www.nature.com/mtna

Probabilistic Bidirectional Promoter Switches: Noncoding RNA Takes Control

Stephen K Anderson[1,2]

The discovery of probabilistic promoter switches in genes that code for class I major histocompatibility complex receptors in mouse and human provides a useful paradigm to explain programmed cell fate decisions. These switches have preset probabilities of transcribing in either the sense or antisense direction, and the characteristics of individual switches are programmed by the relative affinity of competing transcription factor–binding sites. The noncoding RNAs produced from these switches can either activate or suppress gene transcription, based on their location relative to the promoter responsible for gene expression in mature cells. The switches are active in a developmental phase that precedes gene expression by mature cells, thus temporally separating the stochastic events that determine gene activation from the protein expression phase. This allows the probabilistic generation of variegated gene expression in the absence of selection and ensures that mature cells have stable expression of the genes. Programmed probabilistic switches may control cell fate decisions in many developmental systems, and therefore, it is important to investigate noncoding RNAs expressed by progenitor cells to determine if they are expressed in a stochastic manner at the single cell level. This review provides a summary of current knowledge regarding murine and human switches, followed by speculation on the possible involvement of probabilistic switches in other systems of programmed differentiation.

Overview

The ability of a biological organism to sense and react to a complex environment requires the development of an intricate sensory system. The adaptive immune system uses DNA recombination to generate a large array of T-cell and B-cell receptors to sense potentially harmful entities in the environment.[1] In contrast, natural killer (NK) cells of the innate immune system use a smaller group of receptors to survey the organism for loss or alteration of the class I major histocompatibility complex (MHC) proteins that are used by cells to present antigens to the immune system.[2] Since the MHC molecules are highly polymorphic, it is therefore necessary to generate specialized NK cell subsets that are able to sense changes in the expression of a particular MHC molecule.[3] The receptors for class I MHC are expressed in a variegated manner on NK cells, such that the majority of NK cells (~80%) express from 1 to 3 receptors per cell out of a total repertoire of 4–15 receptors depending on the genotype of the individual.[4,5] The variegated, stochastic expression of these receptors is achieved through the use of probabilistic bidirectional promoter switches that generate either sense transcripts linked to gene activation or antisense transcripts that are associated with gene silencing.[6] Each stochastic switch possesses an intrinsic probability of transcribing in either the sense or antisense direction that is programmed by the relative binding affinity of competing transcription factor–binding sites. This results in each receptor type being expressed on a predetermined percentage of cells within the population.

Evolution can therefore select for genes with different frequencies of receptor expression, based on the inherent usefulness of each receptor for survival of the organism.
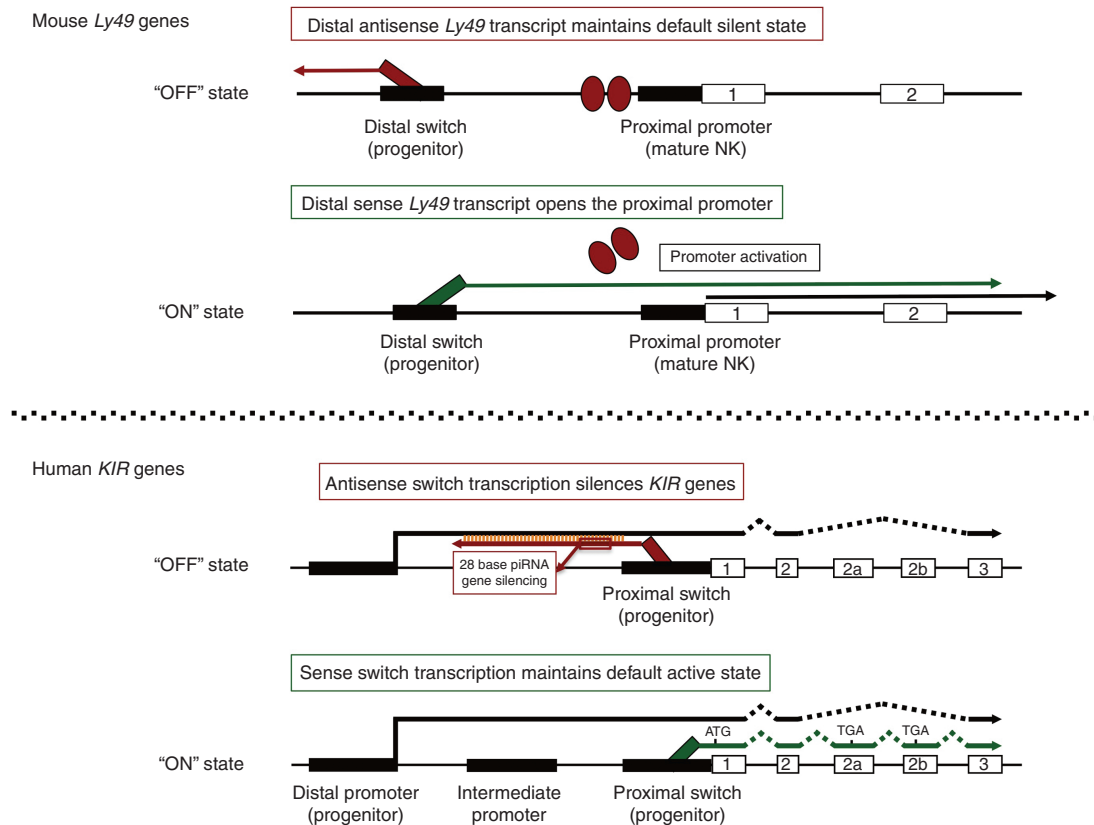
The human and mouse class I MHC receptor systems are separately evolved, and they use the noncoding RNAs generated by the stochastic promoter switches in different ways (**Figure 1**). The murine receptors are members of the C-type lectin–related *Ly49* gene family (originally named *Ly49a* through *Ly49x*, official name *Klra1-n*), and they contain a distal bidirectional promoter (Pro1) located ~5 kb upstream of the *Ly49* promoter used to express the gene in mature NK cells (Pro2).[7,8] In this system, sense noncoding transcripts activate the downstream promoter, whereas antisense noncoding transcripts merely represent the "off" state of the switch. In contrast, antisense transcripts play an active role in gene silencing in the human system. The human *KIR* genes are members of the immunoglobulin (Ig) supergene family (named according to the presence of 2 or 3 Ig domains and short or long cytoplasmic tails, *e.g.*, *KIR2DL1* or *KIR3DS1*), and they possess a stochastic switch 60 bp 5′ of the *KIR* initiation codon. The antisense transcript from the proximal promoter switch produces a piRNA that is associated with gene silencing; however, the sense transcript in this case merely maintains the default "on" state of the gene.

Although the same probabilistic promoter mechanism has been adopted by both the human and mouse receptor systems to achieve variegated receptor expression, the mouse *Ly49* genes use sense lncRNA to activate genes that are in a default "closed" state, whereas the human *KIR* use an

Figure 1 Differential role of lncRNAs produced by mouse versus human stochastic switches. The schematics show the location of stochastic switches and the lncRNAs produced in the mouse *Ly49* genes (upper panel) and the human *KIR* genes (lower panel). Promoters are displayed as black rectangles, and exons are shown as numbered boxes. Sense noncoding transcripts from the switch elements are shown as green lines, and antisense noncoding transcripts are shown in red. The red ovals at the proximal promoter of the *Ly49* genes depict the default closed chromatin state of these genes that is disrupted by sense transcripts from the distal promoter. The inclusion of alternative exons (2a and 2b) in forward transcripts from the *KIR* switch illustrates the observed nontranslatable nature of proximal sense transcripts.

antisense lncRNA to silence active genes. In both human and mouse, active loci are hypomethylated,[9,10] suggesting that DNA methylation plays a role in the maintenance of the chosen state; however, histone acetylation patterns are distinct. All genes of the *KIR* cluster have generally high levels of H3 and H4 acetylation regardless of their expression status,[11] whereas inactive *Ly49* genes have low levels that increase upon gene activation.[10] Consistent with these observations, the DNA methylation inhibitor 5-aza-cytidine alone can induce expression of silent *KIR* genes,[9,11] whereas *Ly49* gene activation requires both 5-aza-cytidine together with the histone deacetylase inhibitor trichostatin-A.[10]

The Ly49 Stochastic Switch

Analysis of *Ly49* transcripts in liver NK cells revealed the presence of a distal promoter that was active only in immature NK cells (Pro1).[7] *In vitro* analysis demonstrated that this element was in fact a bidirectional promoter containing two TATA boxes with overlapping C/EBP-binding sites located ~100 bp apart.[8] Mutational analysis revealed that the competing C/EBP and TATA elements determined the relative sense/antisense activity of the promoter. The switching activity of this element was demonstrated by placing it between two different fluorescent protein cDNAs (CFP and YFP) and

observing its behavior in real time. Remarkably, the element acted as a stable switch, choosing transcription in a single direction and maintaining that choice until a new copy of the element was generated by DNA replication. A cloned cell line containing a single copy of the two-color vector containing the *Ly49g* stochastic switch produced a variegated cell population composed of approximately equal numbers of CFP- or YFP-expressing cells, consistent with the nearly equivalent forward and reverse *Ly49g* switch promoter activity detected *in vitro*. Time lapse imaging of the variegated cell population revealed that a dividing cell could express both YFP and CFP simultaneously prior to division and generate separate CFP- or YFP-expressing daughter cells. This result provides direct evidence that the choice of sense or antisense transcription is not controlled by the relative concentration of transcription factors in the nucleus, since the two copies of the switch element in a dividing cell are able to choose different transcriptional fates. The switch is therefore capable of making a probabilistic choice between the generation of a forward transcript containing the Ly49-coding region or a reverse transcript that extends into the intergenic region and contains no identifiable coding region. It is also of interest to note that only the newly synthesized copy of the switch was capable of choosing to transcribe in the opposite direction, as sorted cells produced a 75–25% mix after one cell division

rather than the 50–50% mix predicted if both copies had the ability to transcribe in either direction. This suggests that the TATA-binding protein remains stably associated with the TATA box, even during the process of DNA replication. TATA-binding protein has been shown to remain bound to condensed chromosomes during cell division, marking active promoters that resume transcription after cell division.[12,13]

Analysis of the transcription properties of the switches contained within various *Ly49* genes revealed an association between the strength of the sense transcriptional activity and the frequency of NK cells expressing that gene.[8] The switches thus display a programmed probability of gene activation that can be ascribed to the relative strength of sense versus antisense promoter elements. Although the process of gene activation is stochastic, a bias toward gene activation or maintenance of the silent state is programmed by differences in switch characteristics, akin to "loaded dice." A model of probabilistic gene activation was proposed wherein the choice of sense transcription led to the production of a forward nontranslated transcript that traversed the downstream *Ly49* promoter, releasing it from its default silent state. The requirement of this transcript for gene activation was demonstrated by the loss of gene expression in either transgenes or endogenous *Ly49* genes that lacked the switch element.[14,15] The antisense switch transcript is a noncoding polyadenylated RNA that presumably represents the "off" state since deletion of the Pro1 element results in a silent gene, arguing against an active role of the antisense in gene silencing.[14]

Each *Ly49* gene represents a separate probabilistic unit with its own stochastic switch. The sense transcript of the switch element spans the entire coding region of the gene and produces a spliced, polyadenylation cDNA that is identical to the mature Ly49-coding mRNA except for the addition of a long 5′ noncoding region that prevents its efficient translation. The antisense switch transcript terminates before it reaches the neighboring *Ly49* gene. The boundary between adjacent *Ly49* genes is marked by a pair of transcriptional termination elements. A large T/C tract (>200 nucleotides) follows the polyA addition site of the *Ly49* mRNA, and ~350 bp downstream, a similar element is present on the other strand. This transcriptional insulator is likely present to prevent antisense lncRNA from the downstream gene from interfering with the expression of the preceding gene.

The Human *KIR* Genes Contain a Stochastic Switch Adjacent to the First Exon

The search for a similar switch controlling the variegated expression of the human *KIR* genes resulted in the surprising finding that the proximal promoter thought to be responsible for KIR expression was in fact bidirectional.[16] There is no significant homology between the *Ly49* and *KIR* bidirectional promoters, and the only shared transcription factor–binding sites are AML/RUNX and Ets. There is a conserved bidirectional core element in most *KIR* genes, and the strength of flanking 5′-YY1– and 3′-SP1–binding sites modulates sense and antisense transcriptional activities in individual *KIR* genes. The antisense transcript is a polyadenylated lncRNA of ~300 bases in length. Sense transcripts from the proximal

promoter are potentially translatable, leading to the production of KIR protein. However, substantial levels of proximal sense transcripts from most *KIR* genes can be detected in NK cells that are in the later stages of differentiation but do not express KIR protein, suggesting that the sense transcripts from the switch in these cells are also lncRNAs. This represents another parallel with the murine system: forward switch transcription does not result in protein expression. The lack of protein expression during the stochastic promoter phase allows for the generation of a variegated array of receptors that will not be subjected to selection until the NK cell matures. The percentage of NK cells expressing a given KIR correlates with the strength of sense relative to antisense transcription from the switch in several genes as well as alleles of a single gene (*KIR3DL1*).[17] The presence or absence of cognate MHC ligands does have some effect on the percentage of NK cells expressing a given receptor, indicating that postexpression selection does occur, but the primary determinant of expression frequency lies within the *KIR* gene itself.[18]

Similar to the *Ly49* genes, a distal transcript is present in immature NK cells; however, the distal *KIR* promoter is unidirectional, and only produces sense transcripts that are associated with locus activation.[19,20] The model for variegated KIR expression proposes that distal transcripts are required for the initial activation of the proximal switch element, similar to the *Ly49* system. However, once activated, the proximal switch element will either transcribe in the sense direction, maintaining the active state, or produce a reverse transcript that leads to resilencing of the proximal promoter region. A search for small RNAs generated from the *KIR* antisense transcripts resulted in the identification of a 28 base piRNA, and gene transduction studies demonstrated that the 28 base region of the antisense was required for *KIR* gene silencing, indicating an autoregulatory role of *KIR* antisense lncRNA.[21] An additional spliced *KIR* antisense noncoding RNA transcribed from a unidirectional antisense promoter in the second exon was discovered in the HEK293 cell line, and this lncRNA was subsequently found to be expressed *in vivo* only by cord blood progenitor cells.[22] The exon 2 antisense lncRNA overlaps with the switch antisense transcript, and it is also able to produce the 28 base piRNA and silence *KIR* loci. This suggests that the *KIR* genes are silenced early in development by an active mechanism to ensure that they are not expressed prematurely.

With the exception of the exon 2 antisense transcript found in progenitor cells, *KIR* antisense transcripts are only present in NK cells just prior to their differentiation into mature KIR-expressing NK cells, and this likely represents the probabilistic phase of *KIR* gene activation where the stochastic promoter switch is active.

The recent study of a rare, weakly expressed allele of the *KIR2DL1* gene has provided some insight into possible mechanisms preventing expression of KIR protein in cells with an active proximal promoter switch.

The weakly expressed variant of the *KIR2DL1* gene contained a distal promoter polymorphism that sharply decreased transcription from a previously unreported intermediate transcription start site located 270 bp upstream from the proximal switch promoter start site.[23] The intermediate transcripts
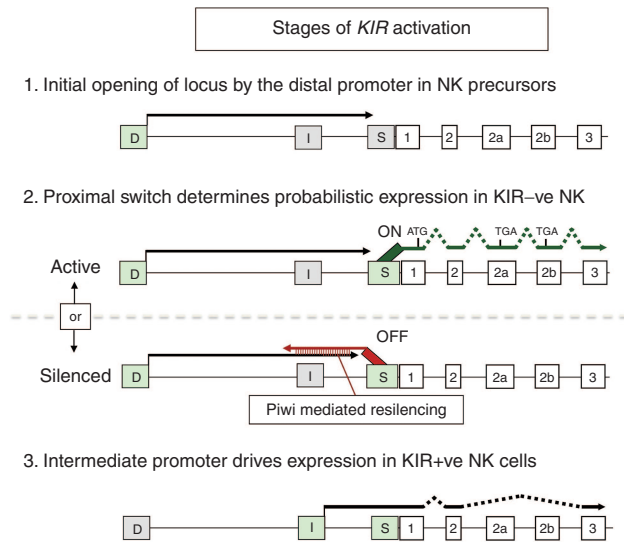
Figure 2 Proposed stages of *KIR* gene activation during NK cell development. A schematic of the 5′ region of the *KIR2DL1* gene is shown. Distal, intermediate, and switch promoters are indicated by rectangles labeled D, I, and S, respectively, and exons are shown as numbered rectangles. Active promoters in each stage are green, and inactive promoters are gray. Transcripts are shown by lines ending in arrows, and intronic regions are shown as dotted lines. The site of initiation of KIR protein translation is indicated by a vertical line labeled "ATG," and the location of premature stop codons in alternative exons found in forward transcripts from Pro-S are indicated by "TGA."

were shown to be translatable, and their levels paralleled KIR protein expression in a series of individuals with differing genotypes. *KIR2DL1* alleles with decreased intermediate transcripts produced proximal transcripts that contained alternative exons with stop codons (exon 2a or 2b; **Figures 1** and **2**). Alternative exon usage was not readily detectable in common *KIR2DL1* alleles with normal intermediate promoter activity, suggesting that intermediate transcription prevents the usage of the alternative exons. Individuals with low levels of intermediate transcript and low KIR protein expression had normal levels of proximal transcripts, suggesting that the novel intermediate promoter is required for protein expression in mature NK cells, and the primary role of the proximal switch is to control stochastic gene activation. The most definitive marker of switch activity is the production of proximal antisense transcripts, and these can only be detected in cells that lack surface KIR expression, suggesting that sense switch transcripts are not translatable, consistent with the discovery of alternatively spliced mRNAs containing stop codons that originate from the proximal switch element. However, additional mechanisms blocking translation of proximal transcripts must exist, since not all proximal transcripts contain alternative exons.

Taken together, the data suggests that there are three distinct phases of *KIR* gene activation: a precursor phase in which the distal transcript activates the proximal switch region; a probabilistic phase wherein the switch promoter determines if the gene will remain in an active state; and a final stage where an intermediate promoter drives KIR expression (**Figure 2**). Although the three promoter system

used by human *KIR* genes is more complex than the two promoter system used by mouse *Ly49* genes to achieve variegated gene expression, it is able to function in a much smaller *KIR* intergenic space (~2 kb) as opposed to the large (~20 kb) intergenic interval found in the *Ly49* genes.

Why Stochastic Switches?

Some have reacted to the concept of random selection of receptors by recalling Einstein's famous quote "God does not play dice," which symbolizes his dislike of Heisenberg's uncertainty principle. However, it is important to note that the phenotypic result of the action of stochastic switches is not random, but is in fact completely reproducible. In inbred strains of mice, individual *Ly49* genes are expressed on the same percentage of NK cells in every mouse with the same genotype. For example, Ly49G is present on 45% of splenic NK cells obtained from every C57BL/6 mouse, but it is only found on 18% of NK cells from NOD mice.[24] This difference can be attributed to weaker sense transcriptional activity of the *Ly49g* switch in NOD mice. Although one cannot know the fate of a single NK cell, the laws of probability dictate the percentage of NK cells that will acquire a given receptor. Thus, it is perhaps more appropriate to say that in the case of variegated gene expression, "evolution takes advantage of probability in the design of complex systems." The probability of gene expression is controlled by the relative affinity of transcription factors for competing binding sites; therefore, the phenotypic result can be programmed through variations in these binding sites. This is a powerful paradigm that can potentially explain some of the decision-making processes that occur during development of an organism. For example, imagine a hematopoietic stem cell that has to make one of several choices: remain as a stem cell; differentiate into a T cell; and differentiate into a B cell. If the lineage-defining transcription factors controlling either B-cell or T-cell fate were under the control of stochastic switches, then either differentiation fate would occur at a predefined frequency, and if neither T-cell– nor B-cell–defining factors were activated, the cell would remain a stem cell. The presence of programmed probabilistic switches governing cell fate choices would provide a mechanism for the generation of appropriate proportions of differentiated cell types required for proper functioning of specialized organ systems.

A genome-wide evaluation of transcription start sites revealed that 11.8% of promoters are bidirectional with start sites separated by less than 1 kb.[25] Of these, 67% have start sites less than 300 bp apart and could thus represent switches. A recent study of RNA expressed by human prefrontal cortex at different postnatal stages provides an indication of how many switches might be operating during the complex process of brain development.[26] High-throughput strand-specific RNA sequencing revealed the presence of a distinct class of bidirectional promoters with a downstream protein coding gene and a 5′ antisense lncRNA, similar to the KIR bidirectional promoters. There were 273 promoters of this class, and they were associated with genes involved in neuronal development and enriched in binding sites for transcription factors associated with neurons, consistent with their potential function as switches controlling brain development.

A key feature of a bidirectional promoter switch is the stable nature of the sense or antisense transcriptional state acquired by the element. Although many bidirectional promoters have been characterized, only a few studies have examined the issue of stable transcriptional states. A careful study of the immediate early enhancer region of murine cytomegalovirus revealed that the divergent transcription of the flanking *ie1/3* and *ie2* genes was mutually exclusive, leading the authors to refer to this element as a genetic switch.[27] A study of the sense and antisense transcripts generated from the expanded hexanucleotide repeats of the *C9ORD72* locus associated with amyotrophic lateral sclerosis revealed cells containing either sense or antisense RNA foci, suggesting stable transcription in either the sense or antisense direction.[28]

Looking forward, it will be of interest to closely examine the promoters of genes that are selectively activated during development. A lineage-defining transcription factor or any gene that is expressed in a variegated manner in a particular tissue would be a good candidate for this type of study. The recent appreciation of the importance of enhancer-derived RNAs[29] and the possibility that some enhancers might represent bidirectional promoters mandate reexamination of enhancers with this new perspective. It is also important to take note of an important lesson learned from the analysis of the MHC receptors: elements controlling differential gene activation may not be directly linked to gene expression, and they may be active prior to the onset of gene expression. It is therefore necessary to study RNA transcripts in precursor cells that do not yet express the gene of interest, in order to identify lncRNAs that may be setting the stage for future gene expression. A detailed analysis of the lncRNAs present in a single precursor cell may reveal variegated expression of sense and antisense transcripts from a bidirectional promoter, leading to the identification of additional probabilistic switches governing developmental choices. Application of this paradigm to additional genes in multiple cell types will undoubtedly reveal many more examples where lncRNA determines the developmental fate of genes.

1. Nemazee, D (2000). Receptor selection in B and T lymphocytes. *Annu Rev Immunol* **18**: 19–51.
2. Lanier, LL (2005). NK cell recognition. *Annu Rev Immunol* **23**: 225–274.
3. Parham, P (2005). MHC class I molecules and KIRs in human history, health and survival. *Nat Rev Immunol* **5**: 201–214.
4. Kubota, A, Kubota, S, Lohwasser, S, Mager, DL and Takei, F (1999). Diversity of NK cell receptor repertoire in adult and neonatal mice. *J Immunol* **163**: 212–216.
5. Yawata, M, Yawata, N, Draghi, M, Partheniou, F, Little, AM and Parham, P (2008). MHC class I-specific inhibitory receptors and their ligands structure diverse human NK-cell repertoires toward a balance of missing self-response. *Blood* **112**: 2369–2380.
6. Pascal, V, Stulberg, MJ and Anderson, SK (2006). Regulation of class I major histocompatibility complex receptor expression in natural killer cells: one promoter is not enough! *Immunol Rev* **214**: 9–21.
7. Saleh, A, Makrigiannis, AP, Hodge, DL and Anderson, SK (2002). Identification of a novel Ly49 promoter that is active in bone marrow and fetal thymus. *J Immunol* **168**: 5163–5169.
8. Saleh, A, Davies, GE, Pascal, V, Wright, PW, Hodge, DL, Cho, EH *et al.* (2004). Identification of probabilistic transcriptional switches in the Ly49 gene cluster: a eukaryotic mechanism for selective gene activation. *Immunity* **21**: 55–66.
9. Santourlidis, S, Trompeter, HI, Weinhold, S, Eisermann, B, Meyer, KL, Wernet, P *et al.* (2002). Crucial role of DNA methylation in determination of clonally distributed killer cell Ig-like receptor expression patterns in NK cells. *J Immunol* **169**: 4253–4261.
10. Rouhi, A, Gagnier, L, Takei, F and Mager, DL (2006). Evidence for epigenetic maintenance of Ly49a monoallelic gene expression. *J Immunol* **176**: 2991–2999.
11. Chan, HW, Miller, JS, Moore, MB and Lutz, CT (2005). Epigenetic control of highly homologous killer Ig-like receptor gene alleles. *J Immunol* **175**: 5966–5974.
12. Christova, R and Oelgeschläger, T (2002). Association of human TFIID-promoter complexes with silenced mitotic chromatin in vivo. *Nat Cell Biol* **4**: 79–82.
13. Chen, D, Hinkley, CS, Henry, RW and Huang, S (2002). TBP dynamics in living human cells: constitutive association of TBP with mitotic chromosomes. *Mol Biol Cell* **13**: 276–284.
14. Tanamachi, DM, Moniot, DC, Cado, D, Liu, SD, Hsia, JK and Raulet, DH (2004). Genomic Ly49A transgenes: basis of variegated Ly49A gene expression and identification of a critical regulatory element. *J Immunol* **172**: 1074–1082.
15. Makrigiannis, AP, Patel, D, Goulet, ML, Dewar, K and Anderson, SK (2005). Direct sequence comparison of two divergent class I MHC natural killer cell receptor haplotypes. *Genes Immun* **6**: 71–83.
16. Davies, GE, Locke, SM, Wright, PW, Li, H, Hanson, RJ, Miller, JS *et al.* (2007). Identification of bidirectional promoters in the human KIR genes. *Genes Immun* **8**: 245–253.
17. Li, H, Pascal, V, Martin, MP, Carrington, M and Anderson, SK (2008). Genetic control of variegated KIR gene expression: polymorphisms of the bi-directional KIR3DL1 promoter are associated with distinct frequencies of gene expression. *PLoS Genet* **4**: e1000254.
18. Parham, P (2006). Taking license with natural killer cell maturation and repertoire development. *Immunol Rev* **214**: 155–160.
19. Stulberg, MJ, Wright, PW, Dang, H, Hanson, RJ, Miller, JS and Anderson, SK (2007). Identification of distal KIR promoters and transcripts. *Genes Immun* **8**: 124–130.
20. Cichocki, F, Hanson, RJ, Lenvik, T, Pitt, M, McCullar, V, Li, H *et al.* (2009). The transcription factor c-Myc enhances KIR gene transcription through direct binding to an upstream distal promoter element. *Blood* **113**: 3245–3253.
21. Cichocki, F, Lenvik, T, Sharma, N, Yun, G, Anderson, SK and Miller, JS (2010). Cutting edge: KIR antisense transcripts are processed into a 28-base PIWI-like RNA in human NK cells. *J Immunol* **185**: 2009–2012.
22. Wright, PW, Huehn, A, Cichocki, F, Li, H, Sharma, N, Dang, H *et al.* (2013). Identification of a KIR antisense lncRNA expressed by progenitor cells. *Genes Immun* **14**: 427–433.
23. Wright, PW, Li, H, Huehn, A, O'Connor, GM, Cooley, S, Miller, JS *et al.* (2014). Characterization of a weakly expressed KIR2DL1 variant reveals a novel upstream promoter that controls KIR expression. *Genes Immun* (epub ahead of print).
24. Belanger, S, Tai, LH, Anderson, SK and Makrigiannis, AP (2008). Ly49 cluster sequence analysis in a mouse model of diabetes: an expanded repertoire of activating receptors in the NOD genome. *Genes Immun* **9**: 509–521.
25. Trinklein, ND, Aldred, SF, Hartman, SJ, Schroeder, DI, Otillar, RP and Myers, RM (2004). An abundance of bidirectional promoters in the human genome. *Genome Res* **14**: 62–66.
26. Hu, HY, He, L and Khaitovich, P (2014). Deep sequencing reveals a novel class of bidirectional promoters associated with neuronal genes. *BMC Genomics* **15**: 457.
27. Simon, CO, Kühnapfel, B, Reddehase, MJ and Grzimek, NK (2007). Murine cytomegalovirus major immediate-early enhancer region operating as a genetic switch in bidirectional gene pair transcription. *J Virol* **81**: 7805–7810.
28. Zu, T, Liu, Y, Bañez-Coronel, M, Reid, T, Pletnikova, O, Lewis, J *et al.* (2013). RAN proteins and RNA foci from antisense transcripts in C9ORF72 ALS and frontotemporal dementia. *Proc Natl Acad Sci USA* **110**: E4968–E4977.
29. Natoli, G and Andrau, JC (2012). Noncoding transcription at enhancers: general principles and functional models. *Annu Rev Genet* **46**: 1–19.