs themselves could be functional, through RNA interference (Han et al., 2007) or Argonaute-independent (Wang et al., 2008) silencing pathways.

Antisense RNA expression in herpesviruses

MiRNAs have been shown to be encoded in all of the three subfamilies of herpesviruses, including herpes simplex virus (HSV; alpha-herpesvirus, Umbach et al., 2008), human cytomegalovirus (beta-herpesvirus; Grey et al., 2005), and Epstein–Barr virus (gamma-herpesvirus; Pfeffer et al., 2004). Although the precise functions of most viral miRNAs remain to be ascertained, an increasing number of their viral and cellular targets have been experimentally confirmed. Herpesviral miRNAs are supposed to play a pivotal role in the latent/lytic switch during the viral life cycle, the evasion of host immune responses, and preventing apoptosis of the host cell (Li et al., 2010). Numerous long *cis*-antisense transcripts have been identified in human cytomegalovirus (Zhang et al., 2007). The latency associated transcript (LAT) of HSV was described as the first viral *cis*-NAT (Stroop et al., 1984). The LAT region encodes multiple transcripts, including the 8.3-kb primary transcript and two stable introns of 1.5 and 2.0 kb (Zwaagstra et al., 1990; Farrell et al., 1991). The LAT plays a key role in establishing and maintaining viral latency (Perng et al., 2000). A spliced 8.4-kb antisense RNA, termed long latency transcript (LLT), has been shown to be synthesized from the complementary DNA strand of *ie180* and *ep0* genes driven by the LAT promoter of PRV, a veterinary alpha-herpesvirus (Cheung, 1989; Priola and Stevens, 1991). The antisense transcript (AST) RNA gene overlapping the *ie180* gene has been suggested to be under the control of a separate promoter, termed antisense promoter (ASP; Boldogköi et al., 2000; **Figure 4**). Antisense RNA expression has also been reported in

some other HSV genes (Ward et al., 1996; Chang et al., 1998; Jovasevic and Roizman, 2010). Furthermore, we have detected an overall antisense RNA transcription in PRV using real-time RT-PCR techniques (unpublished result). Together, antisense transcripts appear to be a pervasive feature of herpesviruses, suggesting that they are fundamental components for the regulation of the viral life cycle.

Antisense RNAs in the Hox gene clusters

Mammalian *Hox* clusters have been shown to contain five miRNA genes intercalated at two homologous positions (Kosman et al., 2004). Furthermore, Mainguy et al. (2007) reported the existence of *cis*-encoded antisense RNAs in the *Hox* gene cluster. Antisense transcripts were found to represent approximately 38% of the total spliced transcripts. It has been shown that *HOTAIR*, an antisense lnc-RNA transcribed from the *HOXC* locus, affects gene expression in the *HOXD* locus residing on a different chromosome (Rinn et al., 2007). Indeed, targeting *HOTAIR* transcripts by RNA interference results in the transcriptional activation of the *Hox*D locus (Rinn et al., 2007), indicating an inhibitory role of these transcripts. *HOTAIR* has been shown to be a regulator of the chromatin state by providing binding surfaces for the assembly of histone modification enzymes (Tsai et al., 2010).

POLYCISTRONIC RNAs

Most of the mRNAs found in eubacteria and archaea is polycistronic, and are transcribed from the structural genes of an operon. These RNAs contain ribosome-binding sites at an upstream location of each gene. On the other hand, eukaryotic mRNAs have only a single site for the initiation of protein synthesis; therefore, with a few exception (e.g., *Hox* genes), they are thought to be monocistronic. *Hox* genes are tandemly arranged in an order collinear with their expression along the anterior–posterior axis. Extensive polycistronism has been reported in mammalian *Hox* gene clusters, where *Hoxc6*, *Hoxc5*, and *Hoxc4* are co-transcribed (Simeone et al., 1988). The mammalian *Hox* clusters are estimated to display at least seven polycistronic clusters (Mainguy et al., 2007). Notably, polycistronic *Hox* transcripts have also been described in a number of crustacean species (Shiga et al., 2006), indicating their importance in diverse evolutionary lineages. Several downstream *Hox* genes on polycistronic mRNAs have been reported to be translated by means of internal ribosomal entry (Oh et al., 1992) or alternative splicing (Simeone et al., 1988; Mainguy et al., 2007). Tandem herpesviral genes are embedded in each other in such a way that the transcription from the upstream genes continues across the transcription termination signals with certain efficiencies. This process is regulated by the ICP27 viral protein, which helps recognize the weak poly(A) signals by RNAP (McGregor et al., 1996); that is, in the absence of ICP27, the parallelly embedded genes will produce long, polycistronic mRNAs with common 3′-termini. Many viral species solved the problem of translating multiple messages from a polycistronic mRNA by developing various mechanisms such as: alternative splicing of primary RNA products in HIV (Schwartz et al., 1990); leaky scanning of *vpu* and *env* transcripts in HIV, as well as the OFR1 and ORF2 genes in rotavirus; ribosomal frame shifting of *gag* and *gag-pol* genes in HIV; and the utilization of

internal ribosomal entry sites (IRES) in picornaviruses and hepatitis C virus (reviewed by Stacey et al., 2000). However, none of the above mechanisms has been described in herpesviruses so far. Thus, it is presently unknown as to whether downstream ORFs of polycistronic RNAs are translated in these viruses (Sokolowski et al., 2003).

### INTERFERENCE BETWEEN THE TRANSCRIPTIONAL MACHINERIES

Transcription regulation in eukaryotic cells is thought to be controlled by transcription factors through binding to the *cis*-regulatory response elements, thereby promoting or inhibiting the efficiency of the pre-initiation complex assembly or the activity of RNAP. Reports have been recently published on a different regulatory mechanism based on the interference between the transcriptional machineries of adjacent genes. The phenomenon of transcriptional collision of RNAP molecules in oppositely oriented genes has been demonstrated in yeast (RHO1 and MRP2 genes, Peterson and Myers, 1993; GAL7, GAL10, and TI genes, Prescott and Proudfoot, 2002). These convergent yeast gene pairs overlap, that is, their transcription termination sequences are embedded within each other's transcribed DNA regions. In these examples, the elongation phase of transcription of both genes was found to be affected, while the transcription initiation was unaltered. In another publication, transcriptional collisions were reported between RNAPs transcribing mRNAs and *cis*-NATs in several murine and human genes (Osato et al., 2007). However, no promoters were identified experimentally or *in silico*, which would control the expression of *cis*-NATs. A similar mechanism was also described in yeasts: lnc-RNA molecules transcribed from a non-protein encoding genomic region were shown to directly interfere with the binding of transcription factors to the promoter of adjacent SER3 gene, thereby preventing this gene from being expressed (Martens et al., 2004). The consequence of the collision of RNAP molecules has not yet been investigated. A theoretical possibility is that one or both RNAPs are released from the template. Alternatively, confronted RNAP molecules may stall and be unable to advance further. Any of the above mechanisms would severely impede the transcription of both genes. TI between the mRNAs and antisense lnc-RNAs has been shown to regulate key developmental decisions, such as where *Hox* genes are expressed (Petruk et al., 2006). Bithoraxoid (*Bxd*) is an lnc-RNA found in *Drosophila*, which silences the expression of the Ultrabithorax (*Ubx*) gene through TI. Furthermore, the amount of transcripts synthesized from tandem genes in the yeasts genome decreases when the transcriptional termination sequences of the upstream gene are removed (Osato et al., 2007), indicating a negative effect of the transcriptional read-through on gene expression. Interaction between the transcriptional apparatuses has not been described in herpesviruses to date.

HYPOTHESIS AND DISCUSSION
#### THE TRANSCRIPTIONAL INTERFERENCE NETWORK HYPOTHESIS

The TIN hypothesis has been put forward in an attempt to provide a conceptual framework for understanding the potential role of the widespread occurrence of gene clusters with tandemly arrayed members, and the overall antisense RNA expression in biological systems belonging to various taxonomic groups. The TIN hypothesis asserts that genes are organized into structural and functional units, termed GMs, which are comprised of tandemly or convergently oriented genes, or more typically, combinations of them. The hypothesis suggests that the transcription of neighboring genes overlaps in different ways (head-to-tail, head-to-head, tail-to-tail, or full overlap; **Figure 3**), resulting in the confrontation of their transcriptional machineries at the overlapping region. Furthermore, it is proposed that this form of genetic regulation occurs at a genome-wide scale and, therefore, plays an essential role in controlling global gene expression in both cellular organisms and viruses. Convergently overlapping genes can be protein-coding genes or RNA genes, or the combination of them. True convergently overlapping protein-coding genes, especially those with ORF overlaps, are rare in various genomes. However, the TIN hypothesis proposes that in many gene pairs, transcription from convergently positioned genes is not terminated with perfect efficiency at their poly(A) sequences. Instead,
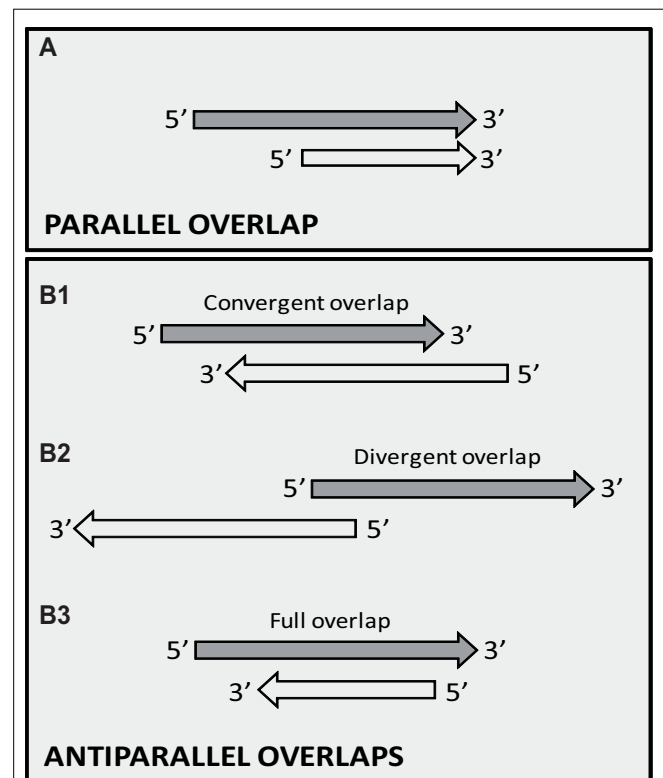
FIGURE 3 | Possible overlaps producing antiparallel transcripts.
(A) Tandem (parallel, head-to-tail, 5′–3′) overlap. Transcription of the upstream genes of a tandem gene cluster often fails to stop at their poly(A) signal, thereby producing long polycistronic RNA molecules. **(B)** Convergent (antiparallel) overlaps. (**B1**) Head-to-head (5′–5′) overlap between two divergently positioned genes. (**B2**) Tail-to-tail (3′–3′) overlap between convergently oriented genes. Two opposing genes can overlap at their 3′-UTR regions (a) or can exhibit a 3′-UTR–ORF (b), or ORF–ORF overlap. (**B3**) Transcriptional read-through overlap occurs when the major poly(A) signal occasionally fails to terminate by the advancing RNAP (a). The use of alternative polyadenylation provides a mechanism similar to that of transcriptional read-through for the spatiotemporal control of gene expression (b).
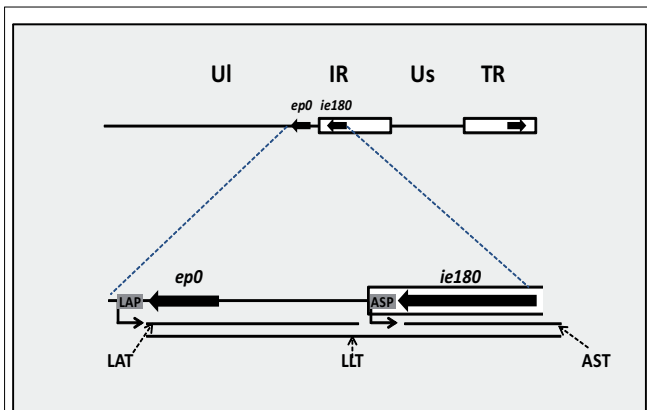
FIGURE 4 | Antisense transcripts regulating the two key transcriptional activators of pseudorabies virus. The immediate early 180 (IE180) and early protein 0 (EP0) proteins are the major regulators of global gene expression of PRV. The latency-associated transcript (LAT), the antisense transcript (AST) and the long latency transcripts (LLT) are RNA molecules transcribed from the complementary DNA strands of these genes from their own promoters. The antisense transcripts are driven by LAP (latency-associated transcript promoter) and ASP (antisense promoter).

transcription occasionally continues at the locus of the opposing gene, which leads to the generation of long RNA molecules containing mRNA sequences from the given gene and antisense RNA sequences from the oppositely oriented gene. The process of alternative polyadenylation can also lead to differential transcriptional overlaps. The long transcripts, generated by the read-through mechanism, appear to be the mere by-products of this process; however, it cannot be excluded that they are further utilized at other levels of genetic regulation. Likewise, downstream-positioned genes of polycistronic RNAs, if untranslated, are also superfluous by-products of the transcriptional read-through between tandemly oriented genes. The potential use of these polycistronic RNAs, if at all, is unknown. Another way of TI is based on the competition of transcription apparatuses on the divergent promoters to determine the direction of transcription. According to the TIN hypothesis, transcription of neighboring genes can depend on each other in either a one-way (parallelly overlapping genes) or two-way direction (convergently overlapping genes). The interaction between the transcriptional machineries provides a mechanism for the coordinate control of the expression of functionally linked genes using a simple genetic algorithm: "keep a gene silent while the adjacent gene is transcribed through the process of transcription itself," which would be a very complicated task to accomplish in a large gene cluster by means of transcription factors alone. A TIN regulates the transcription switch between two or more genes and acts thereafter to maintain the turned-on state of a gene and the turned-off state of the neighboring genes for a given period of time. The interdependence of gene activities provides an in-built automatism for producing a tightly choreographed succession of expression patterns of co-localized genes. To coordinate gene expression at a broader scale, GMs themselves can also interact with each other in various ways, including the TI within the divergent overlaps of upstream genes of adjacent GMs or the transcriptional

read-through from distant genes belonging to other GMs. The operation of TINs will now be discussed in the alpha-herpesviruses, especially in the PRV, which is a useful model organism for studying the molecular pathogenesis of herpesviruses (Boldogköi et al., 2004) and is also a beneficial tool in neuroscience (Boldogköi et al., 2009).

The Waterfall model

Tandemly overlapping genes. Functionally coupled genes are often arrayed in tandem along a DNA segment in various cellular organisms and viruses. Many of these genes have been generated by gene duplication, but the conservation of this architecture over a long evolutionary timescale requires an explanation. In an attempt to explain this phenomenon, Huvet et al. (2007) proposed a hypothesis that was based on the collision of the replication and transcription machineries; however, the experimental verification of this model, at least in human cells, failed (Necsulea et al., 2009). The Waterfall model of the TIN hypothesis provides an alternative, but not necessarily exclusive, explanation for this subgenomic organization, which is based on the TI between tandemly overlapping genes. Accordingly, the upstream-positioned genes exert an inhibitory effect on the expression of downstream genes through their unterminated transcription. In herpesviruses, this process is regulated by the ICP27 protein, which helps recognize the weak transcription termination signals of the RNAP molecules (McGregor et al., 1996). As a result of transcriptional read-through, parallelly oriented genes generate long polycistronic mRNAs with common 3′-termini. Most of the tandem gene clusters of PRV are composed of genes with similar transcription kinetics (Tombácz et al., 2009). We assume, however, that the transcription of these genes is slightly shifted in time relative to each other, which is difficult to verify experimentally, even with highly accurate techniques such as real-time RT-PCR. The reason for this is that downstream genes are located on both monocistronic and polycistronic transcripts, but in the latter case, these genes are very likely untranslated; that is, the "effective" (translated) amount of transcripts is unknown. The transcriptional machineries of tandem genes collide and, as a result, transcription of the upstream genes suppresses the expression of downstream genes by inhibiting the assembly of their transcription pre-initiation complex. Thus, a gene with early expression kinetics can effectively block the transcription of a downstream gene. Following the cease of the blocking effect by the upstream gene (termination of the transcription or the read-through), the downstream gene is transcribed with a delayed kinetic shift. Likewise, an upstream gene with late expression kinetics can terminate the transcription of downstream-positioned early genes. Poly(A) site selection, as a form of post-transcriptional regulation, has been described to be influenced both by the inherent poly(A) site "strength" and by the concentrations of processing factors (Lassman and Milcarek, 1992). Together, the expression properties of a gene are not solely determined via the interaction between its *cis*-regulatory elements (promoters and enhancers) and transcription factors, but also via the activity or silence of upstream genes and the control of the transcriptional read-through efficiency. In the Waterfall model, the downstream genes do not exert any direct effect on the transcription of upstream genes;

therefore, this type of interaction results in a one-way dependence of gene expression (here comes the name "waterfall"). Nonetheless, producing polycistronic RNAs can slow down the transcription of upstream genes due to the longer time needed for the production of polycistronic RNAs compared to monocistronic ones. The downstream genes of *Hox* clusters are controlled by the transcription factors encoded by the upstream genes, thereby resulting in a spatiotemporally well-controlled stepwise activation of these genes. According to the Waterfall model, the same pattern of gene expression can be achieved through the operation of TINs.
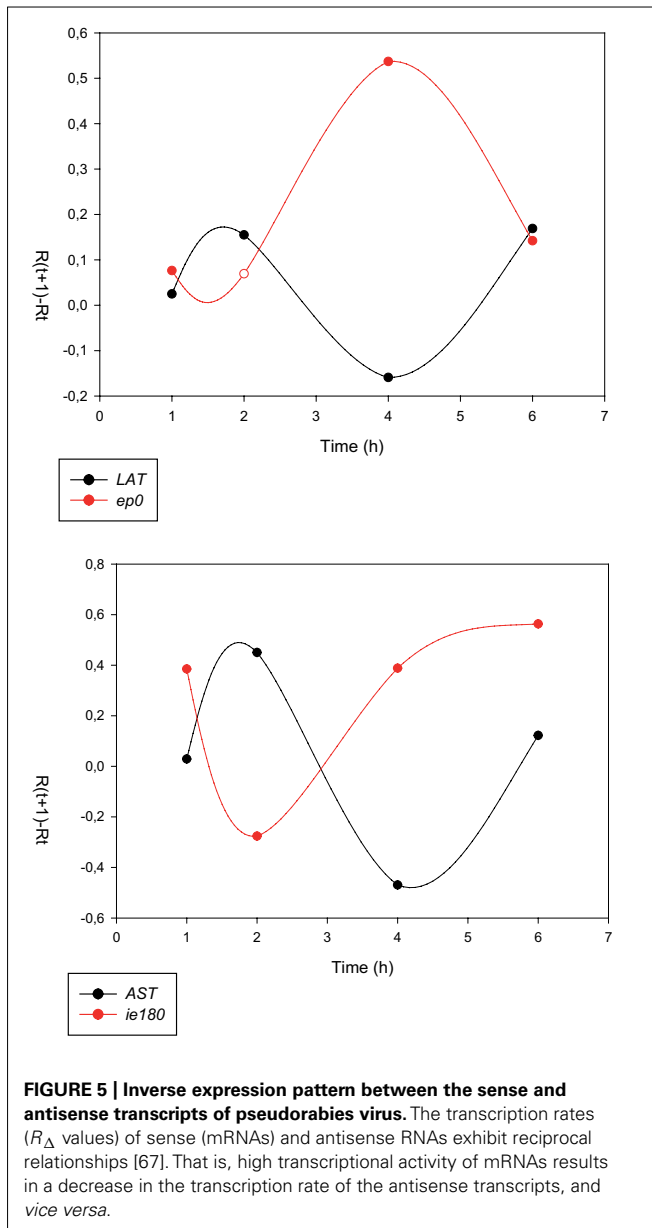
The Seesaw model

The Seesaw model of the TIN hypothesis proposes a mutually interdependent mode for regulating the expression of neighboring genes. The seesaw mechanism is based on the head-on collision of RNAP molecules during the transcription of two or more overlapping genes positioned in convergent or divergent orientations relative to each other.

Divergently overlapping genes.   Divergently oriented genes can often overlap and generate transcripts with 5′ overlapping regions (Wei et al., 2011). All of the 11 divergent gene pairs of PRV transcriptionally overlap in the following manners: 5′–5′-UTR (untranslated region) overlap in the *ul41–ul42* gene pair; 5′-UTR–ORF overlap in the *ul37–ul38* gene pair; and ORF–ORF overlaps in seven PRV gene pairs, intriguingly, four of them overlap with two nucleotides at the AUG translation initiation codons. Furthermore, 2 of the 11 divergent PRV genes have a common TATA-box (see in Gu et al., 2009). The Seesaw model predicts that the transcription of one gene inhibits the transcription of the divergently overlapping partner by preventing the assembly of its transcription initiation complex and/or dislocating its RNAP molecules from the DNA template. At a later phase of viral replication, the direction of transcription alters due to the formation of a new combination of transcription factors and/or the activation or repression of these factors. Therefore, the gene, which was earlier suppressed, is transcribed, which then results in the inhibition of the transcription of the opposing gene (hence, the name "seesaw"). This mechanism operates according to the "Winner takes all" principle, where the "winner" is the gene with the higher initial rate of expression. Our analysis revealed that four of the nine divergent overlaps occur between the L genes (*ul5–ul6*, *ul32–ul33*, *ul37–ul38*, and *ul41–ul42*), one between the E/L genes (*ul14–ul15*), and four between an E and an L gene (*ul9–ul10*, *ul23–ul24*, *ul49.5–ul50*, and *ul51–ul52*), while there is no overlap between the E genes (Tombácz et al., 2009). Although overlapping of divergent genes contributes to genome compaction by eliminating DNA sequences between genes, we believe that its function also includes a TI-based regulation of gene expression. The Seesaw model predicts that divergently overlapping genes all belong to different kinetic classes or subclasses using a finer scale categorization. Additionally, the identification of *cis*-acting elements, both experimentally and *in silico*, is much more difficult than that of the coding sequences in eukaryotic cells; therefore, the real extent of potential divergent overlaps is currently unknown.

Convergently overlapping protein-coding genes.   The presence of overlaps between oppositely oriented genes has been described in diverse species, such as yeast (Peterson and Myers, 1993; Prescott and Proudfoot, 2002), *C. elegans* (Osato et al., 2007), and PRV (Szpara et al., 2011). Convergent genes can produce mRNAs with an overlap in their 3′-UTRs. However, very few convergent ORF–ORF overlapping has been detected, possibly because the determination of protein structure imposes a severe constraint on codon usage in the coding DNA strand, leaving limited variations of the evolution of protein-coding sequences on the other DNA strand. Convergent PRV genes are thought to be non-overlapping with their oppositely oriented partners. The only true convergently overlapping gene pair is the *ul30* and *ul31* (175-nucleotide total overlap with 75-nt overlap in the ORFs). Apparently, the transcriptional machineries of the opposing genes must collide at the overlapping region unless their expressions are temporally well separated. According to the Seesaw model, TI at the overlapping region plays a pivotal role in the temporal separation of the expression of overlapping genes. The underlying mechanism is based on either the dislocation of RNAP molecules, which results in the inhibition of the assembly of transcription pre-initiation complex and/or the premature termination of the transcription. This process provides an automatism through which a gene with a high expression rate suppresses the expression of the opposing gene(s) with lower activity. The alteration of the transcriptional direction in a locus is initiated by transcription factors (through a change in their composition and/or activity).

Convergent overlap between a protein-coding gene and an RNA gene.   The overlaps between protein-coding genes and RNA genes have been described in several cellular organisms and viruses (Izant and Weintraub, 1985; Katayama et al., 2005; Lapidot and Pilpel, 2006). In this type of interaction, there is no constraint for protein coding in one of the overlapping partners, which allows the formation of long overlaps. There is increasing recognition that the antisense RNAs within the *Hox* cluster play an important role in the regulation of gene expression (Mainguy et al., 2007). The early protein 0 gene (*ep0*)-LAT and the *ie180*-AST transcript pairs serve as examples for this type of transcriptional overlap in PRV. Additionally, the LLT overlaps both *ep0* and *ie180* mRNAs (**Figure 4**). The HSV LAT has been shown to encode miRNAs (Umbach et al., 2008). It is still unknown whether the sole function of these antisense transcripts is being the precursor of miRNAs. In an earlier report, we investigated the alteration in the amount of sense (mRNA) and antisense transcripts during viral infection using real-time RT-PCR, and obtained an inverse relationship between them (Tombácz et al., 2009; **Figure 5**), indicating that the transcription rates of these molecules were interrelated. The LAT and AST transcripts of PRV are considered non-coding. Intriguingly, however, the complementary DNA strands contain long ORFs, which had led to the speculation of their potential in encoding antisense proteins. However, we showed that these ORFs were generated by the selective accumulation of G and C bases in the silent codon positions of these genes, thus eliminating stop codons from the fourth antisense reading frame on the complementary DNA strand

FIGURE 5 | Inverse expression pattern between the sense and antisense transcripts of pseudorabies virus. The transcription rates (R_Δ values) of sense (mRNAs) and antisense RNAs exhibit reciprocal relationships [67]. That is, high transcriptional activity of mRNAs results in a decrease in the transcription rate of the antisense transcripts, and *vice versa*.

(Boldogköi et al., 1995). We proposed that these antisense ORFs are the simple by-products of a selective GC-pressure, which, if they have any function, are involved in something else other than the generation of protein-coding sequences on the complementary DNA strand.

Convergent read-through overlap between protein-coding genes.
Many genes in an organism are oriented in an opposite direction relative to each other. The herpesvirus genome appears to be organized in a modular fashion, each module (GM) being composed of two convergently oriented tandem gene clusters. The Seesaw model proposes the existence of a mechanism that allows an extensive transcriptional read-through across the oppositely oriented genes, which is the result of the inefficient termination of transcription at the terminal poly(A) signals of the nested gene clusters.

This mechanism is supposedly under the control of the ICP27 protein, resulting in the generation of long transcripts containing sequences that are complementary to the minus DNA strand of a gene and antisense sequences transcribed from the plus strand of the opposing gene. While most of the genes within the tandem gene cluster belong to the same kinetic class, the convergent partners usually belong to different temporal classes (Tombácz et al., 2009). A special type of overlap occurs in the *ul15*, *ul16*, and *ul17* genes, where the two latter genes are embedded in the former one in a convergent orientation (**Figure 1**). The *ul15* gene has to produce a long transcript encompassing the other two genes, which are removed during mRNA processing by splicing. The short Us region appears to be differently organized than the long Ul region of the PRV genome: all the Us genes are tandemly oriented. However, the *us1* gene located on the inverted repeats is transcribed in an opposite direction (**Figure 1**). The Us genes are transcribed in an E/L kinetics (except *us3*, which is an E gene), while the *us1* (*icp22*) appears to be an L gene in PRV (Tombácz et al., 2009) and an IE gene in HSV (Prod'hon et al., 1996). It is predicted that a long transcript encompassing the entire Us region is generated from the *us1* gene as a result of transcriptional read-through from this gene. The "Winner takes all" principle can also be applied to the convergent read-through mechanisms. The long transcripts containing the antisense RNA and the mRNA of the opposite gene can be separated by cleavage at the poly(A) site of the mRNA or by removing the antisense transcript by RNA splicing.

The Extended Seesaw model
The TIN hypothesis suggests the following scenario for controlling herpesvirus replication through the TINs. We considered prototypic GMs, which are composed of two opposing tandem gene clusters. Each tandem gene cluster is composed of genes with different kinetic subclasses and the two convergent gene clusters belong to different kinetic classes (**Figure 6**). However, real GMs can be built up somewhat differently: (1) the first step of the transcription cascade of PRV is the production of IE180 mRNA, which does not need *de novo* viral protein synthesis. The transcription of this transactivator is regulated, among others, by the RNA genes AST and LLT through a seesaw mechanism (**Figure 5**), and by its own protein product (auto-feedback; Tombácz et al., 2009). (2) Next, several E genes are induced by the IE180 transactivator in association with specific cellular factors. At this stage of viral infection, the upstream E genes inhibit both their tandem gene partners through a waterfall mechanism and the convergent L genes through a seesaw mechanism. Furthermore, these upstream E genes also inhibit the expression of the divergently oriented genes belonging to different kinetic (sub)classes. The *ie180* and the *ep0* genes encode the two key transcriptional activators of PRV. Therefore, these genes and also their antisense partners require a complex regulation, which could explain why LAT, LLT, and AST are controlled by their own promoters. This is in contrast to the rest of the antisense RNAs, which, according to the Seesaw model, are controlled by the promoter of opposite genes and produced via a transcriptional read-through mechanism. (3) In the next step, the ICP27 protein and/or other factors recognize the poly(A) signal of the upstream E genes, thereby allowing the transcription of
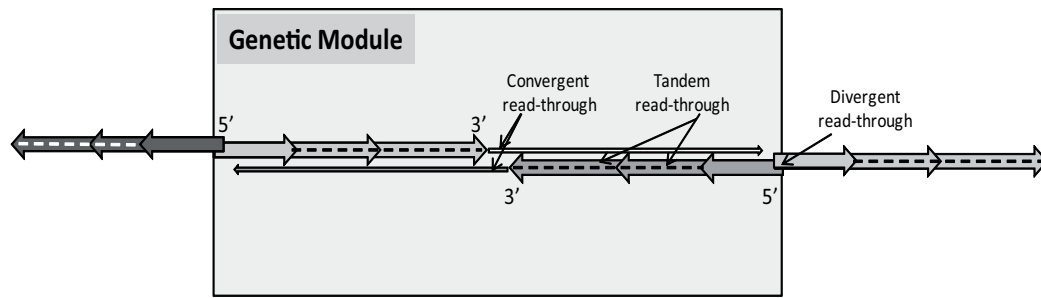
FIGURE 6 | A prototypical genetic module of the herpesviruses. The PRV genome exhibits a modular design composed of tandemly overlapping genes, as opposed to another tandem gene cluster in the prototypical Genetic Modules. Many tandem genes are assumed to be transcriptionally overlapping in a parallel fashion (e.g., *ul42*, *ul43*, and *ul44*); however, no reports have been published on this to date.

downstream E genes from their own promoters. (4) Subsequently, gene expression swaps from the E to L stage (capsizing the arms of the seesaw), in which the ICP27 protein may play a critical role by efficiently recognizing the poly(A) signals at the 3′ co-terminals leading to a decreased convergent read-through from the E genes. Furthermore, poly(A) signals at the 3′ co-terminal of the late genes may not be recognized by the RNAP molecule, causing a transcriptional read-through from the L genes toward the E genes. In other words, the poly(A) signals might be differentially recognized by ICP27 and/or other factors, thereby dictating a temporal order in the succession of the ON/OFF switching of gene expression. Other viral proteins, such as VHS and EP0, may also play roles in the kinetic shift, presumably through modulating the expression and/or activity of the IE180 transactivator (Tombácz et al., 2011, 2012). (5) Following the switch, the transcription of the L genes are initiated by the IE180 protein and other factors, which results in the inhibition of the E gene expression through the above-described read-through mechanism. In principle, the direction of the transcription can be changed more than once between convergent nested gene clusters in a hypothetical configuration of E1L1–L2E2 in the following temporal order: (step 1) E1 gene inhibits both the tandem L1 gene and the convergent L2 and E2 genes; (step 2) E2 inhibits all of the other three genes; (step 3) L1 blocks L2 and E2 genes; (step 4) L2 inhibits E1 and L1 (E1 and E2 genes are silent at steps 3 and 4 due to the absence of transcriptional initiation). The cascade-like events of viral gene expression is explained in terms of the mechanisms based on the combination of parallel and antiparallel collisions of RNAP molecules at the overlapping regions, hence, the name "Extended Seesaw model." Adjacent GMs may influence one another's transcription via transcriptional read-through extending beyond the boundaries separating two GMs. The long RNA molecules synthesized by the above processes may be difficult to detect by "size-sensitive" techniques, such as Northern blot analysis, because they may be randomly terminated and thus, produce smears on the gel.

The Promoter Competition model

Bidirectional promoters can also function in the regulation of the temporal orderliness of gene expression. Two of the divergent gene pairs of PRV (*ul20–ul21* and *ul29–ul30* genes) overlap at their TATA boxes. In our earlier publication, we categorized the *ul20* gene as E/L, and the other three genes as E genes. It is assumed that these genes are expressed in a slightly different kinetics, which is, at least partly, regulated by competition between overlapping promoters for the transcriptional initiation factors. The mechanisms influencing the preferential binding of transcription factors to the two regions of these promoters are currently unknown.

TINs IN CELLULAR ORGANISMS

The operating mechanism of TINs has been demonstrated through use of a herpesvirus model. The question may be raised of whether the same or similar principles can be applied to cellular organisms too. The structure and operation of genes of metazoan organisms are more complex than those of viral genes, due to the presence of introns, and the extended leader and trailer sequences, the need for a multitude of *cis*-regulatory sequences and transcription factors for the proper expressions, the essential roles of epigenetic modifications, and the expansive intergenic regions. Although viruses are intracellular parasites which exploit the biochemical and genetic apparatuses of the cell, it is theoretically possible that they utilize essentially different TIN-based mechanisms for the control of their gene expressions. The differences and similarities between these biological systems are explored below.

Parallel read-through by the tandem genes

The tandem organization of genes is common throughout the genome in all species (Pan and Zhang, 2008). For example, the vertebrate *Hox* genes are all positioned in a tandem orientation relative to one other (**Figure 2**). An extensive transcriptional read-through between the *Hox* genes has been reported in various species (Simeone et al., 1988; Shiga et al., 2006), indicating the importance of this mechanism in diverse evolutionary lineages. Furthermore, it has been demonstrated that, besides the "regular" transcriptional start site (TSS), many yeast genes contain an additional upstream TSS, which controls the generation of short CUT molecules transcribed in the same direction as those of mRNAs, which overlap the promoter and 5′-UTRs of the mRNAs (Neil et al., 2009). It has been shown that the mutation of the TATAA box of the yeast *TPI1* gene affects the transcription of the mRNA, but not that of the CUT, which indicates the existence

of two independent pre-initiation complexes on the promoter (Scott and Baker, 1993). Furthermore, the same study revealed a reverse correlation between the expressions of CUTs and the mRNAs of downstream glycolytic genes, which may reflect the operation of a mechanism similar to that described in the herpesviruses in the Waterfall model. The difference between these two systems is that in yeasts the upstream transcripts are not read-through products of upstream protein-coding genes, but are from RNA genes producing CUTs. Additionally, global run-on sequencing (Gro-seq) analysis has revealed that human RNAP extends more than 10 kb pairs on average beyond the poly(A) site of genes before termination (Core et al., 2008), and may interact with the transcription of closely positioned tandem genes, if they overlap.

Convergent gene overlaps

The estimated extent of the overlapping transcripts varies between the different organisms: 29% in mouse (34.9% in the protein-coding regions; Katayama et al., 2005), 5% in rat (Sun et al., 2006), 22% in human (Chen et al., 2004), 15% in fruit fly (Misra et al., 2002), and 9% in *Arabidopsis* (Wang et al., 2005). Sun et al. (2006) modified the data for mouse (12%) and fruit fly (17%). On the other hand, Cheng et al. (2005), using RACE/tiling arrays, found that 61% of all human transcribed regions have a counterpart on the opposite strand. These data demand further clarification, in view of the significant variations in the results obtained for the same species in different laboratories. Furthermore, it is difficult to accept that the mouse and rat differ to such a great extent as indicated above. The discrepancies in the data could be a consequence of the different techniques used by the different research groups. CUTs generated from a bidirectional promoter can overlap and thereby, interfere with a gene oriented in divergent fashion relative to a gene having a bidirectional promoter.

Convergent and divergent overlaps between protein-coding genes.

The *ul30 and ul31* genes are the only true convergently overlapping protein-coding gene pair of PRV. Convergent overlaps between oppositely oriented genes, as in the yeast RHO1/MRP2 (Peterson and Myers, 1993) and GAL7/GAL10 gene pairs (Prescott and Proudfoot, 2002), appear to be rare in cellular organisms too. The reason for this could be that this type of interaction imposes a severe constraint on the expression of genes. Similarly,
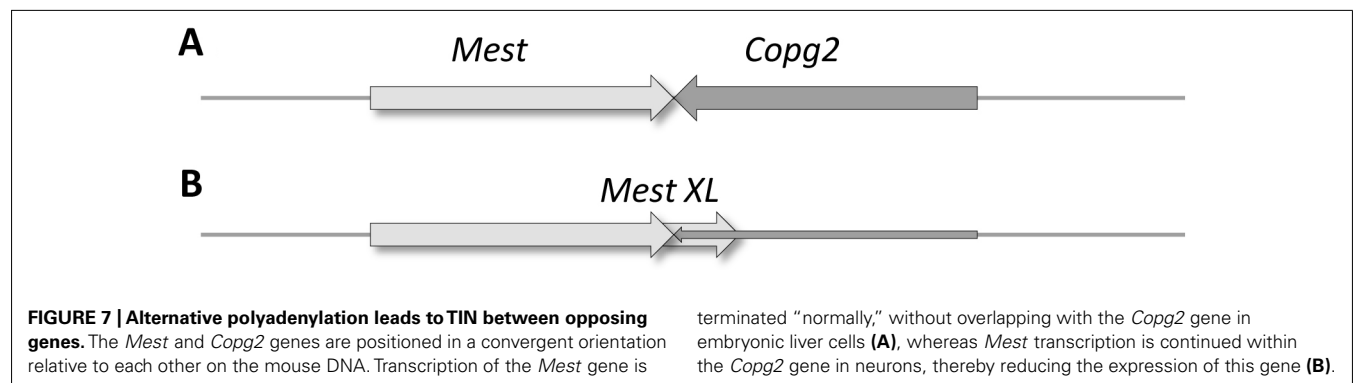
divergent gene overlaps are probably also rare, since gene expression is regulated by a multitude of other mechanisms in complex genomes.

Convergent read-trough overlaps between protein-coding genes.

The idea of convergent transcriptional read-trough assumes the existence of a main transcriptional termination/poly(A) signal, which is occasionally overlooked by the RNAP enzyme. Eukaryotic genes do not necessarily form convergent read-through overlaps to achieve a similar regulatory effect. Instead, they can also overlap as a result of alternative polyadenylation, which is recognized as a major post-transcriptional regulatory mechanism in eukaryotes. Differential polyadenylation depends on a variety of factors, including cell type (such as the *Mest* and *Cop* genes of the mouse; MacIsaac et al., 2011; **Figure 7**), embryonic stage, environment and disease (Audic and Claverie, 1997). The 3′-UTRs are largely unannotated, and the extent of alternative polyadenylation in the various genomes is therefore still unknown. Through use of a novel genomic 3′RACE technique, Mangone et al. (2010) detected the presence of at least one 3′-UTR isoform for every protein-coding gene of *C. elegans*, which were differentially expressed during development. The extents of the various overlaps produced by alternative polyadenylation are currently unknown.

Convergent overlaps between a protein-coding gene and an RNA gene.

The *Hox* gene cluster is one of the best-known groups of genes, but it is still not available a comprehensive knowledge relating to the transcription from these DNA regions. It was recently estimated that the unannotated transcripts can account for up to 60% of the total transcriptional output of a cluster (Mainguy et al., 2007). The generation of antisense transcripts has been reported in *Drosophila Hox* genes, which act in both *trans* (Rinn et al., 2007) and in *cis;* this latter represents 14 antisense transcripts overlapping the *Hox* genes (Mainguy et al., 2007). Antisense CUTs generated from a bidirectional promoter can, in principle, convergently overlap with the 3′-UTR of an upstream tandem gene over a certain length. This potential mechanism could provide a feedback from the downstream gene toward the upstream gene, which would provide a seesaw-like mechanism. An overall antisense transcription has been described in the mammalian genome. A perturbation of antisense transcription has been shown to affect the expression of mRNAs, which suggests that the

FIGURE 7 | Alternative polyadenylation leads to TIN between opposing genes. The *Mest* and *Copg2* genes are positioned in a convergent orientation relative to each other on the mouse DNA. Transcription of the *Mest* gene is terminated "normally," without overlapping with the *Copg2* gene in embryonic liver cells **(A)**, whereas *Mest* transcription is continued within the *Copg2* gene in neurons, thereby reducing the expression of this gene **(B)**.

antisense transcripts and/or the process generating them are functional (Katayama et al., 2005).

Convergent overlaps between RNA genes. A classical example of the interaction between two lnc-RNAs involves the *Xist* and *Tsix* RNA genes. Xist is the major factor of X chromosome inactivation, while Tsix is a *cis*-acting repressor of Xist. A close correlation has been established between Xist downregulation and Tsix upregulation during X inactivation (Rougeulle and Avner, 2004). A genome-wide occurrence of overlaps between the RNA genes has been reported in the mammalian genome (Katayama et al., 2005), which provides a great number of possibilities for the operation of TINs. It is reasonable to assume that the RNA genes, and especially those encoding lnc-RNAs, are regulated by the interaction of the transcriptional machineries according to the Waterfall, Seesaw, or Promoter Competition models.

Bidirectional promoters. A large proportion of eukaryotic promoters are intrinsically bidirectional (Wei et al., 2011), though the explanation of this is not yet known. When a bidirectional promoter controls the transcription of two divergent protein-coding genes, the purpose could be to provide either a synchronized expression or, in contrast, an inversely synchronized expression, as predicted by the Promoter Competition model. In most cases, however, bidirectional promoters generate the short, unstable transcripts called CUTs in the direction opposite to the genes. If a CUT does not overlap with a gene, it may result as a by-product of promoter competition (see above for other possible mechanisms). If a CUT overlaps with an upstream gene, then, depending on the relative orientation of this gene relative to the downstream gene, the interaction between the transcriptional machineries of the upstream gene and the CUT can be unidirectional (tandem orientation) or bidirectional (convergent orientation), as explained by the Waterfall model or the Seesaw model, respectively. Indeed, serial analysis of gene expression (SAGE) has shown that the 3′ ends of CUTs can be located in close proximity to the ORFs in both sense and antisense configurations (Neil et al., 2009).

A major difference between viruses and cellular organisms is that the interacting partner of a protein-coding gene in a cellular organism, and especially in a eukaryote, appears to be more frequently an RNA gene, rather than another protein-coding gene as in a virus. The reason for this could be that viruses contain extremely compact and efficient genomes, since they have to utilize every detail of the DNA. In PRV, only the two most important transactivator genes (*ie180* and *ep0*) are regulated by antisense transcription from RNA genes (LAT and ASP). Additionally, at the moment, it is not known whether CUTs are also expressed in viruses.

THE FUNCTION OF TINs
The TIN hypothesis claims that TIN-based mechanisms are not only curiosities with peripheral significance, but control the operation of a wide range of gene clusters throughout the genome. The temporal control of gene expression is assumed to be especially important in a variety of mechanisms, including virus life cycle,

embryogenesis, cell differentiation, tissue regeneration, cell signaling pathways, cellular stress, operation of gene networks, and metabolic pathways, among others. TINs can also participate in simple regulatory mechanisms requiring mutually exclusive gene expression of neighboring genes. Besides the temporal regulation, TINs are assumed to be involved in the spatial control of transcription in multicellular organism, which includes the regulation of tissue-specific gene expression.

THE POTENTIAL FUNCTIONS OF READ-THROUGH TRANSCRIPTS
Antisense transcripts and polycistronic RNAs are the by-products of the operation of TINs. However, this does not inevitably mean that they are not utilized at other levels of genetic regulation or in other processes. Antisense RNAs can serve as miRNA precursors (Umbach et al., 2008), as well as translational regulators, by binding the complementary mRNA strand, thereby forming double-stranded RNAs. Indeed, the presence of double-stranded transcripts has been detected in cells infected with HSV (Jacquemont and Roizman, 1975), a mouse polyoma virus (overlapping poly(A) signals; Osato et al., 2007) and in *Drosophila* (Lai, 2002; Lewis et al., 2005). One might speculate about the function of long double-stranded viral transcripts in herpesvirus-infected cells. The mRNA/antisense RNA duplexes normally evoke the interferon-induced apoptotic pathway in mammalian cells. However, herpesviruses have developed several anti-apoptotic mechanisms as a defense strategy (Li and Ramchandran, 2010). Even though double-stranded transcripts are tolerated by infected cells, the viral life cycle is too rapid for the completion of RNA interference. Alternatively, antisense transcripts may act as epigenetic regulators in processes such as the methylation of promoters and the conversion of the chromosome structure (Wutz et al., 1997; Reik and Walter, 2001; Tufarelli et al., 2003). Additionally, downstream genes of polycistronic RNAs can be translated through any of the abovementioned mechanisms.

CONCLUSION
Protein and RNA molecules are needed in optimal amounts in order to fulfill their biological functions; therefore, the regulation of their availability on a spatiotemporal scale is a critical point for the normal operation of cells and the organisms themselves. The TIN hypothesis suggests the genome-wide existence of GMs coordinating the expression of its genes through the interaction between their transcriptional machineries. The interplay among genes within and between GMs forms a TI network, which represents a novel layer of gene regulatory control in various forms of life. The TIN hypothesis suggests that the role of a TIN includes controlling the ON/OFF switching of gene expression in a temporally successive manner and maintaining the ON or OFF state for a given period of time within a cluster of functionally related genes through disrupting each other's transcription at the initiation and/or elongation phase. The transcription machineries clash with each other at overlapping regions in the collision-based (Waterfall and Seesaw) models and compete with each other in the promoter-competition-based model. In the collision-based models, the genes with higher initial expression rates and/or transcriptional read-through efficiencies suppress the transcription

of genes with lower initial expression and/or efficiency, thereby further reducing the transcription of the latter genes close to the zero level. The Waterfall model predicts a one-way inhibitory effect (from upstream to downstream genes) among tandem genes, while the Seesaw model envisages a reciprocal downregulatory effect on the transcription of convergently oriented genes. The latter system results in an inverse synchronization of transcription between convergent genes through a self-adjusting automatism, thereby greatly simplifying the coordination of gene expression. The Extended Seesaw model proposes a complex regulatory pattern between genes forming both tandem and convergent clusters in a genomic locus. In the latter case, the direction of transcription can be repeatedly altered. In this study, we focused on the operation of TINs in alpha-herpesviruses. The question is whether the large-scale organization of the genomes of cellular organisms resembles that of the herpesviruses. The number of true convergent overlaps between protein-coding genes is low in the genomes of various cellular organisms, which appears to imply a peripheral significance of transcriptional collision in controlling genome-wide transcription. However, convergent orientation of gene clusters is common in these organisms, which potentially allow transcriptional read-through and alternative transcriptional termination mechanisms. Moreover, the intrinsic bidirectionality of the eukaryotic promoters provides an additional means for neighboring genes to exert a mutual effect on each other's expression.

REFERENCES

Audic, S., and Claverie, J. M. (1997). The significance of digital gene expression profiles. *Genome Res.* 7, 986–995.

Al-Shahrour, F., Minguez, P., Marqués-Bonet, T., Gazave, E., Navarro, A., and Dopazo, J. (2010). Selection upon genome architecture: conservation of functional neighborhoods with changing genes. *PLoS Comput. Biol.* 6, e1000953. doi: 10.1371/journal.pcbi.1000953

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233.

Bertone, P., Stolc, V., Royce, T. E., Rozowsky, J. S., Urban, A. E., Zhu, X., Rinn, J. L., Tongprasit, W., Samanta, M., Weissman, S., Gerstein, M., and Snyder, M. (2004). Global identification of human transcribed sequences with genome tiling arrays. *Science* 306, 2242–2246.

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., Andrews, R. M., Flicek, P., Boyle, P. J., Cao, H., Carter, N. P., Clelland, G. K., Davis, S., Day, N., Dhami, P., Dillon, S. C., Dorschner, M. O., Fiegler, H., Giresi, P. G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K. D., Johnson, B. E., Johnson, E. M., Frum, T. T., Rosenzweig, E. R., Karnani, N., Lee, K., Lefebvre, G. C., Navas, P. A., Neri, F., Parker, S. C., Sabo, P. J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F. S., Dekker, J., Lieb, J. D., Tullius, T. D., Crawford, G. E., Sunyaev, S., Noble, W. S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I. L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H. A., Sekinger, E. A., Lagarde, J., Abril, J. F., Shahab, A., Flamm, C., Fried, C., Hackermüller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J. S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M. C., Thomas, D. J., Weirauch, M. T., Gilbert, J., Drenkow, J., Bell, I., Zhao, X., Srinivasan, K. G., Sung, W. K., Ooi, H. S., Chiu, K. P., Foissac, S., Alioto, T., Brent, M., Pachter, L., Tress, M. L., Valencia, A., Choo, S. W., Choo, C. Y., Ucla, C., Manzano, C., Wyss, C., Cheung, E., Clark, T. G., Brown, J. B., Ganesh, M., Patel, S., Tammana, H., Chrast, J., Henrichsen, C. N., Kai, C., Kawai, J., Nagalakshmi, U., Wu. J., Lian, Z., Lian, J., Newburger, P., Zhang, X., Bickel, P., Mattick, J. S., Carninci, P., Hayashizaki, Y., Weissman, S., Hubbard, T., Myers, R. M., Rogers, J., Stadler, P. F., Lowe, T. M., Wei, C. L., Ruan, Y., Struhl, K., Gerstein, M., Antonarakis, S. E., Fu, Y., Green, E. D., Karaöz, U., Siepel, A., Taylor, J., Liefer, L. A., Wetterstrand, K. A., Good, P. J., Feingold, E. A., Guyer, M. S., Cooper, G. M., Asimenos, G., Dewey, C. N., Hou, M., Nikolaev, S., Montoya-Burgos, J. I., Löytynoja, A., Whelan, S., Pardi, F., Massingham, T., Huang, H., Zhang, N. R., Holmes, I., Mullikin, J. C., Ureta-Vidal, A., Paten, B., Seringhaus, M., Church, D., Rosenbloom, K., Kent, W. J., Stone, E. A.; NISC Comparative Sequencing Program; Baylor College of Medicine Human Genome Sequencing Center; Washington University Genome Sequencing Center; Broad Institute; Children's Hospital Oakland Research Institute,; Batzoglou, S., Goldman, N., Hardison, R. C., Haussler, D., Miller, W., Sidow, A., Trinklein, N. D., Zhang, Z. D., Barrera, L., Stuart, R., King, D. C., Ameur, A., Enroth, S., Bieda, M. C., Kim, J., Bhinge, A. A., Jiang, N., Liu, J., Yao, F., Vega, V. B., Lee, C. W., Ng, P., Shahab, A., Yang, A., Moqtaderi, Z., Zhu, Z., Xu, X., Squazzo, S., Oberley, M. J., Inman, D., Singer, M. A., Richmond, T. A., Munn, K. J., Rada-Iglesias, A., Wallerman, O., Komorowski, J., Fowler, J. C., Couttet, P., Bruce, A. W., Dovey, O. M., Ellis, P. D., Langford, C. F., Nix, D. A., Euskirchen, G., Hartman, S., Urban, A. E., Kraus, P., Van Calcar, S., Heintzman, N., Kim, T. H., Wang, K., Qu, C., Hon, G., Luna, R., Glass, C. K., Rosenfeld, M. G., Aldred, S. F., Cooper, S. J., Halees, A., Lin, J. M., Shulha, H. P., Zhang, X., Xu, M., Haidar, J. N., Yu, Y., Ruan, Y., Iyer, V. R., Green, R. D., Wadelius, C., Farnham, P. J., Ren, B., Harte, R. A., Hinrichs, A. S., Trumbower, H., Clawson, H., Hillman-Jackson, J., Zweig, A. S., Smith, K., Thakkapallayil, A., Barber, G., Kuhn, R. M., Karolchik, D., Armengol, L., Bird, C. P., de Bakker, P. I., Kern, A. D., Lopez-Bigas, N., Martin, J. D., Stranger, B. E., Woodroffe, A., Davydov, E., Dimas, A., Eyras, E., Hallgrímsdóttir, I. B., Huppert, J., Zody, M. C., Abecasis, G. R., Estivill, X., Bouffard, G. G., Guan, X., Hansen, N. F., Idol, J. R., Maduro, V. V., Maskeri, B., McDowell, J. C., Park, M., Thomas, P. J., Young, A. C., Blakesley, R. W., Muzny, D. M., Sodergren, E., Wheeler, D. A., Worley, K. C., Jiang, H., Weinstock, G. M., Gibbs, R. A., Graves, T., Fulton, R., Mardis, E. R., Wilson, R. K., Clamp, M., Cuff, J., Gnerre, S., Jaffe, D. B., Chang, J. L., Lindblad-Toh, K., Lander, E. S., Koriabine, M., Nefedov, M., Osoegawa, K., Yoshinaga, Y., Zhu, B., and de Jong, P. J. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447, 799–816.

Boldogköi, Z., Bálint, K., Awatramani, G. B., Balya, D., Busskamp, V., Viney, T. J.,Lagali, P. S., Duebel, J., Pásti, E., Tombácz, D., Tóth, J. S., Takács, I. F., Scherf, B. G., and Roska, B. (2009). Genetically timed, activity-sensor and rainbow transsynaptic viral tools. *Nat. Methods* 6, 127–130.

Boldogköi, Z., Erdélyi, F., and Fodor, I. (2000). A putative latency promoter/enhancer [P(LAT2)] region of pseudorabies virus contains a virulence determinant. *J. Gen. Virol.* 81(Pt 2), 415–420.

Boldogköi, Z., Murvai, J., and Fodor, I. (1995). G and C accumulation at silent positions of codons produces additional ORFs. *Trends Genet.* 11, 125–126.

Boldogköi, Z., and Nógrádi, A. (2003). Gene and cancer therapy – pseudorabies virus: a novel research and therapeutic tool? *Curr. Gene Ther.* 3, 155–182.

Boldogköi, Z., Sík, A., Dénes, A., Reichart, A., Toldi, J., Gerendai, I., Kovács, K. J., and Palkovits, M. (2004). Novel tracing paradigms–genetically engineered herpesviruses as tools for mapping functional circuits within the CNS: present status and future prospects. *Prog. Neurobiol.* 72, 417–445.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., Kodzius, R., Shimokawa, K., Bajic, V. B., Brenner, S. E., Batalov, S., Forrest, A. R., Zavolan, M., Davis, M. J., Wilming, L. G., Aidinis, V., Allen, J. E., Ambesi-Impiombato, A., Apweiler, R., Aturaliya, R. N., Bailey, T. L., Bansal, M., Baxter, L.,

Beisel, K. W., Bersano, T., Bono, H., Chalk, A. M., Chiu, K. P., Choudhary, V., Christoffels, A., Clutterbuck, D. R., Crowe, M. L., Dalla, E., Dalrymple, B. P., de Bono, B., Della Gatta, G., di Bernardo, D., Down, T., Engstrom, P., Fagiolini, M., Faulkner, G., Fletcher, C. F., Fukushima, T., Furuno, M., Futaki, S., Gariboldi, M., Georgii-Hemming, P., Gingeras, T. R., Gojobori, T., Green, R. E., Gustincich, S., Harbers, M., Hayashi, Y., Hensch, T. K., Hirokawa, N., Hill, D., Huminiecki, L., Iacono, M., Ikeo, K., Iwama, A., Ishikawa, T., Jakt, M., Kanapin, A., Katoh, M., Kawasawa, Y., Kelso, J., Kitamura, H., Kitano, H., Kollias, G., Krishnan, S. P., Kruger, A., Kummerfeld, S. K., Kurochkin, I. V., Lareau, L. F., Lazarevic, D., Lipovich, L., Liu, J., Liuni, S., McWilliam, S., Madan Babu, M., Madera, M., Marchionni, L., Matsuda, H., Matsuzawa, S., Miki, H., Mignone, F., Miyake, S., Morris, K., Mottagui-Tabar, S., Mulder, N., Nakano, N., Nakauchi, H., Ng, P., Nilsson, R., Nishiguchi, S., Nishikawa, S., Nori, F., Ohara, O., Okazaki, Y., Orlando, V., Pang, K. C., Pavan, W. J., Pavesi, G., Pesole, G., Petrovsky, N., Piazza, S., Reed, J., Reid, J. F., Ring, B. Z., Ringwald, M., Rost, B., Ruan, Y., Salzberg, S. L., Sandelin, A., Schneider, C., Schönbach, C., Sekiguchi, K., Semple, C. A., Seno, S., Sessa, L., Sheng, Y., Shibata, Y., Shimada, H., Shimada, K., Silva, D., Sinclair, B., Sperling, S., Stupka, E., Sugiura, K., Sultana, R., Takenaka, Y., Taki, K., Tammoja, K., Tan, S. L., Tang, S., Taylor, M. S., Tegner, J., Teichmann, S. A., Ueda, H. R., van Nimwegen, E., Verardo, R., Wei, C. L., Yagi, K., Yamanishi, H., Zabarovsky, E., Zhu, S., Zimmer, A., Hide, W., Bult, C., Grimmond, S. M., Teasdale, R. D., Liu, E. T., Brusic, V., Quackenbush, J., Wahlestedt, C., Mattick, J. S., Hume, D. A., Kai, C., Sasaki, D., Tomaru, Y., Fukuda, S., Kanamori-Katayama, M., Suzuki, M., Aoki, J., Arakawa, T., Iida, J., Imamura, K., Itoh, M., Kato, T., Kawaji, H., Kawagashira, N., Kawashima, T., Kojima, M., Kondo, S., Konno, H., Nakano, K., Ninomiya, N., Nishio, T., Okada, M., Plessy, C., Shibata, K., Shiraki, T., Suzuki, S., Tagami, M., Waki, K., Watahiki, A., Okamura-Oho, Y., Suzuki, H., Kawai, J., Hayashizaki, Y.; FANTOM Consortium; RIKEN Genome Exploration Research Group; and Genome Science Group (Genome Network Project Core Group). (2005). The transcriptional landscape of the mammalian genome. *Science* 309, 1559–1563.

Carthew, R. W. (2006). Gene regulation by microRNAs. *Curr. Opin. Genet. Dev.* 16, 203–208.

Chang, Y. E., Menotti, L., Filatov, F., Campadelli-Fiume, G., and Roizman, B. (1998). UL27.5 is a novel gamma2 gene antisense to the herpes simplex virus 1 gene encoding glycoprotein B. *J. Virol.* 72, 6056–6064.

Chen, J., Sun, M., Kent, W. J., Huang, X., Xie, H., Wang, W., Zhou, G., Shi, R. Z., and Rowley, J. D. (2004). Over 20% of human transcripts might form sense–antisense pairs. *Nucleic Acids Res.* 32, 4812–4820.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., Sementchenko, V., Piccolboni, A., Bekiranov, S., Bailey, D. K., Ganesh, M., Ghosh, S., Bell, I., Gerhard, D. S., and Gingeras, T. R. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* 308, 1149–1154.

Cheung, A. K. (1989). Detection of pseudorabies virus transcripts in trigeminal ganglia of latently infected swine. *J. Virol.* 63, 2908–2913.

Core, L. J., Waterfall, J. J., and Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845–1848.

DeLong, J. P., Okie, J. G., Moses, M. E., Sibly, R. M., and Brown, J. H. (2010). Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. *Proc. Natl. Acad. Sci. U.S.A.* 107, 12941–12945.

Ebisuya, M., Yamamoto, T., Nakajima, M., and Nishida, E. (2008). Ripples from neighbouring transcription. *Nat. Cell. Biol.* 10, 1106–1113.

Farrell, M. J., Dobson, A. T., and Feldman, L. T. (1991). Herpes simplex virus latency-associated transcript is a stable intron. *Proc. Natl. Acad. Sci. U.S.A.* 88, 790–794.

Friedman, R. C., Farh, K. K., Burge, C. B., and Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105.

Grey, F., Antoniewicz, A., Allen, E., Saugstad, J., McShea, A., Carrington, J. C., and Nelson, J. (2005). Identification and characterization of human cytomegalovirus-encoded microRNAs. *J. Virol.* 79, 12095–12099.

Gu, R., Zhang, Z., DeCerbo, J. N., and Carmichael, G. G. (2009). Gene regulation by antisense overlap of polyadenylation signals. *RNA* 15, 1154–1163.

Han, J., Kim, D., and Morris, K. V. (2007). Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells. *Proc. Natl. Acad. Sci. U.S.A.* 104, 12422–12427.

Hermsen, R., ten Wolde, P. R., and Teichmann, S. (2008). Chance and necessity in chromosomal gene distributions. *Trends Genet.* 24, 216–219.

Hurst, L. D., Pál, C., and Lercher, M. J. (2004). The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* 5, 299–310.

Huvet, M., Nicolay, S., Touchon, M., Audit, B., d'Aubenton Carafa, Y., Arneodo, A., and Thermes, C. (2007). Human gene organization driven by the coordination of replication and transcription. *Genome Res.* 17, 1278–1285.

Izant, J. G., and Weintraub, H. (1985). Constitutive and conditional suppression of exogenous and endogenous genes by anti-sense RNA. *Science* 229. 345–352.

Jacquemont, B., and Roizman, B. (1975). RNA synthesis in cells infected with herpes simplex virus. X. Properties of viral symmetric transcripts and of double-stranded RNA prepared from them. *J. Virol.* 15, 707–713.

Janowski, B. A., Kaihatsu, K., Huffman, K. E., Schwartz, J. C., Ram, R., Hardy, D., Mendelson, C. R., and Corey, D. R. (2005). Inhibiting transcription of chromosomal DNA with antigene peptide nucleic acids. *Nat. Chem. Biol.* 1, 210–215.

Jovasevic, V., and Roizman, B. (2010). The novel HSV-1 US5-1 RNA is transcribed off a domain encoding US5, US4, US3, US2 and alpha22. *Virol. J.* 7, 103.

Kaliman, A. V., Boldogköi, Z., and Fodor, I. (1994). Large and small subunits of the Aujeszky's disease virus ribonucleotide reductase: nucleotide sequence and putative structure. *Biochim. Biophys. Acta* 1219, 151–156.

Kapranov, P., Cheng, J., Dike, S., Nix, D. A., Duttagupta, R., Willingham, A. T., Stadler, P. F., Hertel, J., Hackermüller, J., Hofacker, I. L., Bell, I., Cheung, E., Drenkow, J., Dumais, E., Patel, S., Helt, G., Ganesh, M., Ghosh, S., Piccolboni, A., Sementchenko, V., Tammana, H., and Gingeras, T. R. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316, 1484–1488.

Kapranov, P., and St. Laurent, G. (2012). Dark matter RNA: existence, function, controversy. *Front. Genet.* 3:60. doi: 10.3389/fgene.2012.00060

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C. C., Suzuki, M., Kawai, J., Suzuki, H., Carninci, P., Hayashizaki, Y., Wells, C., Frith, M., Ravasi, T., Pang, K. C., Hallinan, J., Mattick, J., Hume, D. A., Lipovich, L., Batalov, S., Engström, P. G., Mizuno, Y., Faghihi, M. A., Sandelin, A., Chalk, A. M., Mottagui-Tabar, S., Liang, Z., Lenhard, B., Wahlestedt, C.; RIKEN Genome Exploration Research Group; Genome Science Group (Genome Network Project Core Group); FANTOM Consortium. (2005). Antisense transcription in the mammalian transcriptome. *Science* 309, 1564–1566.

Kosman, D., Mizutani, C. M., Lemons, D., Cox, W. G., McGinnis, W., and Bier, E. (2004). Multiplex detection of RNA expression in *Drosophila* embryos. *Science* 305, 846.

Lai, E. C. (2002). Micro RNAs are complementary to 39 UTR sequence motifs that mediate negative post-transcriptional regulation. *Nat. Genet.* 30, 363–364.

Lapidot, M., and Pilpel, Y. (2006). Genome-wide natural antisense transcription: coupling its regulation to its different regulatory mechanisms. *EMBO Rep.* 7, 1216–1222.

Lassman, C. R., and Milcarek, C. (1992). Regulated expression of the mouse gamma 2b Ig H chain gene is influenced by poly(A) site order and strength. *J. Immunol.* 148, 2578–2585.

Lee, J. M., and Sonnhammer, E. L. (2003). Genomic gene clustering analysis of pathways in eukaryotes. *Genome Res.* 13, 875–882. Erratum in (2004): *Genome Res.* 14, 2510.

Lehner, B., Williams, G., Campbell, R. D., and Sanderson, C. M. (2002). Antisense transcripts in the human genome. *Trends Genet.* 18, 63–65.

Lercher, M. J., Urrutia, A. O., and Hurst, L. D. (2002). Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat Genet.* 31, 180–183.

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15–20.

Li, K., and Ramchandran, R. (2010). Natural antisense transcript: a concomitant engagement with protein-coding transcript. *Oncotarget* 1, 447–452.

Li, S., Carpenter, D., Hsiang, C., Wechsler, S. L., and Jones, C. (2010). Herpes simplex virus type 1 latency-associated transcript inhibits apoptosis and promotes neurite sprouting in neuroblastoma cells following serum starvation by maintaining protein kinase B (AKT) levels. *J. Gen. Virol.* 91(Pt 4), 858–866.

Li, W.-H. Gu, Z., Wang, H., and Nekrutenko, A. (2001). Evolutionary analyses of the human genome. *Nature* 409, 847–849.

Li, Y. Y., Qin, L., Guo, Z. M., Liu, L., Xu, H., Hao, P., Su, J., Shi, Y., He, W.-Z., and Li, Y.-X. (2006). *In silico* discovery of human natural antisense transcripts. *BMC Bioinformatics* 7, 18. doi: 10.1186/1471-2105-7-18

Louro, R., Smirnova, A. S., and Verjovski-Almeida, S. (2009). Long intronic noncoding RNA transcription: expression noise or expression choice? *Genomics* 93, 291–298.

Lynch, M. (2007a). The evolution of genetic networks by non-adaptive processes. *Nat. Rev. Genet.* 8, 803–813.

Lynch, M. (2007b). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl. Acad. Sci. U.S.A.* 104(Suppl. 1), 8597–8604.

Lynch, M., and Conery, J. S. (2003). The origins of genome complexity. *Science* 302, 1401–1404.

MacIsaac, J. L., Bogutz, A. B., Morrissy, A. S., and Lefebvre, L. (2011). Tissue-specific alternative polyadenylation at the imprinted gene Mest regulates allelic usage at Copg2. *Nucleic Acids Res.* 40, 1523–1535.

Mainguy, G., Koster, J., Woltering, J., Jansen, H., and Durston, A. (2007). Extensive polycistronism and antisense transcription in the mammalian Hox clusters. *PLoS ONE* 2, e356. doi: 10.1371/journal.pone.0000356

Mangone, M., Manoharan, A. P., Thierry-Mieg, D., Thierry-Mieg, J., Han, T., Mackowiak, S. D., Mis, E., Zegar, C., Gutwein, M. R., Khivansara, V., Attie, O., Chen, K., Salehi-Ashtiani, K., Vidal, M., Harkins, T. T., Bouffard, P., Suzuki, Y., Sugano, S., Kohara, Y., Rajewsky, N., Piano, F., Gunsalus, K. C., and Kim, J. K. (2010). The landscape of *C. elegans* 3′UTRs. *Science* 329, 432–435.

Martens, J. A., Laprade, L., and Winston, F. (2004). Intergenic transcription is required to repress *Saccharomyces cerevisiae* SER3 gene. *Nature* 429, 571–574.

Mavrich, T. N., Jiang, C., Ioshikhes, I. P., Li, X., Venters, B. J., Zanton, S. J., Tomsho, L. P., Qi, J., Glaser, R. L., Schuster, S. C., Gilmour, D. S., Albert, I., and Pugh, B. F. (2008). Nucleosome organization in the *Drosophila* genome. *Nature* 453, 358–362.

Mattick, J. (2010). Non-coding RNAs and eukaryotic evolution – a personal view. *BMC Biol.* 8, 67. doi: 10.1186/1741-7007-8-67

Mattick, J. S. (2004). RNA regulation: A new genetics? *Nat. Rev. Genet.* 5, 316–323.

Mattick, J. S., and Makunin, I. V. (2006). Non-coding RNA. *Hum. Mol. Genet.* 15, R17–29.

McGregor, F., Phelan, A., Dunlop, J., and Clements, J. B. (1996). Regulation of herpes simplex virus poly(A) site usage and the action of immediate-early protein IE63 in the early–late switch. *J. Virol.* 70, 1931–1940.

McLauchlan, J., Phelan, A., Loney, C., Sandri-Goldin, R. M., and Clements, J. B. (1992). Herpes simplex virus IE63 acts at the posttranscriptional level to stimulate viral mRNA 3′ processing. *J. Virol.* 66, 6939–6945.

Meier, J. L., and Straus, S. E. (1993). Varicella-zoster virus DNA polymerase and major DNA-binding protein genes have overlapping divergent promoters. *J. Virol.* 67, 7573–7581.

Michalak, P. (2008). Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* 91, 243–248.

Misra, S., Crosby, M. A., Mungall, C. J., Matthews, B. B., Campbell, K. S., Hradecky, P., Huang, Y., Kaminker, J. S., Millburn, G. H., Prochnik, S. E., Smith, C. D., Tupy, J. L., Whitfield, E. J., Bayraktaroglu, L., Berman, B. P., Bettencourt, B. R., Celniker, S. E., de Grey, A. D. N. J., Drysdale, R. A., Harris, N. L., Richter, J., Russo, S., Schroeder, A. J., Shu, S. Q., Stapleton, M., Yamada, C., Ashburner, M., Gelbart, W. M., Rubin, G. M., and Lewis, S. E. (2002). Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. *Genome Biol.* 3: RESEARCH0083.

Morris, K. V., and Vogt, P. K. (2010). Long antisense non-coding RNAs and their roles in transcription and oncogenesis. *Cell Cycle* 13, 2544–2547.

Napoli, S., Pastori, C., Magistri, C. M., Carbone, G. M., and Catapano, C. V. (2009). Promoter-specific transcriptional interference and c-myc gene silencing by siRNAs in human cells. *EMBO J.* 28, 1708–1719.

Necsulea, A., Guillet, C., Cadoret, J. C., Prioleau, M. N., and Duret, L. (2009). The relationship between DNA replication and human genome organization. *Mol. Biol. Evol.* 26, 729–741.

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L. M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457, 1038–1042.

Oh, S. K., Scott, M. P., and Sarnow, P. (1992). Homeotic gene antennapedia mRNA contains 5′ noncoding sequences that confer translational initiation by internal ribosome binding. *Genes Dev.* 6, 1643–1653.

Osato, N., Suzuki, Y., Ikeo, K., and Gojobori, T. (2007). Transcriptional interferences in cis natural antisense transcripts of humans and mice. *Genetics* 176, 1299–1306.

Pan, D., and Zhang, L. (2008). Tandemly arrayed genes in vertebrate genomes. *Comp. Funct. Genomics* 545269.

Peri, S., Navarro, J. D., Amanchy, R., Kristiansen, T. Z., Jonnalagadda, C. K., Surendranath, V., Niranjan, V., Muthusamy, B., Gandhi, T. K., Gronborg, M., Ibarrola, N., Deshpande, N., Shanker, K., Shivashankar, H. N., Rashmi, B. P., Ramya, M. A., Zhao, Z., Chandrika, K. N., Padma, N., Harsha, H. C., Yatish, A. J., Kavitha, M. P., Menezes, M., Choudhury, D. R., Suresh, S., Ghosh, N., Saravana, R., Chandran, S., Krishna, S., Joy, M., Anand, S. K., Madavan, V., Joseph, A., Wong, G. W., Schiemann, W. P., Constantinescu, S. N., Huang, L., Khosravi-Far, R., Steen, H., Tewari, M., Ghaffari, S., Blobe, G. C., Dang, C. V., Garcia, J. G., Pevsner, J., Jensen, O. N., Roepstorff, P., Deshpande, K. S., Chinnaiyan, A. M., Hamosh, A., Chakravarti, A., and Pandey, A. (2003). Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res.* 13, 2363–2371.

Perng, G. C., Slanina, S. M., Yukht, A., Ghiasi, H., Nesburn, A. B., and Wechsler, S. L. (2000). The latency-associated transcript gene enhances establishment of herpes simplex virus type 1 latency in rabbits. *J. Virol.* 74, 1885–1891.

Peterson, J. A., and Myers, A. M. (1993). Functional analysis of mRNA 3′ end formation signals in the convergent and overlapping transcription units of the S. cerevisiae genes RHO1 and MRP2. *Nucleic Acids Res.* 21, 5500–5508.

Petkov, P. M. Graber, J. H., Churchill, G. A., DiPetrillo, K., King, B. L., and Paigen, K. (2005). Evidence of a large-scale functional organization of mammalian chromosomes. *PLoS Genet.* 1, e33. doi: 10.1371/journal.pgen.0010033

Petruk, S., Sedkov, Y., Riley, K. M., Hodgson, J., Schweisguth, F., Hirose, S., Jaynes, J. B., Brock, H. W., and Mazo, A. (2006). Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. *Cell* 127, 1209–1221.

Pfeffer, S., Zavolan, M., Grässer, F. A., Chien, M., Russo, J. J., Ju, J., John, B., Enright, A. J., Marks, D., Sander, C., and Tuschl, T. (2004). Identification of virus-encoded microRNAs. *Science* 304, 734–736.

Pomeranz, L. E., Reynolds, A. E., and Hengartner, C. J. (2005). Molecular biology of pseudorabies virus: impact on neurovirology and veterinary medicine. *Microbiol. Mol. Biol. Rev.* 69, 462–500.

Prescott, E. M., and Proudfoot, N. J. (2002). Transcriptional collision between convergent genes in budding yeast. *Proc. Natl. Acad. Sci. U.S.A.* 99, 8796–8801.

Priola, S. A., and Stevens, J. G. (1991). The 5′ and 3′ limits of transcription in the pseudorabies virus latency associated transcription unit. *Virology* 182, 852–856.

Prod'hon, C., Machuca, I., Berthomme, H., Epstein, A., and Jacquemont, B. (1996). Characterization of regulatory functions of the HSV-1 immediate-early protein ICP22. *Virology* 226, 393–402.

Reik, W., and Walter, W. J. (2001). Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.* 2, 21–32.

Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., and Chang, H. Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323.

Roeszler, K. N., Itman, C., Sinclair, A. H., and Smith, C. A. (2012). The long non-coding RNA, MHM, plays a role in chicken embryonic development, including gonadogenesis. *Dev. Biol.* 366, 317–326.

Roizman, B. (1996). The function of herpes simplex virus genes: a primer for genetic engineering of novel vectors. *Proc. Natl. Acad. Sci. U.S.A.* 93, 11307–11312.

Rougeulle, C., and Avner, P. (2004). The role of antisense transcription in the regulation of X-inactivation. *Curr. Top. Dev. Biol.* 63, 61–89.

Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L., Nelson, C. R., Hariharan, I. K., Fortini, M.

E., Li, P. W., Apweiler, R., Fleischmann, W., Cherry, J. M., Henikoff, S., Skupski, M. P., Misra, S., Ashburner, M., Birney, E., Boguski, M. S., Brody, T., Brokstein, P., Celniker, S. E., Chervitz, S. A., Coates, D., Cravchik, A., Gabrielian, A., Galle, R. F., Gelbart, W. M., George, R. A., Goldstein, L. S., Gong, F., Guan, P., Harris, N. L., Hay, B. A., Hoskins, R. A., Li, J., Li, Z., Hynes, R. O., Jones, S. J., Kuehl, P. M., Lemaitre, B., Littleton, J. T., Morrison, D. K., Mungall, C., O'Farrell, P. H., Pickeral, O. K., Shue, C., Vosshall, L. B., Zhang, J., Zhao, Q., Zheng, X. H., and Lewis, S. (2000). Comparative genomics of the eukaryotes. *Science* 287, 2204–2215.

Shackelton, L. A., and Holmes, E. C. (2004). The evolution of large DNA viruses: combining genomic information of viruses and their hosts. *Trends Microbiol.* 12, 458–465.

Schwartz, S., Felber, B. K., Benko, B. M., Fenyo, E. M., and Pavlakis, G. N. (1990). Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. *J. Virol.* 64, 2519–2529.

Scott, E. W., and Baker, H. V. (1993). Concerted action of the transcriptional activators REB1, RAP1, and GCR1 in the high-level expression of the glycolytic gene TPI. *Mol. Cell. Biol.* 13, 543–550.

Seila, A. C., Calabrese, J. M., Levine, S. S., Yeo, G. W., Rahl, P. B., Flynn, R. A., Young, R. A., and Sharp, P. A. (2008). Divergent transcription from active promoters. *Science* 322, 1849–5181.

Shiga, Y., Sagawa, K., Takai, R., Sakaguchi, H., Yamagata, H., and Hayashi, S. (2006). Transcriptional readthrough of Hox genes Ubx and Antp and their divergent post-transcriptional control during crustacean evolution. *Evol. Dev.* 8, 407–414.

Simeone, A., Pannese, M., Acampora, D., D'Esposito, M., and Boncinelli, E. (1988). At least three human homeoboxes on chromosome 12 belong to the same transcription unit. *Nucleic Acids Res.* 16, 5379–5390.

Sokolowski, M., Scott, J. E., Heaney, R. P., Patel, A. H., and Clements, J. B. (2003). Identification of herpes simplex virus RNAs that interact specifically with regulatory protein ICP27 *in vivo*. *J. Biol. Chem.* 278, 33540–33549.

Stacey, S. N., Jordan, D., Williamson, A. J. K., Brown, M., Coote, J. H., and Arrand, J. R. (2000). Leaky scanning g is the predominant mechanism for translation of human paplillomavirus type 16 E7 oncoprotein from E6/E7 bicistronic mRNA. *J. Virol.* 74, 7284–7297.

Stolt, P., and Zillig, W. (1993). Antisense RNA mediates transcriptional processing in an archaebacterium, indicating a novel kind of RNase activity. *Mol. Microbiol.* 7, 875–882.

Stroop, W. G., Rock, D. L., and Fraser, N. W. (1984). Localization of herpes simplex virus in the trigeminal and olfactory systems of the mouse central nervous system during acute and latent infections by in situ hybridization. *Lab. Invest.* 51, 27–38.

Struhl, K. (2007). Transcriptional noise and the fidelity of initiation by RNA polymerase II. *Nat. Struct. Mol. Biol.* 14, 103–105.

Sun, M., Hurst, L. D., Carmichael, G. G., and Chen, J. (2006). Evidence for variation in abundance of antisense transcripts between multicellular animals but no relationship between antisense transcription and organismic complexity. *Genome Res.* 16, 922–933.

Szpara, M. L., Tafuri, Y. R., Parsons, L., Shamim, S. R., Verstrepen, K. J., Legendre, M., and Enquist, L. W. (2011). A wide extent of inter-strain diversity in virulent and vaccine strains of alphaherpesviruses. *PLoS Pathog.* 7, E1002282. doi: 10.1371/journal.ppat.1002282

Tan-Wong, S. M., Wijayatilake, H. D., and Proudfoot, N. J. (2009). Gene loops function to maintain transcriptional memory through interaction with the nuclear pore complex. *Genes Dev.* 23, 2610–2624.

Teichmann, S. A., and Veitia, R. A. (2004). Genes encoding subunits of stable complexes are clustered on the yeast chromosomes: an interpretation from a dosage balance perspective. *Genetics* 167, 2121–2125.

Thomas, J. H. (2006). Analysis of homologous gene clusters in *Caenorhabditis elegans* reveals striking regional cluster domains. *Genetics* 172, 127–143.

Tombácz, D., Tóth, J. S., and Boldogköi, Z. (2011). Deletion of the virion host shut: off gene of pseudorabies virus results in selective upregulation of the expression of early viral genes in the late stage of infection. *Genomics* 98, 15–25.

Tombácz, D., Tóth, J. S., and Boldogköi, Z. (2012). Effects of deletion of the early protein 0 gene of pseudorabies virus on the overall viral gene expression. *Gene* 493, 235–242.

Tombácz, D., Tóth, J. S., Petrovszki, P., and Boldogköi, Z. (2009). Whole-genome analysis of pseudorabies virus gene expression by real-time quantitative RT-PCR assay. *BMC Genomics* 10, 491.

Trinklein, N. D., Aldred, S. F., Hartman, S. J., Schroeder, D. I., Otillar, R. P., and Myers, R. M. (2004). An abundance of bidirectional promoters in the human genome. *Genome Res.* 14, 62–66.

Tsai, M. C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J. K., Lan, F., Shi, Y., Segal, E., and Chang, H. Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689–693.

Tufarelli, C., Stanley, J. A., Garrick, D., Sharpe, J. A., Ayyub, H., Wood, W. G., and Higgs, D. R. (2003). Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat. Genet.* 34, 157–165.

Umbach, J. L., Kramer, M. F., Jurak, I., Karnowski, H. W., Coen, D. M., and Cullen, B. R. (2008). MicroRNAs expressed by herpes simplex virus 1 during latent infection regulate viral mRNAs. *Nature* 454, 780–783.

Wang, X., Arai, S., Song, X., Reichart, D., Du, K., Pascual, G., Tempst, P., Rosenfeld, M. G., Glass, C. K., and Kurokawa, R. (2008). Induced ncRNAs allosterically modify RNA-binding proteins in *cis* to inhibit transcription. *Nature* 454, 126–130.

Wang, X. J., Gaasterland, T., and Chua, N. H. (2005). Genome-wide prediction and identification of cis-natural antisense transcripts in *Arabidopsis thaliana*. *Genome Biol.* 6, R30

Ward, P. L., Barker, D. E., and Roizman, B. (1996). A novel herpes simplex virus 1 gene, UL43.5, maps antisense to the UL43 gene and encodes a protein which colocalizes in nuclear structures with capsid proteins. *J. Virol.* 70, 2684–2690.

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

Weinberg, Z., Perreault, J., Meyer, M. M., and Breaker, R. R. (2009). Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature* 462, 656–659.

Wilusz, J. E., Sunwoo, H., and Spector, D. L. (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* 23, 1494–1504.

Wutz, A., Smrzka, O. W., Schweifer, N., Schellander, K., Wagner, E. F., and Barlow, D. P. (1997). Imprinted expression of the Igf2r gene depends on an intronic CpG island. *Nature* 389, 745–749.

Zhang, G., Raghavan, B., Kotur, M., Cheatham, J., Sedmak, D., Cook, C., Waldman, J., and Trgovcich, J. (2007). Antisense transcription in the human cytomegalovirus transcriptome. *J. Virol.* 81, 11267–11281.

Zwaagstra, J. C., Ghiasi, H., Slanina, S. M., Nesburn, A. B., Wheatley, S. C., Lillycrop, K., Wood, J., Latchman, D. S., Patel, K., and Wechsler, S. L. (1990). Activity of herpes simplex virus type 1 latency-associated transcript (LAT) promoter in neuron-derived cells: evidence for neuron specificity and for a large LAT transcript. *J. Virol.* 64, 5019–5028.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L. M. (2009). Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033–1037.

Yi, G., Sze, S. H., and Thon, M. R. (2007). Identifying clusters of functionally related genes in genomes. *Bioinformatics* 23, 1053–1060.