nature genetics

A strand-specific switch in noncoding transcription switches the function of a Polycomb/Trithorax response element

Veronika A Herzog[1,7], Adelheid Lempradl[1,2,7], Johanna Trupke[1], Helena Okulski[1], Christina Altmutter[1], Frank Ruge[1], Bernd Boidol[3], Stefan Kubicek[3], Gerald Schmauss[4], Karin Aumayr[4], Marius Ruf[2], Andrew Pospisilik[2], Andrew Dimond[1,5], Hasene Basak Senergin[1], Marcus L Vargas[6], Jeffrey A Simon[6] & Leonie Ringrose[1]

Polycomb/Trithorax response elements (PRE/TREs) can switch their function reversibly between silencing and activation by mechanisms that are poorly understood. Here we show that a switch in forward and reverse noncoding transcription from the *Drosophila melanogaster vestigial* (*vg*) PRE/TRE switches the status of the element between silencing (induced by the forward strand) and activation (induced by the reverse strand). *In vitro*, both noncoding RNAs inhibit PRC2 histone methyltransferase activity, but, *in vivo*, only the reverse strand binds PRC2. Overexpression of the reverse strand evicts PRC2 from chromatin and inhibits its enzymatic activity. We propose that the interaction of RNAs with PRC2 is differentially regulated *in vivo*, allowing regulated inhibition of local PRC2 activity. Genome-wide analysis shows that strand switching of noncoding RNAs occurs at several hundred Polycomb-binding sites in fly and vertebrate genomes. This work identifies a previously unreported and potentially widespread class of PRE/TREs that switch function by switching the direction of noncoding RNA transcription.

The Polycomb (PcG) and Trithorax Group (TrxG) proteins are essential for development and differentiation in flies and vertebrates[1]. In flies, these proteins provide epigenetic maintenance of repressed (PcG) or activated (TrxG) gene expression states that are initially determined by transcription factors[2]. PcG and TrxG proteins act antagonistically through PRE/TREs[3]. PRE/TREs are switchable regulatory DNA elements that can preserve the memory of an activated or silenced state of their associated genes over several cell generations[2]. How PRE/TREs switch between active and silent states and how this function is modulated in different developmental contexts are poorly understood. Many PcG/TrxG-binding sites are transcribed into long noncoding RNAs[4], and several noncoding RNAs are involved in PcG/TrxG function, although the mechanisms by which they act are currently the subject of debate[5,6].

Here we show that the *Drosophila vg* PRE/TRE is transcribed in a bidirectional and developmentally regulated manner. Embryonic transcription of the PRE/TRE reverse strand correlates with *vg* gene activation, whereas larval transcription of the PRE/TRE forward strand correlates with *vg* gene repression. Ectopic transcription of each strand leads to *trans*-activation or *trans*-silencing at the endogenous *vg* locus. Thus, *in vivo*, two different strands transcribed from the same regulatory element have opposite effects on the properties of the element.

TrxG and PcG proteins have been shown to bind to several noncoding RNAs, and it has been proposed that these RNAs specifically recruit the TrxG and PcG proteins to their target genes[7–12] (reviewed in ref. 6). However, it has been shown that the Polycomb repressive complex 2 (PRC2) binds RNA *in vitro* with high affinity but no specificity[13], challenging the hypothesis that PRC2 is recruited to its target genes via specific interactions with noncoding RNAs. In contrast to these findings, it has recently been reported that some RNAs show specificity in both binding to PRC2 and inhibiting its activity *in vitro*[14]. Thus, the specificity of PRC2–noncoding RNA interactions, in terms of both binding and inhibition, is currently disputed.

Here we report that RNA promiscuously inhibits the enzymatic activity of PRC2 *in vitro* at equimolar RNA-PRC2 concentrations and with no specificity for the different RNAs tested, including those from the *vg* PRE/TRE. However, *in vivo*, the reverse (but not the forward) *vg* PRE/TRE strand binds with high specificity to the PRC2 component enhancer of zeste (E(Z)). Overexpression of the reverse strand leads to both local inhibition of the catalytic activity of PRC2 and the eviction of E(Z) from chromatin. Thus, we show that, despite nonspecific interaction *in vitro*, the interaction of different noncoding RNAs and PcG proteins *in vivo* is highly specific and has the potential to inhibit PRC2 at its site of action in a developmentally regulated manner. In summary, this work has broad implications for understanding the dynamic, dual nature of PRE/TREs and the role of specific noncoding RNAs in switching their function.

[1]IMBA (Institute of Molecular Biotechnology), Vienna, Austria. [2]Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany. [3]CeMM (Research Center for Molecular Medicine), Vienna, Austria. [4]IMP (Institute of Molecular Pathology), Vienna, Austria. [5]The Babraham Institute, Babraham Research Campus, Cambridge, UK. [6]Department of Genetics, Cell Biology and Development, University of Minnesota, Minneapolis, Minnesota, USA. [7]These authors contributed equally to this work. Correspondence should be addressed to L.R. (leonie.ringrose@imba.oeaw.ac.at).

Received 20 April; accepted 14 July; published online 10 August 2014; doi:10.1038/ng.3058

RESULTS

Opposite RNA strands of the *vg* PRE/TRE correlate inversely with *vg* mRNA

The *Drosophila vg* gene is essential for determining correct cell identity in the wing and haltere[15,16] and in specific muscle lineages[15,17,18], and it is also expressed in the embryonic central nervous system[16,19]. The *vg* gene is a target of PcG regulation in embryos[20,21] and in the wing[22–24]. The locus contains two sites that are enriched for PcG proteins in chromatin immunoprecipitation (ChIP) experiments: one at the promoter and one downstream of the gene (**Fig. 1a**)[20,25]. The latter site contains a PRE/TRE that has been characterized in several transgenic assays and can maintain both silencing and activation of reporter gene activity in a PcG- and TrxG-dependent manner[22,23,26].

We detected transcripts from the downstream *vg* PRE/TRE in tissues in which *vg* mRNA is expressed (**Supplementary Fig. 1**). Mapping of the 5′ and 3′ ends of the PRE/TRE transcripts by 5′ RACE, primer walking and cDNA cloning in embryos, larval brains and wing discs identified several RNA polymerase II (Pol II) transcripts arising from both strands of the PRE/TRE (**Fig. 1b**, **Supplementary Figs. 2** and **3**, and **Supplementary Table 1**). Additional mapping showed that the forward-strand transcripts were distinct from the *vg* mRNA itself, as no transcription was detected between the 3′ end of *vg* and the 5′ ends of the PRE/TRE transcripts. Interestingly, we observed a developmental switch in strand preference for the PRE/TRE transcripts: in embryos, both strands of the PRE/TRE were transcribed, whereas, in larval wing discs and brains, only transcripts from the forward strand were detected (**Fig. 1b** and **Supplementary Fig. 2**). Thus, PRE/TRE transcripts are highly developmentally regulated.

Double *in situ* hybridizations to each strand of the PRE/TRE and *vg* mRNA showed that the reverse-strand transcripts of the PRE/TRE and *vg* mRNA became visible at 10 h of embryogenesis. Both of these transcripts were present in the somatic muscle lineage[27] (**Fig. 1c–j**). The PRE/TRE transcript was detectable in muscle founder cells before the appearance of *vg* mRNA, both in the shared cytoplasm of fused

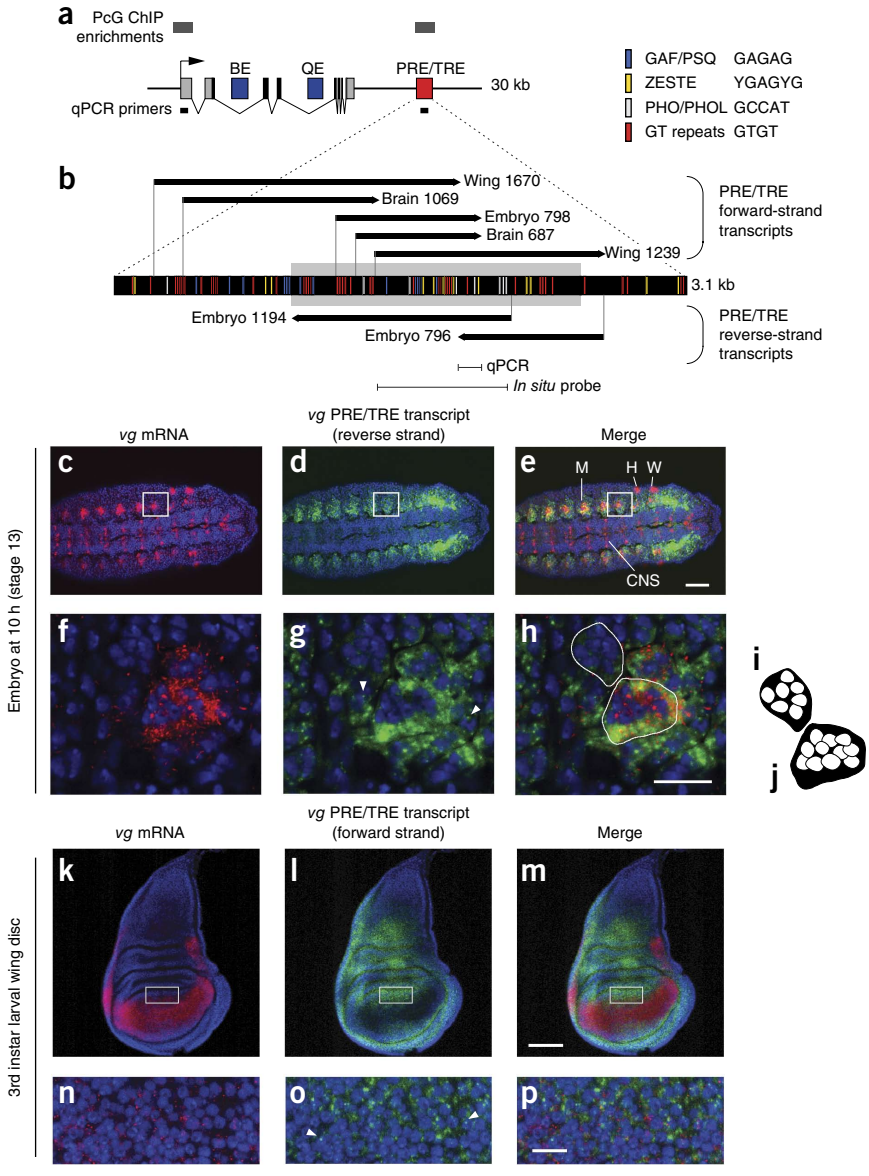

Figure 1 Coding and noncoding transcription at the *vg* locus. (**a**) *vg* locus showing the *vg* intron and exons, the PRE/TRE (red) and enhancers (blue). BE, boundary enhancer[38]; QE, quadrant enhancer[39]. Sites of ChIP enrichment[20,25] (**Supplementary Figs. 2** and **10**) are shown above the schematic (gray bars). The positions of the quantitative PCR primers used in **Supplementary Figure 1** are shown below the schematic. (**b**) Noncoding transcripts at the *vg* PRE/TRE (see also **Supplementary Figs. 2** and **3**). The tissue in which each transcript was detected and transcript length in nucleotides are given. Colored bars indicate PRE/TRE DNA motifs (in the ZESTE motif, Y = C or T). The gray box represents the 1.6-kb core PRE/TRE[26] used for transgenes. (**c**–**p**) Double *in situ* hybridization on embryos (**c**–**h**) and 3rd instar larval wing discs (**k**–**p**) showing *vg* mRNA (red), PRE/TRE reverse- or forward-strand transcript (green) and 4′,6-diamidino-2-phenylindole (DAPI; blue). Anterior is to the right, posterior is to the left; the ventral view is shown for embryos. (**c**–**e**) Embryos at 10 h (stage 13) with staining for *vg* mRNA (**c**) and the reverse strand of the *vg* PRE/TRE transcript (**d**) together with the merged image (**e**). W, wing primordia; H, haltere primordia; M, muscle precursors; CNS, central nervous system. Scale bar, 75 µm. (**f**–**h**) Single optical slice at high magnification of the boxes in **c**–**e** showing staining for *vg* mRNA (**f**) and the reverse strand of the *vg* PRE/TRE transcript (**g**) together with the merged image (**h**) in muscle precursors with several nuclei. Arrowheads indicate nuclear dots. Scale bar, 20 µm. (**i**,**j**) Diagram of the multinucleate precursors outlined in **h**, showing a founder colony (**i**) and an advanced colony (**j**). (**k**–**p**) Wing discs. Single optical slices showing that *vg* mRNA and the PRE/TRE forward-strand transcript are expressed in a complementary pattern. (**k**–**m**) Staining for *vg* mRNA (**k**) and the forward strand of the *vg* PRE/TRE transcript (**l**) together with the merged image (**m**). Scale bar, 100 µm. (**n**–**p**) High magnification of the boxes in **k**–**m** showing staining for *vg* mRNA (**n**) and the forward strand of the *vg* PRE/TRE transcript (**o**) together with the merged image (**p**). Arrowheads indicate nuclear dots. Scale bar, 20 µm. Embryo and wing disc *in situ* hybridizations were performed 3 times, and 5–10 discs or 10–20 embryos were analyzed.

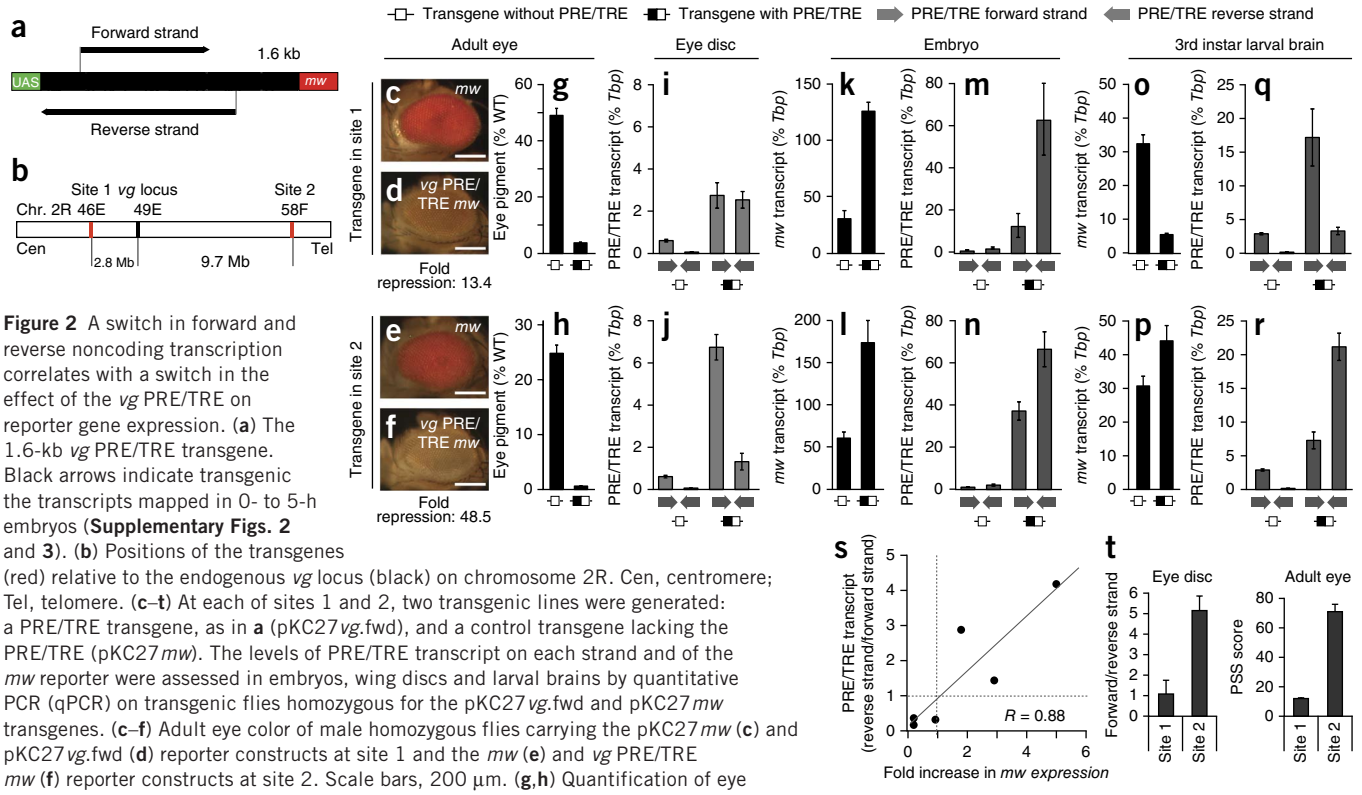

Figure 2 A switch in forward and reverse noncoding transcription correlates with a switch in the effect of the *vg* PRE/TRE on reporter gene expression. (**a**) The 1.6-kb *vg* PRE/TRE transgene. Black arrows indicate transgenic the transcripts mapped in 0- to 5-h embryos (**Supplementary Figs. 2** and **3**). (**b**) Positions of the transgenes (red) relative to the endogenous *vg* locus (black) on chromosome 2R. Cen, centromere; Tel, telomere. (**c**–**t**) At each of sites 1 and 2, two transgenic lines were generated: a PRE/TRE transgene, as in **a** (pKC27*vg*.fwd), and a control transgene lacking the PRE/TRE (pKC27*mw*). The levels of PRE/TRE transcript on each strand and of the *mw* reporter were assessed in embryos, wing discs and larval brains by quantitative PCR (qPCR) on transgenic flies homozygous for the pKC27*vg*.fwd and pKC27*mw* transgenes. (**c**–**f**) Adult eye color of male homozygous flies carrying the pKC27*mw* (**c**) and pKC27*vg*.fwd (**d**) reporter constructs at site 1 and the *mw* (**e**) and *vg* PRE/TRE *mw* (**f**) reporter constructs at site 2. Scale bars, 200 µm. (**g**,**h**) Quantification of eye pigment in **c**–**f** (ref. 26) for transgenes at site 1 (**g**) and site 2 (**h**). Error bars show the maximum and minimum of two independent measurements for each line. (**i**,**j**) Strand-specific RT-qPCR analysis of PRE/TRE transcription in 3rd instar larval eye discs for transgenes at site 1 (**i**) and site 2 (**j**), shown as the percentage of *Tbp* (TATA-binding protein). (**k**–**n**) RT-qPCR of *mw* reporter transcript for transgenes at site 1(**k**) and site 2 (**l**) and PRE/TRE transcript for transgenes at site 1 (**m**) and site 2 (**n**) on each strand in embryos. (**o**–**r**) RT-qPCR of *mw* reporter transcript for transgenes at site 1 (**o**) and site 2 (**p**) and PRE/TRE transcript for transgenes at site 1 (**q**) and site 2 (**r**) on each strand in 3rd instar larval brains. For all RT-qPCR analyses, the mean and s.e.m. of four independent cDNAs prepared from two independent RNAs are shown. (**s**) For each transgenic line and tissue, the relative proportion of PRE/TRE transcripts from each strand is shown on the *y* axis as the reverse strand/forward strand ratio. The *x* axis shows the fold increase in expression of the *mw* reporter, calculated as the *mw* level of the PRE/TRE transgene divided by the *mw* level of the control transgene. (**t**) Left, ratio of the transgenic forward- and reverse-strand transcripts in 3rd instar larval eye discs (data from **i**,**j**) at sites 1 and 2. The mean and s.e.m. of four independent cDNAs prepared from two independent RNAs are shown. Right, PSS scores of transgenes at sites 1 and 2 (from ref. 26). PSS scores were calculated from adult eye pigment levels as the heterozygote/homozygote ratio for each transgenic line, normalized to the same ratio calculated for the no PRE/TRE control line at the same landing site (see **Fig. 3d** in ref. 26). Error bars show the maximum and minimum of two independent measurements for each line.

nuclei and as nuclear dots (**Fig. 1f**–**j** and **Supplementary Fig. 4**). In later embryos, both the PRE/TRE reverse strand and *vg* mRNA were expressed at high levels in all somatic muscles (**Supplementary Fig. 4j**–**l**). Thus, the reverse strand of the *vg* PRE/TRE is transcribed specifically in cells of the muscle lineage in which the expression of *vg* mRNA later becomes activated.

In contrast, the spatial pattern of PRE/TRE transcription from the forward strand was reciprocal to that of *vg* mRNA, with the highest levels of *vg* mRNA corresponding to the lowest levels of PRE/TRE transcripts and vice versa. This reciprocal relationship was observed to some extent in late embryos (**Supplementary Fig. 5**) and, most strikingly, in 3rd instar larval wing discs (**Fig. 1k**–**p**) and brains (**Supplementary Fig. 6**). In summary, these results demonstrate that transcripts from opposite strands of the *vg* PRE/TRE correlate inversely with *vg* mRNA levels.

PRE/TRE strand switching correlates with reporter switching

To gain further insight into the role of strand-specific PRE/TRE transcription, we used a transgenic reporter assay in which the central 1.6 kb of the downstream *vg* PRE/TRE was linked to a *miniwhite* (*mw*) reporter gene (**Fig. 2a**). This 1.6-kb *vg* PRE/TRE confers PcG-dependent repression and strong pairing-sensitive silencing (PSS) of the *mw* reporter in adult eyes[26]. ChIP followed by sequencing (ChIP-seq) analysis of the endogenous *vg* locus in *Drosophila* cells[25] and embryos (**Supplementary Fig. 2b**) showed substantial overlap of the 1.6-kb PRE/TRE with regions enriched for Polycomb (PC), E(Z) and histone H3 lysine 27 trimethylation (H3K27me3). Most notably, we chose to use this PRE/TRE subfragment because it contains a subset of embryonic and larval promoters, namely those driving transcription through the core segment of the PRE/TRE (**Fig. 1b** and **Supplementary Fig. 2a**). PRE/TRE constructs and control constructs lacking the PRE/TRE were each independently integrated at two distinct genomic locations[26] (**Fig. 2a**,**b**). Mapping of the 5′ and 3′ ends of the transgenic transcripts was performed in early (0- to 5-h) embryos, in which there was no background of endogenous PRE/TRE transcripts. At both integration sites, the transgene produced transcripts with the same 5′ ends as the endogenous embryonic ones, confirming that the transgenic PRE/TRE uses its own promoters and is not simply transcribed from promoters flanking the integration site (**Fig. 2a** and **Supplementary Fig. 3**). However, the levels of the transgenic transcripts were higher than the endogenous levels and were different at the two sites, probably owing to the effects of flanking enhancers[26,28] (**Fig. 2i**,**j**,**m**,**n**,**q**,**r**). These differences enabled comparison of the levels of each PRE/TRE strand with reporter activity at both sites.

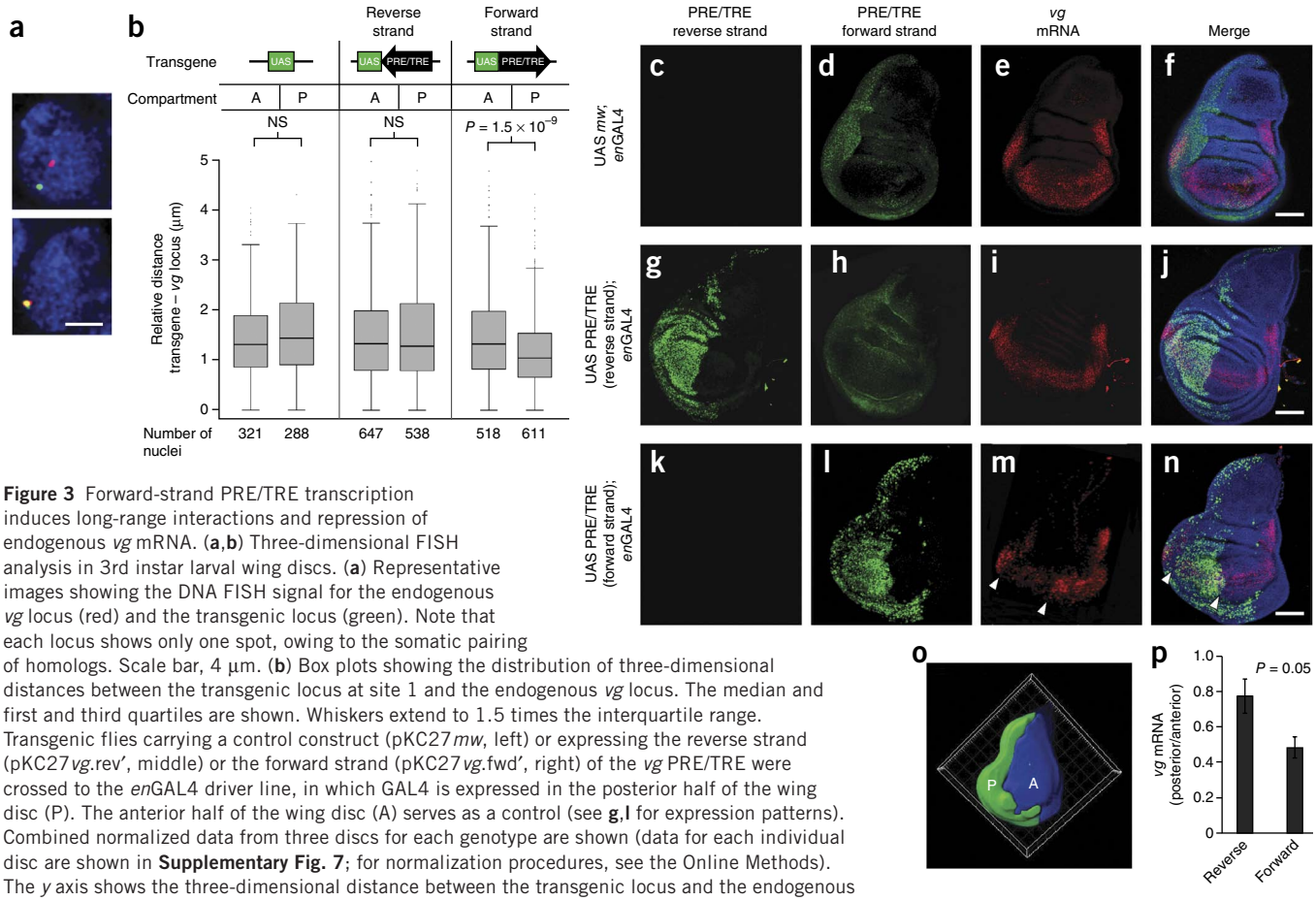

Figure 3 Forward-strand PRE/TRE transcription induces long-range interactions and repression of endogenous *vg* mRNA. (**a**,**b**) Three-dimensional FISH analysis in 3rd instar larval wing discs. (**a**) Representative images showing the DNA FISH signal for the endogenous *vg* locus (red) and the transgenic locus (green). Note that each locus shows only one spot, owing to the somatic pairing of homologs. Scale bar, 4 µm. (**b**) Box plots showing the distribution of three-dimensional distances between the transgenic locus at site 1 and the endogenous *vg* locus. The median and first and third quartiles are shown. Whiskers extend to 1.5 times the interquartile range. Transgenic flies carrying a control construct (pKC27*mw*, left) or expressing the reverse strand (pKC27*vg*.rev′, middle) or the forward strand (pKC27*vg*.fwd′, right) of the *vg* PRE/TRE were crossed to the *en*GAL4 driver line, in which GAL4 is expressed in the posterior half of the wing disc (P). The anterior half of the wing disc (A) serves as a control (see **g**,**l** for expression patterns). Combined normalized data from three discs for each genotype are shown (data for each individual disc are shown in **Supplementary Fig. 7**; for normalization procedures, see the Online Methods). The *y* axis shows the three-dimensional distance between the transgenic locus and the endogenous *vg* locus. The average nuclear radius was 3.6 µm. Induced transcription of the forward strand but not the reverse strand leads to a significant decrease in the distances between the transgene and the endogenous locus in the posterior compartment in comparison to the anterior compartment (NS, *P* value ≥ 0.1, Wilcoxon rank-sum test). (**c**–**n**) Double *in situ* hybridization to detect the reverse strand or the forward strand of the *vg* PRE/TRE and *vg* mRNA in the wing discs of transgenic 3rd instar larvae carrying the reporters and control constructs integrated at site 1, when crossed to the *en*GAL4 driver. Posterior is to the left, anterior is to the right. Scale bars, 100 µm. (**c**–**f**) Control transgene without the PRE/TRE (pKC27*mw*) with staining for the reverse strand (**c**), forward strand (**d**) and *vg* mRNA (**e**) together with the merged image (**f**). **c** and **d** show different discs, **d** and **e** show the same disc, and **f** shows a merged image of **d** and **e**. (**g**–**j**) Transgene overexpressing the PRE/TRE reverse strand (pKC27*vg*.rev′) with staining for the reverse strand (**g**), forward strand (**h**) and *vg* mRNA (**i**) together with the merged image (**j**). **g** and **h** show different discs, **g** and **i** show the same disc, and **j** shows a merged image of **g** and **i**. (**k**–**n**) Transgene overexpressing the forward strand (pKC27*vg*.fwd′) with staining for the reverse strand (**k**), forward strand (**l**) and *vg* mRNA (**m**) together with the merged image (**n**). **k** and **l** show different discs, **l** and **m** show the same disc, and **n** shows a merge of **l** and **m**. Blue, DAPI. Arrowheads indicate the borders of the domain of downregulation of *vg* mRNA (see also **Supplementary Video 1**). Each *in situ* hybridization was performed at least twice, and 4–10 discs were analyzed for each line. (**o**) For each disc, the domain of *en*GAL4 expression was masked in three dimensions using the noncoding RNA signal (P), and the remaining volume was separately masked (A). (**p**) *vg* mRNA signal intensity was quantified in three dimensions within the anterior and posterior domains for individual discs, and the posterior/anterior ratio was calculated. The mean and s.e.m. of the posterior/anterior ratio for four discs for each line are shown. *P* values were calculated by one-tailed Student's *t* test.

In adult eyes, as shown previously[26], the PRE/TRE conferred strong repression of the *mw* reporter. Repression was stronger at site 2 (48.5-fold) than at site 1 (13.4-fold; **Fig. 2c**–**h**)[26]. Interestingly, in the eye imaginal discs of 3rd instar larvae, the levels of forward-strand transcripts correlated with the extent of repression at each site (**Fig. 2g**–**j**). Remarkably, in embryos, both the regulatory properties and the strand preference of the PRE/TRE were reversed. At both sites, we observed robust activation of the reporter, accompanied by higher levels of the reverse PRE/TRE strand than the forward strand (**Fig. 2k**–**n**). Finally, in larval brains, the transgene at site 1 showed higher forward- than reverse-strand PRE/TRE transcript levels, accompanied by repression of the reporter (**Fig. 2o**,**q**). Conversely, at site 2, the reverse strand of the PRE/TRE was present at higher levels than the forward strand, and the reporter was activated (**Fig. 2p**,**r**). In summary, this analysis shows that lines and tissues in which transcripts for the forward PRE/TRE strand are present at higher levels than reverse-strand transcripts show repression of the *mw* reporter, whereas those with higher levels of reverse-strand transcripts show activation of the *mw* reporter (summarized in **Fig. 2s**). Thus, these data document a remarkable correlation between the switch in forward and reverse noncoding transcription and the switch in the effect of the *vg* PRE/TRE on reporter gene expression. We name this element a 'GEAR box' (gene expression alternating RNA) element.

The *vg* PRE/TRE forward strand represses endogenous *vg* mRNA

To determine whether forward and reverse GEAR box transcripts themselves have a role in regulating gene expression, we examined the effects of ectopically overexpressing each strand from the transgene on the endogenous *vg* locus. PRE/TREs have been shown to mediate

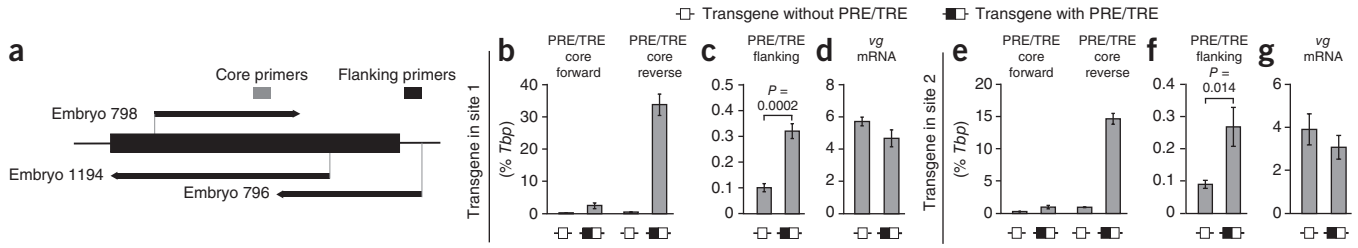

Figure 4 The reverse-strand PRE/TRE transcript activates the endogenous *vg* PRE/TRE. (**a**) Map of the endogenous *vg* PRE/TRE locus showing the position of the sequence present in transgenes (thick bar). (**b**–**g**) RT-qPCR on cDNA from 0- to 5-h embryos. (**b**–**d**) Transgenes integrated at site 1 with qPCR using core primers (**b**), flanking primers (**c**) or primers detecting *vg* mRNA (**d**). (**e**–**g**) Transgenes integrated at site 2 with qPCR using core primers (**e**), flanking primers (**f**) or primers detecting *vg* mRNA (**g**). Error bars show the s.e.m. of 2–6 independent cDNAs prepared from at least 2 independent RNAs. *P* values were calculated by one-tailed Student's *t* test. The positions of the PCR primers used are indicated in **a**.

long-range interactions[29]. We observed that the relative levels of the *vg* PRE/TRE forward strand in eye discs correlated with the extent of PSS at each site (**Fig. 2t**)[26], suggesting that the forward strand might be involved in mediating or stabilizing pairing between homologous PRE/TREs and might thus facilitate PSS. To determine whether the transgenic *vg* PRE/TRE interacts physically with the endogenous *vg* locus upon induced transcription of either the forward or reverse noncoding RNA strands, we performed three-dimensional DNA FISH (**Fig. 3a**,**b**). We used the GAL4-UAS system in combination with transgenes designed to overexpress either the reverse or forward strand, integrated at site 1 (**Figs. 2a**,**b** and **3b**). *engrailed* GAL4 (*en*GAL4, which drives GAL4 from the *engrailed* promoter) was used to drive elevated transcription of the transgenic PRE/TRE forward or reverse strand in the posterior compartment of the wing disc, while the anterior compartment of the same disc lacking *en*GAL4 served as a negative control (see **Fig. 3g**,**l** for the schematic of the *en*GAL4 expression domain).

Induced transcription of the control transgene lacking the PRE/TRE or induction of the PRE/TRE reverse-strand transcript had no significant effect on the distance distribution between the endogenous *vg* locus and the transgene (compare A and P in **Fig. 3b** and **Supplementary Fig. 7**), showing that transcription *per se* has little effect on the localization of the transgene relative to the endogenous *vg* locus. In contrast, induction of the PRE/TRE forward strand led to a significant increase in the proximity of the transgene to the endogenous locus (**Fig. 3b**) in two of the three discs examined (**Supplementary Fig. 7**). This effect was dependent on transcription and was strand specific, suggesting that the forward strand but not the reverse strand might induce or stabilize long-range interactions between two copies of the PRE/TRE.

To ascertain whether transgenic PRE/TRE transcription also affects the regulation of endogenous *vg* mRNA levels, we used *en*GAL4 to drive transcription of either the forward or reverse strand from site 1 and analyzed the levels of *vg* mRNA (**Fig. 3c**–**p**). In control lines in which *en*GAL4 was used to drive expression of the UAS-*mw* transgene lacking the PRE/TRE (**Fig. 3c**–**f**) or to drive transcription of the reverse PRE/TRE strand (**Fig. 3g**–**j**), the *vg* mRNA expression pattern was identical to that for wild-type (non-transgenic) discs (**Fig. 1k**). In contrast, *en*GAL4-driven overexpression of the PRE/TRE forward strand from the transgene caused substantial downregulation of *vg* mRNA levels within the *en*GAL4 expression domain (**Fig. 3k**–**p**; see also **Supplementary Video 1**). Analysis of *en*GAL4 expression levels, using RNA interference (RNAi) for *vg* as a sensor, showed that this domain corresponded with the highest levels of *en*GAL4 expression (**Supplementary Fig. 8**). Taken together, these results demonstrate that ectopic expression of the *vg* PRE/TRE forward strand causes both long-range interactions and repression of the endogenous *vg* locus.

The PRE/TRE reverse strand activates endogenous PRE/TRE

The above experiments demonstrate a role for the forward strand in *vg* regulation, but the reverse strand had little effect in these assays. Because the reverse strand is not normally expressed in larval tissues (**Fig. 1b**), we reasoned that its principal function might be in embryos, where it is normally present. To address this possibility, we examined the effects of transgenic transcripts on the endogenous *vg* locus in embryos (**Fig. 4**). In early embryos (0–5 h), neither endogenous *vg* mRNA nor PRE/TRE transcripts were detectable (**Supplementary Fig. 1c**), allowing the unambiguous identification of induced transcripts from the endogenous locus. We took advantage of the fact that the transgenic reverse-strand promoter was active at this early stage without the need for GAL4-induced transcription (**Fig. 4a**,**b**,**e**). To distinguish between transgenic and endogenous PRE/TRE transcripts, we designed PCR primers to specifically detect the 5′ end of the 'embryo 796' transcript (the 796-nucleotide embryonic transcript;

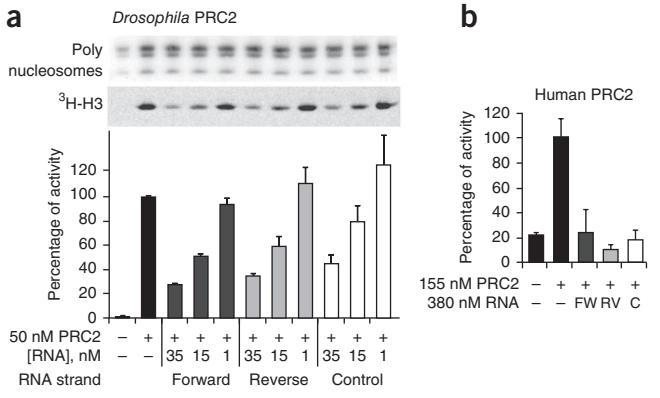

Figure 5 RNAs inhibit PRC2 HMTase activity *in vitro*. (**a**) *In vitro* HMTase assay. *Drosophila* PRC2 was incubated with HeLa polynucleosomes and ^{3}H-labeled S-adenosylmethionine together with a 1.6-kb RNA *in vitro* transcribed from the 1.6-kb PRE/TRE shown in **Figure 4a**, in forward or reverse orientation, or a bacterial control RNA of the same length. The experiment was performed three times. The mean and s.d. of two quantifications of a representative experiment are shown. (**b**) Human PRC2 was incubated with an H3 tail peptide (residues 21–40) and with RNAs as in **a**; detection of H3K27me3 was performed using the Dissociation-Enhanced Lanthanide Fluorescent Immunoassay (DELFIA). The experiment was performed three times. The mean and s.e.m. of three technical replicates within a representative experiment are shown. All RNAs show similar inhibition of *Drosophila* and human PRC2 HMTase activity.

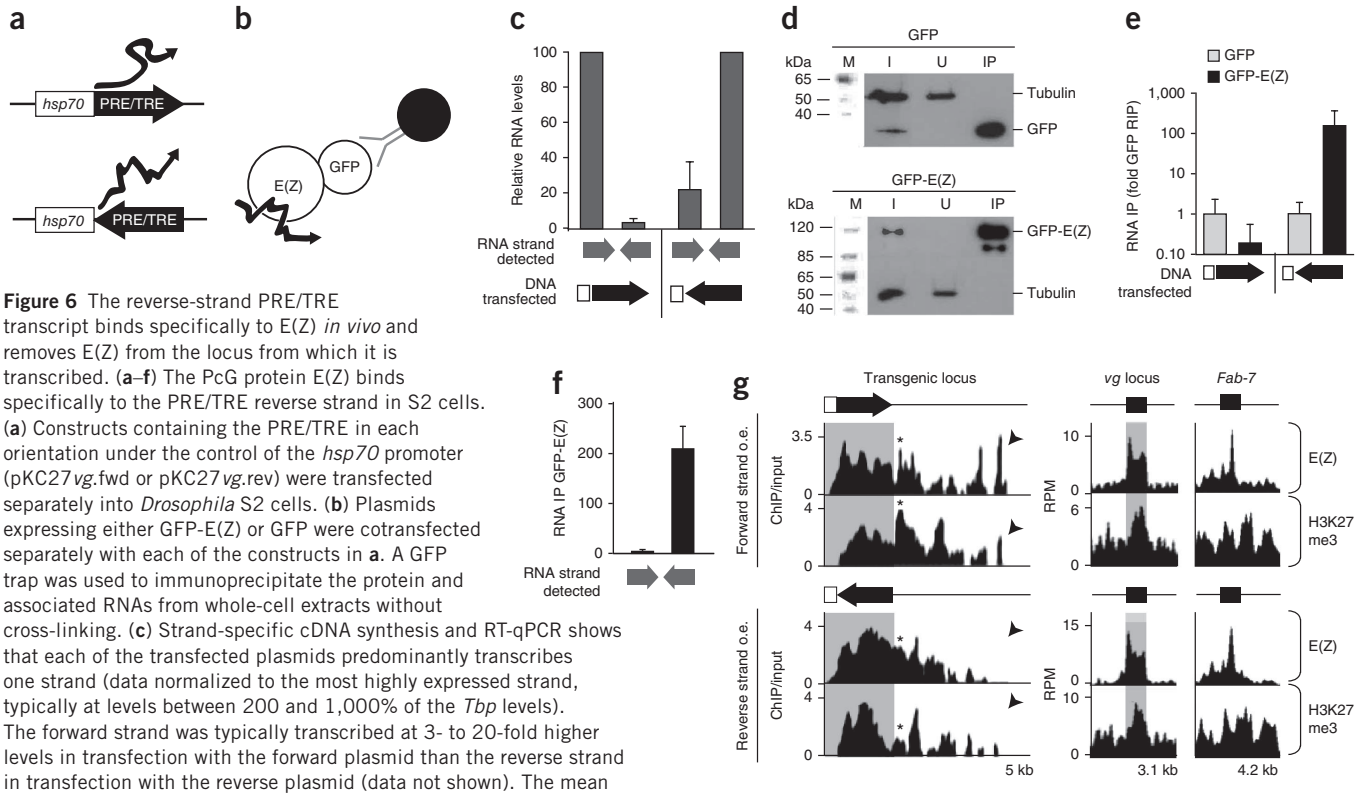

Figure 6 The reverse-strand PRE/TRE transcript binds specifically to E(Z) *in vivo* and removes E(Z) from the locus from which it is transcribed. (**a**–**f**) The PcG protein E(Z) binds specifically to the PRE/TRE reverse strand in S2 cells. (**a**) Constructs containing the PRE/TRE in each orientation under the control of the *hsp70* promoter (pKC27*vg*.fwd or pKC27*vg*.rev) were transfected separately into *Drosophila* S2 cells. (**b**) Plasmids expressing either GFP-E(Z) or GFP were cotransfected separately with each of the constructs in **a**. A GFP trap was used to immunoprecipitate the protein and associated RNAs from whole-cell extracts without cross-linking. (**c**) Strand-specific cDNA synthesis and RT-qPCR shows that each of the transfected plasmids predominantly transcribes one strand (data normalized to the most highly expressed strand, typically at levels between 200 and 1,000% of the *Tbp* levels). The forward strand was typically transcribed at 3- to 20-fold higher levels in transfection with the forward plasmid than the reverse strand in transfection with the reverse plasmid (data not shown). The mean and s.e.m. of three independent immunoprecipitation experiments are shown. (**d**) Protein blot using antibody to GFP and tubulin, showing GFP (top) and GFP-E(Z) (bottom) present in input (I) but not in the unbound fraction (U) and enriched after the GFP trap (IP). (**e**) Enrichment of PRE/TRE RNA after the GFP trap on cells transfected with either plasmid as indicated and with GFP or GFP-E(Z). Immunoprecipitate levels were normalized to input and were further normalized to *GFP* mRNA levels, and they were controlled to exclude DNA contamination. The mean and s.e.m. of at least three independent immunoprecipitation experiments are shown. cDNA synthesis was performed by random priming. Whereas RNA does not coimmunoprecipitate with GFP-E(Z) after transfection with the forward-strand construct, GFP-E(Z) shows PRE/TRE RNA enrichment of over 100-fold with transfection with the reverse-strand plasmid, which predominantly transcribes the reverse strand (**c**), after immunoprecipitation (fourth column). (**f**) Immunoprecipitate was subjected to strand-specific cDNA analysis, showing that only the reverse strand is detectable. The mean and s.d. of two experiments are shown; the scale shows immunoprecipitation as a percentage of the *Tbp* levels in the input. (**g**) ChIP-seq analysis of E(Z) and H3K27me3 in 0- to 16-h embryos overexpressing forward-strand (top; pKC27*vg*.fwd′) or reverse-strand (bottom; pKC27*vg*.rev′) noncoding RNAs. Left, the transgenic locus. The black arrows indicate the position and orientation of the 1.6-kb PRE/TRE. The gray box represents the 1.6-kb PRE/TRE sequence identical between the transgene and the endogenous *vg* locus. The white area denotes the sequence unique to the transgenic locus, corresponding to the ampicillin resistance gene of pKC27. The asterisks and arrowheads indicate differences between forward- and reverse-strand overexpression. Data are shown as ChIP/input ratios to correct for different copy numbers for PRE/TRE and the flanking sequence. Right, no difference was detectable at the endogenous *vg* locus (chr. 2R: 8,791,079–8,794,193; see also **Supplementary Fig. 10**) or at the *Fab-7* PRE/TRE (chr. 3R: 12,723,848–12,728,057). Thick lines indicate the position of the PRE/TRE. RPM, reads per million. The experiment was performed three times for H3K27me3 on pKC27*vg*.fwd′, once for H3K27me3 on pKC27*vg*.rev′ and once for E(Z) on both pKC27*vg*.fwd′ and pKC27*vg*.rev′. Tracks from one representative ChIP experiment are shown.

Figs. 1b and **4a**). This sequence is not present on the transgene and is therefore unique to the endogenous locus (**Fig. 4a**, flanking primers). In early (0- to 5-h) embryos carrying control transgenes lacking the PRE/TRE, endogenous PRE/TRE transcription was barely detectable (**Fig. 4b**,**e**). In contrast, the PRE/TRE transgene was highly transcribed on the reverse strand (**Fig. 4b**,**e**, fourth column). Remarkably, this ectopic reverse-strand transcription was accompanied by a threefold upregulation of transcript levels from the endogenous flanking site (**Fig. 4c**,**f**), although the levels of *vg* mRNA itself were unaffected (**Fig. 4d**,**g**). Additional mapping by primer walking confirmed that these ectopically induced endogenous transcripts corresponded to the endogenous reverse-strand transcript embryo 796 (**Fig. 4a** and **Supplementary Fig. 3**). In summary, these results show that early embryonic transcription of the reverse strand of the transgenic *vg* PRE/TRE activates reverse-strand transcription at the endogenous *vg* PRE/TRE.

RNA promiscuously inhibits PRC2 H3K27 methyltransferase activity

Our *in vivo* data show that the forward and reverse transcripts of the *vg* PRE/TRE have different properties. *In vitro* binding assays with recombinant vertebrate and fly PRC2 proteins (EZH2-EED or EZH2 and E(Z) alone) have reported specific binding to some RNAs and not to others[8,12] and specific inhibition by some RNAs[14]. However, a recent study reports that recombinant human PRC2 binds RNA *in vitro* with high affinity and with no specificity for different RNAs[13]. To determine whether there is any difference in the interaction of the different strands of the *vg* PRE/TRE with PRC2 *in vitro*, we asked whether the forward and reverse strands might inhibit or stimulate the enzymatic activity of recombinant PRC2 to different extents. To this end, we assayed the *in vitro* H3K27 methyltransferase (HMTase) activity of recombinant *Drosophila* and human PRC2 in the presence of RNAs. We tested the *vg* PRE/TRE forward or reverse strand or

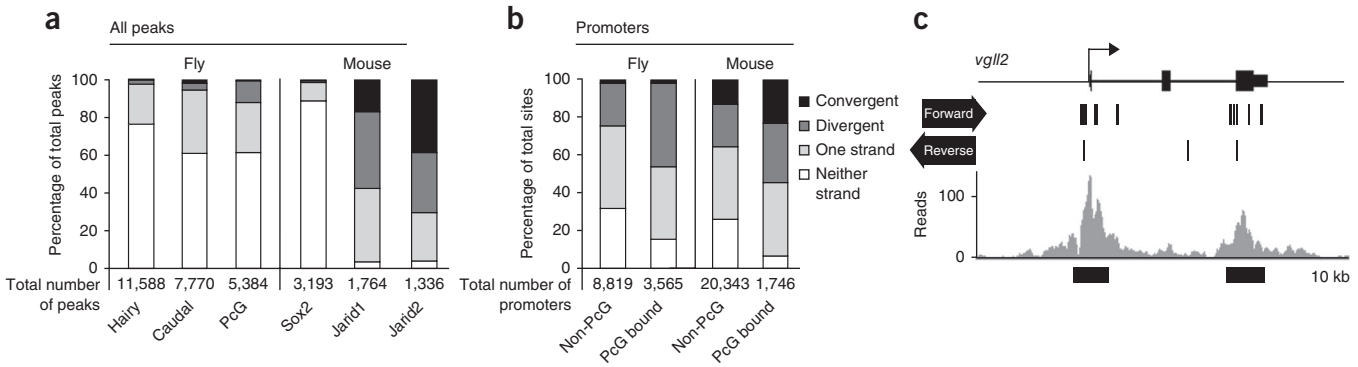

Figure 7 Elements with GEAR box potential are widespread. (**a**) Fly: ChIP peaks from ref. 21 with CAGE tags from ref. 31. Mouse: ChIP peaks from refs. 33,34 compared with CAGE tags from the FANTOM3 CAGE data set[32]. Suz12 comparison gave similar results (**Supplementary Table 3**). The percentage of total peaks overlapping with CAGE tags in the categories shown is indicated. For convergent CAGE tags, tags on opposite strands are transcribed toward each other; for divergent CAGE tags, tags on opposite strands are transcribed away from each other. (**b**) The CAGE tag status of PcG ChIP peaks at promoters was compared to that of non-PcG-bound promoters. Fly: data as for **a**. Mouse: combined Jarid2 and Suz12 data from ref. 34 were used for PcG peaks. In most cases, bidirectional transcription occurs significantly more frequently at PcG-bound sites than at other sites. The frequency of the occurrence of the category of interest (convergent or divergent) in comparison to all other categories, including no transcription, was compared for pairs of data sets (for each comparison, the PcG-bound sites were compared to non-PcG-bound sites). *P* values were calculated by one-sided Yates χ^2 test. In all cases except caudal (convergent) and Jarid1 (divergent), the category of interest was more abundant in PcG-bound sites than in non-PcG-bound sites. For fly promoters (convergent), $P = 0.19$; for all other categories, $P < 4 \times 10^{-6}$. See also the Online Methods and **Supplementary Tables 2** and **3**. (**c**) Mouse *Vgll2* locus: gray, Suz12 binding in ref. 34; black lines, CAGE tag occurrence[32]. Black bars below the plot indicate potential GEAR box elements.

control RNAs (of the same length but derived from bacterial plasmid sequences) (**Fig. 5** and **Supplementary Fig. 9**). This analysis showed that all RNAs tested caused inhibition of both *Drosophila* and human PRC2 HMTase activity to equal extents. A similar effect was also observed for DNA (**Supplementary Fig. 9**). Robust inhibition occurred at approximately equimolar concentrations of PRC2 and RNA. This finding suggests that there is no inherent difference in the ability of different RNAs to inhibit PRC2 HMTase activity. These results are consistent with the observed lack of RNA binding specificity for human PRC2 (ref. 13) and the proposed inhibition of PRC2 by RNA[14,30], and they raise the question of whether the two strands interact differently with PRC2 *in vitro* and *in vivo*.

E(Z) binds to the reverse strand of the *vg* PRE/TRE *in vivo*

Experiments in which the interaction of different RNAs with PRC2 components was assayed by incubation with nuclear extracts have shown a high specificity for one strand in preference to its antisense counterpart for several noncoding RNAs[7,8,12]. To evaluate whether the two strands of the *vg* PRE/TRE bind differentially to PRC2 components in living cells, we performed RNA immunoprecipitation from *Drosophila* S2 cells. Cells were cotransfected with plasmid expressing GFP or GFP-E(Z) and a plasmid encoding either the forward or reverse PRE/TRE strand (**Fig. 6a**,**b**). Surprisingly, this analysis identified a highly strand-specific interaction between E(Z) and the reverse RNA strand of the PRE/TRE (**Fig. 6c**–**f**). Thus, although PRC2 does not interact specifically with different RNAs *in vitro* (**Fig. 5** and **Supplementary Fig. 9**)[13], the PRC2 component E(Z) binds with high specificity to only one strand of the *vg* PRE/TRE *in vivo* (**Fig. 6e**,**f**). These results indicate that other factors present in the cell, rather than an intrinsic property of PRC2 itself, mediate the specificity of binding by facilitating or preventing interactions between PRC2 and specific RNAs.

The reverse strand inhibits and evicts E(Z) from chromatin

The fact that both strands can inhibit E(Z) activity *in vitro* but only the reverse strand can bind *in vivo* suggests that the reverse strand might activate the PRE/TRE *in vivo* by inhibiting local E(Z) HMTase activity, whereas the forward strand is unable to bind and has no inhibitory effect. To test this idea, we performed ChIP on embryos overexpressing either the forward or reverse strand of the *vg* PRE/TRE from transgenes integrated at site 1, under the control of the ubiquitous *daughterless* GAL4 (*da*GAL4) driver, which drives GAL4 expression from the *daughterless* promoter. We examined H3K27me3 and E(Z) levels at both the transgenic and endogenous *vg* loci (**Fig. 6g**). The transgene contains sequences that are also found elsewhere in the genome (the 1.6-kb PRE/TRE itself and the markers *white* and *yellow*). Any effects occurring within the 1.6-kb PRE/TRE would not be discernible because the transgenic and endogenous PRE/TRE sequences are identical, but the sequences unique to each locus would be informative. The sequence that is unique to the transgene is the downstream sequence immediately flanking the PRE/TRE, which is bacterially derived (**Fig. 6g**, left, unique region shown on a white background, 1.6-kb PRE/TRE shown on a gray background). The sequences that are unique to the endogenous *vg* locus are those flanking the 1.6-kb PRE/TRE (**Fig. 6g**, middle).

For the transgenic locus, embryos overexpressing the forward strand showed similar enrichment profiles across the transgene-flanking region for both E(Z) binding and H3K27me3 (**Fig. 6g**, top). In contrast, embryos overexpressing the reverse strand showed two striking effects in this flanking region (**Fig. 6g**, bottom). Distal to the PRE/TRE, both the E(Z) and H3K27me3 levels were lower in comparison to those in the line overexpressing the forward strand (**Fig. 6g**, arrowhead). This finding suggests that the reverse strand might titrate E(Z) away from the locus. In addition, the flanking sequence directly adjacent to the PRE/TRE displayed high enrichment for E(Z) binding but low H3K27me3 levels (**Fig. 6g**, asterisk). This result is consistent with the inhibition of E(Z) activity by the reverse strand (however, it might also reflect the removal of nucleosomes from this site). There was no detectable difference between the overexpression of the forward and reverse strands at the endogenous *vg* locus (**Fig. 6g**, middle and **Supplementary Fig. 10**). This finding is consistent with the lack of upregulation of *vg* mRNA observed in early embryos expressing the reverse strand (**Fig. 4d**,**g**). We also observed no detectable difference at several other targets of E(Z) regulation (**Fig. 6g**, right and data not

Figure 8 Proposed model for switching PRE/TRE properties upon strand switching. Black box, PRE/TRE; red, forward-strand transcript; green, reverse-strand transcript; yellow circle, protein binding the forward-strand transcript, proposed to prevent interaction with E(Z) and to facilitate PRE/TRE pairing. See the main text for details.

shown). In summary, this analysis suggests that the forward strand does not inhibit the enzymatic activity of E(Z) *in vivo*, whereas the reverse strand might act both by inhibiting activity and by removing E(Z) from the locus from which it is transcribed.

GEAR box elements are potentially widespread

To determine whether further GEAR box elements might exist in the fly and mouse genomes, we mined genome-wide data sets[20,21,31–34] (**Fig. 7** and **Supplementary Tables 2** and **3**). This analysis identified several hundred PcG-binding sites in the fly and mouse genomes that showed transcription from both strands (either convergent or divergent) passing through the PcG-binding site (**Fig. 7a**,**b**; see **Supplementary Tables 2** and **3** for full listings). Furthermore, although bidirectional and antisense transcription is a common feature of many genomes[35], comparison with binding sites for other transcription factors (**Fig. 7a**) or non-PcG-bound promoters (**Fig. 7b**) showed a highly significant enrichment of bidirectional transcription at PcG-bound sites in most cases, indicating that this feature is favored at PcG-regulated genes (*P* values are given in **Supplementary Table 3c**).

We note that the occurrence of bidirectional transcription is in general less common in fly than in mouse. This difference might be owing to the fact that fly transcript data were available only from embryos or cell lines, whereas the mouse cap analysis gene expression (CAGE) data were generated from multiple different tissues. For the mouse data set, we were able to evaluate tissue and strand specificity, finding that over 99% of the PcG sites that transcribe opposite strands do so in different tissues. A notable example is the mouse *Vgll2* gene, a functional homolog of the fly *vg* gene[36] (**Fig. 7c**). The PcG-binding sites at this locus transcribe the forward strand in muscle tissue, where the gene is active, and the reverse strand in embryonic tissues, where the gene is repressed. In summary, this analysis identifies many further sites of PcG regulation in fly and vertebrate genomes that might switch their function between activation and silencing by switching the strand of their noncoding transcription.

DISCUSSION

We demonstrate both by analysis of endogenous transcripts and by using ectopic overexpression strategies *in vivo* that transcripts from opposite strands of the *vg* PRE/TRE have opposite effects on PRE/TRE status. However, *in vitro*, both noncoding RNAs have equivalent inhibitory effects on the HMTase activity of PRC2. Taking the *in vitro* and *in vivo* data together, we propose that the specificity of the noncoding RNA interaction with PcG proteins *in vivo* is not a result of inherently different affinities of PRC2 for different noncoding RNAs but of the availability of a given noncoding RNA (regulated by interactions of that RNA with other molecules) to interact with PRC2 and inhibit its enzymatic activity. We propose that the forward-strand noncoding RNA promotes silencing by facilitating pairing between PRE/TREs. PRE/TRE pairing has been shown to be essential for maximum silencing by the *vg* PRE/TRE[26], and this silencing is genetically dependent on PcG[23,26]. Thus, we propose that forward strand–induced pairing might facilitate or stabilize PcG-mediated pairing-dependent silencing (**Fig. 8**, left). E(Z) is detected at the *vg* PRE/TRE in ChIP analyses[25] (**Supplementary Figs. 2b** and **10**) but does not interact with the forward strand (**Fig. 6e**,**f**). Thus, we propose that E(Z) binds at the silenced PRE/TRE independently of RNA. The forward-strand noncoding RNA might facilitate or stabilize pairing by binding to additional bridging proteins (**Fig. 8**, yellow). These or other proteins might also prevent binding and inhibition of E(Z) by the RNA.

Upon switching to the active state (**Fig. 8**, right), transcription of the reverse PRE/TRE strand would be incompatible with forward-strand transcription because the reverse transcript runs through the forward-strand promoter (**Figs. 1b** and **2a**). Reverse-strand transcription might thus destabilize pairing, enabling the activation of the PRE/TRE (**Fig. 4c**,**f**). In addition, the reverse strand binds E(Z) (**Fig. 6e**,**f**) and, upon binding, would inhibit E(Z) HMTase activity (**Figs. 5a** and **6g**), and it might also remove E(Z) from the locus (**Fig. 6g**). In this way, multiple self-reinforcing events could contribute to the stable switching of the PRE/TRE into an active state. The *vg* PRE/TRE responds to TrxG mutations by loss of activation[23]. Whether TrxG-dependent activation acts via the noncoding RNAs will be a key question for future studies.

Genome-wide analysis identifies many sites that might share functional features with the *vg* PRE/TRE. We propose that these elements can exist in a 'neutral' state, in which neither they nor their associated genes are transcribed (this is consistent with our observation that the *vg* PRE/TRE is not transcribed in tissues or at embryonic stages that do not express *vg* mRNA; **Supplementary Fig. 1**). However, in tissues that have the potential to transcribe the gene, the element might be switched either to the forward or reverse mode, thereby boosting either silencing or activation. This might serve to sharpen spatial expression boundaries, to stabilize gene expression states or to accelerate the kinetics of activation or repression[37].

In conclusion, this work provides a new paradigm linking forward and reverse noncoding transcription to dynamic and developmentally regulated switching of PRE/TRE properties and thus to the maintenance of cell identities during development. Furthermore, the demonstration that any RNA is a potent inhibitor of PRC2 enzymatic activity *in vitro* but that only specific RNAs are able to bind and inhibit PRC2 *in vivo* strongly implies that specific RNAs are masked *in vivo* from interacting with PRC2. This provides an enormous potential for the regulated and reversible RNA-mediated inhibition of local PRC2 activity.

URLs. Vienna Campus Support Facility (CSF), http://www.csf.ac.at/; Austrian Academy of Sciences, http://www.oeaw.ac.at/; European Union Framework Programme 6 Network of Excellence 'The Epigenome', http://www.epigenome-noe.net/; European Union Framework Programme 7 Network of Excellence 'Epigenesys', http://www.epigenesys.eu/; FWF Austrian Science Fund, http://www.fwf.ac.at/en/; Definiens XD software, http://www.definiens.com/product-services/definiens-xd-product-suite.html; Vienna *Drosophila* Resource Center (VDRC), http://stockcenter.vdrc.at/; modENCODE, http://www.modencode.org/; FANTOM3 CAGE summary data set, http://fantom31p.gsc.riken.jp/cage_analysis/export/mm5/; liftOver tool, https://genome.ucsc.edu/cgi-bin/hgLiftOver; R/Bioconductor, http://www.R-project.org/; dm3 and mm9 RefSeq gene annotation, http://genome.ucsc.edu/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. ChIP-seq data have been deposited in the Gene Expression Omnibus (GEO) and are available through accession GSE59192.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank S. Gasser for discussions and for critical reading of the manuscript. We thank M. Rehmsmeier and members of our laboratories for discussions, P. Pasierbek for advice and training on imaging, P.A. Steffen (IMBA) for the *GFP* and *E(z)*::*GFP* constructs, C. Ehrhardt and E. Dworschak for technical assistance, B. Dickson (Janelia Farm) for the *en*GAL4 driver line, J.M. Dura (Institute of Human Genetics, Montpelier, France) for the *da*GAL4 driver line, I. Tamir (CSF Vienna) for the bioinformatics analysis of ChIP-seq data and sharing the 'Fuge' algorithm, F. Bantignies (Institute of Human Genetics, Montpelier, France) for advice on three-dimensional DNA FISH, R. Jones (Dedman College, Southern Methodist University) for providing antibody to E(Z) and the Vienna Campus Support Facility (CSF) for library preparation, deep sequencing and the purification of *Drosophila* PRC2. This work was funded by the Austrian Academy of Sciences, by European Community grants European Union Framework Programme 6 Network of Excellence 'The Epigenome' (to L.R.) and the European Union Framework Programme 7 Network of Excellence 'Epigenesys' (to L.R.), and by an FWF Austrian Science Fund grant (P21525-B20 to L.R.).

AUTHOR CONTRIBUTIONS

A.L. and L.R. initiated the project. L.R., V.A.H. and A.L. designed the experiments. J.T. performed bioinformatics analysis of genome-wide data sets. G.S. and K.A. performed automated image analysis of three-dimensional DNA FISH. V.A.H., A.L., H.O., C.A., F.R., B.B., A.D., M.R., H.B.S. and L.R. conducted the experiments and analyzed the data. S.K. supervised B.B. A.P. supervised M.R. M.L.V. and J.A.S. provided purified *Drosophila* PRC2 and polynucleosome substrates. L.R. prepared the manuscript with input from V.A.H., J.T., A.L., B.B. and H.O.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Di Croce, L. & Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat. Struct. Mol. Biol.* **20**, 1147–1155 (2013).
2. Steffen, P.A. & Ringrose, L. What are memories made of? How Polycomb and Trithorax proteins mediate epigenetic memory. *Nat. Rev. Mol. Cell Biol.* **15**, 340–356 (2014).
3. Müller, J. & Kassis, J.A. Polycomb response elements and targeting of Polycomb group proteins in *Drosophila*. *Curr. Opin. Genet. Dev.* **16**, 476–484 (2006).
4. Hekimoglu, B. & Ringrose, L. Non-coding RNAs in polycomb/trithorax regulation. *RNA Biol.* **6**, 129–137 (2009).
5. Spitale, R.C., Tsai, M.C. & Chang, H.Y. RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. *Epigenetics* **6**, 539–543 (2011).
6. Brockdorff, N. Noncoding RNA and Polycomb recruitment. *RNA* **19**, 429–442 (2013).
7. Rinn, J.L. *et al.* Functional demarcation of active and silent chromatin domains in human *HOX* loci by noncoding RNAs. *Cell* **129**, 1311–1323 (2007).
8. Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J. & Lee, J.T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756 (2008).
9. Kanhere, A. *et al.* Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol. Cell* **38**, 675–688 (2010).
10. Zhao, J. *et al.* Genome-wide identification of polycomb-associated RNAs by RIP-seq. *Mol. Cell* **40**, 939–953 (2010).
11. Wang, K.C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011).
12. Guil, S. *et al.* Intronic RNAs mediate EZH2 regulation of epigenetic targets. *Nat. Struct. Mol. Biol.* **19**, 664–670 (2012).
13. Davidovich, C., Zheng, L., Goodrich, K.J. & Cech, T.R. Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* **20**, 1250–1257 (2013).
14. Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K. & Lee, J.T. Regulatory interactions between RNA and Polycomb repressive complex 2. *Mol. Cell* doi:10.1016/j.molcel.2014.05.009 (29 May 2014).
15. Cohen, B., Simcox, A.A. & Cohen, S.M. Allocation of the thoracic imaginal primordia in the *Drosophila* embryo. *Development* **117**, 597–608 (1993).
16. Williams, J.A., Bell, J.B. & Carroll, S.B. Control of *Drosophila* wing and haltere development by the nuclear *vestigial* gene product. *Genes Dev.* **5**, 2481–2495 (1991).
17. Deng, H., Hughes, S.C., Bell, J.B. & Simmonds, A.J. Alternative requirements for Vestigial, Scalloped, and Dmef2 during muscle differentiation in *Drosophila melanogaster*. *Mol. Biol. Cell* **20**, 256–269 (2009).
18. Bernard, F. *et al.* Control of *apterous* by *vestigial* drives indirect flight muscle development in *Drosophila*. *Dev. Biol.* **260**, 391–403 (2003).
19. Guss, K.A., Mistry, H. & Skeath, J.B. Vestigial expression in the *Drosophila* embryonic central nervous system. *Dev. Dyn.* **237**, 2483–2489 (2008).
20. Enderle, D. *et al.* Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. *Genome Res.* **21**, 216–226 (2011).
21. Kharchenko, P.V. *et al.* Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**, 480–485 (2011).
22. Lee, N., Maurange, C., Ringrose, L. & Paro, R. Suppression of Polycomb group proteins by JNK signalling induces transdetermination in *Drosophila* imaginal discs. *Nature* **438**, 234–237 (2005).
23. Pérez, L. *et al.* Enhancer-PRE communication contributes to the expansion of gene expression domains in proliferating primordia. *Development* **138**, 3125–3134 (2011).
24. Oktaba, K. *et al.* Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in *Drosophila*. *Dev. Cell* **15**, 877–889 (2008).
25. Schwartz, Y.B. *et al.* Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat. Genet.* **38**, 700–705 (2006).
26. Okulski, H., Druck, B., Bhalerao, S. & Ringrose, L. Quantitative analysis of polycomb response elements (PREs) at identical genomic locations distinguishes contributions of PRE sequence and genomic environment. *Epigenetics Chromatin* **4**, 4 (2011).
27. Bate, M. The embryonic development of larval muscles in *Drosophila*. *Development* **110**, 791–804 (1990).
28. DeVido, S.K., Kwon, D., Brown, J.L. & Kassis, J.A. The role of Polycomb-group response elements in regulation of *engrailed* transcription in *Drosophila*. *Development* **135**, 669–676 (2008).
29. Bantignies, F. *et al.* Polycomb-dependent regulatory contacts between distant *Hox* loci in *Drosophila*. *Cell* **144**, 214–226 (2011).
30. Kaneko, S., Son, J., Shen, S.S., Reinberg, D. & Bonasio, R. PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat. Struct. Mol. Biol.* **20**, 1258–1264 (2013).
31. Hoskins, R.A. *et al.* Genome-wide analysis of promoter architecture in *Drosophila melanogaster*. *Genome Res.* **21**, 182–192 (2011).
32. Katayama, S., Kanamori, M. & Hayashizaki, Y. Integrated analysis of the genome and the transcriptome by FANTOM. *Brief. Bioinform.* **5**, 249–258 (2004).
33. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).
34. Peng, J.C. *et al.* Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* **139**, 1290–1302 (2009).
35. Ietswaart, R., Wu, Z. & Dean, C. Flowering time control: another window to the connection between antisense RNA and chromatin. *Trends Genet.* **28**, 445–453 (2012).
36. Chen, H.H., Maeda, T., Mullett, S.J. & Stewart, A.F. Transcription cofactor Vgl-2 is required for skeletal muscle differentiation. *Genesis* **39**, 273–279 (2004).
37. Pelechano, V. & Steinmetz, L.M. Gene regulation by antisense transcription. *Nat. Rev. Genet.* **14**, 880–893 (2013).
38. Williams, J.A., Paddock, S.W., Vorwerk, K. & Carroll, S.B. Organization of wing formation and induction of a wing-patterning gene at the dorsal/ventral compartment boundary. *Nature* **368**, 299–305 (1994).
39. Kim, J. *et al.* Integration of positional signals and regulation of wing formation and identity by *Drosophila vestigial* gene. *Nature* **382**, 133–138 (1996).

ONLINE METHODS

PRE/TRE transcript mapping. Mapping of the 5′ ends of PRE/TRE transcripts (**Fig. 1b** and **Supplementary Fig. 2**) was performed using a 5′ RACE kit (version 2.0, Invitrogen) according to the manufacturer's instructions. Each transcript was mapped from at least two independent cDNA clones generated from at least two different gene-specific primers (GSP1). The sequences for the primers used for 5′ RACE analysis are given in **Supplementary Table 1**. For 3′ end mapping, the A-rich nature of the PRE/TRE transcripts precluded the use of 3′ RACE with oligo(dT) primers. Instead, the 3′ ends were mapped by primer walking. Strand-specific cDNA synthesis was performed using a single 5′ primer in combination with different 3′ primers, designed to anneal consecutively toward the 3′ end of each transcript at intervals of 50 to 100 bp (**Supplementary Fig. 3** and **Supplementary Table 1**). The 3′ end was designated as the position of the last primer that gave a product.

Double RNA *in situ* hybridization. PCR products from embryonic cDNA for the *vg* gene or the *vg* PRE/TRE noncoding RNA were cloned in both orientations into PCR-XL-TOPO (Invitrogen). For *in situ* hybridization, RNA probes were *in vitro* transcribed from each cDNA strand. One probe was labeled with digoxigenin (Boehringer), detected with alkaline phosphatase–conjugated sheep antibody to digoxigenin (Roche, 11093274910; 1:1,500 dilution) and visualized with FastRed (Sigma). The other probe was labeled with fluorescein, detected with primary antibody (mouse antibody to fluorescein, Roche, 11426320; 1:500 dilution) and secondary antibody (HRP-conjugated goat antibody to mouse, Invitrogen), and visualized with Alexa Fluor 488 Tyramide (Invitrogen) using the Tyramide Signal Amplification kit (TSA, Invitrogen) according to the manufacturer's instructions. Label swaps were performed to ensure specificity. In the images shown, *vg* mRNA was detected with FastRed and PRE/TRE transcripts were detected with TSA. Images were acquired using confocal microscopy, with an LSM 700 Axioimager (larval tissues) or an LSM 510 Axiovert 200M (embryos) (Zeiss).

Three-dimensional quantification of RNA *in situ* hybridization signals. Three-dimensional quantification (**Fig. 3o**,**p**) was performed using IMARIS x64 7.4.0 (14 December 2011), Bitplane Scientific Software. *z*-stack images were uploaded to the software. Surfaces were generated by defining contours around areas expressing the *en*GAL4 driver or areas not expressing the driver. The contours were drawn manually in each slice of the *z* stack. On the basis of these contours, the software calculated the volume of the surface in the *z* stack and quantified the intensity sum of the *vg* mRNA signal within this volume. For the animation shown in **Supplementary Video 1**, the *z*-stack image was again uploaded to the software and key frames were added manually (for example, rotation and the removal of surfaces and channels).

Transgenic flies. The transgenic *vg* PRE/TRE lines and control lines are described in ref. 26 or are derived from those constructs as described below. All constructs contained a UAS, an *hsp70* minimal promoter, a PRE/TRE (or no PRE/TRE in controls) and an *mw* reporter. The landing site was marked with *yellow*. Landing site 1 in this study corresponded to landing site 2 in ref. 26 (cytological location on chr. 2R, 46E1, genomic position 5,965,083). Landing site 2 in this study corresponded to landing site 3 in ref. 26 (cytological location on chr. 2R, 58F4, genomic position 18,549,410). pKC27*vg*.fwd (**Figs. 2** and **6a–f**, and **Supplementary Fig. 3**) carried 1.6 kb of the *vg* PRE/TRE cloned into pKC27*mw* downstream of the UAS cassette and *hsp70* promoter (described in detail in ref. 26). pKC27*vg*.rev (**Fig. 6a–f**) was generated by introducing XbaI sites upstream and downstream of the *vg* PRE/TRE in the pKC27*vg*.fwd plasmid by site-directed mutagenesis (QuikChange II Site-Directed Mutagenesis kit, Stratagene) and subsequent excision and re-ligation of the *vg* PRE, such that the PRE/TRE was in the opposite orientation to that in pKC27*vg*.fwd (direction was validated by sequencing the XbaI junction sites). In pKC27*vg*.fwd′ (**Figs. 3b**,**k–n** and **6g**, and **Supplementary Figs. 7** and **10**) and pKC27*vg*.rev′ (**Figs. 3b**,**g–j** and **6g**, and **Supplementary Fig. 10**), the UAS-PRE/TRE cassette was inverted with respect to the *mw* reporter such that induced transcription drives away from *mw*. These constructs were assembled by single-step *in vitro* recombination[40] and were validated by Sanger sequencing. Primer sequences and constructs are available upon request.

RNA extraction, cDNA synthesis and quantitative PCR analysis. Total RNA was extracted from dissected larval tissues and dechorionated embryos using the High Pure RNA isolation kit (Roche). Embryonic RNA was additionally treated with DNase (Turbo DNase, Ambion). cDNA was synthesized using the SuperScript II kit (Invitrogen) according to the manufacturer's instructions. For random priming (**Fig. 6c**,**e** and **Supplementary Fig. 1**), priming was performed with equal amounts of oligo(dT) and random decamer primers (Ambion). For strand-specific cDNA synthesis (all other figures), cDNA was primed with a cocktail containing 0.2 pmol of each of the appropriate primers for the detection of the *Tbp* forward strand (for normalization) and other transcripts of interest (see **Supplementary Table 1** for primer sequences). For the transactivation experiment (**Fig. 4**), primers for the *Timp* transcript were included as a control for embryo age. For RNA immunoprecipitation (**Fig. 6e**,**f**), primers to *GFP* were included to control for transfection efficiency. For all strand-specific cDNAs, appropriate controls for DNA contamination and mispriming were included. Annealing and cDNA synthesis were performed at 50 °C, and, after heat inactivation of the reverse transcriptase, RNA was removed by treatment with RNase H (NEB) before PCR analysis. qPCR analysis was performed using SYBR Green JumpStart Taq ReadyMix (Sigma) in a Realplex MasterCycler (Eppendorf).

Three-dimensional DNA FISH. Probe labeling was carried out using the FISH Tag DNA Orange kit and FISH Tag DNA Green kit (Invitrogen) following the protocol without modifications. Whole-fly genomic DNA was used as the PCR template. See **Supplementary Table 1** for primer sequences. Homozygous transgenic fly lines were crossed to homozygous *en*GAL4 flies as described above, and 3rd instar larval wing discs from the progeny (heterozygous for both the transgene and the GAL4 driver) were dissected in PBS and transferred to an Eppendorf tube for immediate fixation in 4% paraformaldehyde in PBT (PBS containing 0.1% Tween-20) for 5 min on a rotating wheel. Prehybridization was carried out as described[41], except that the rehydration steps were omitted and we proceeded directly to RNase A treatment. Hybridization and posthybridization washes were performed as described[41]. Samples were subsequently counterstained with DAPI (1:1,000 dilution in PBT) for 10 min on a rotating wheel. Samples were washed once with PBT, rinsed once in PBS and mounted in 25 µl of ProLong Antifade kit (Invitrogen). Three-dimensional stacks were imaged by confocal microscopy. Deconvolved images were analyzed by automated image analysis implemented in Definiens XD software.

For each disc, at least three non-overlapping images were acquired within the posterior *en*GAL4 expression domain and at least three images were acquired within the opposite anterior domain not expressing *en*GAL4. The positions of the domains were defined from morphological features with reference to *in situ* images such as those shown in **Figure 3g**,**l**. For each disc, data from the three anterior images and the three posterior images were combined, giving a single data set for each half of each disc (these data are shown in **Supplementary Fig. 7**). Within single discs, tissue and cell growth occurs uniformly[42,43]; thus, we reason that the comparison of distances in the anterior and posterior compartments is informative. However, we observed that different discs had different sizes and might thus have different cell sizes, which in turn might affect nuclear size and the average distances between loci for different discs. To account for such differences, data for a given genotype were normalized using the mean of distances observed in the anterior compartment. Thus, for disc 1 in each panel of **Supplementary Figure 7**, the distances shown are the actual observed distances, whereas those for discs 2 and 3 were multiplied by a factor representing the difference in mean distance between the anterior compartment of each disc to that of disc 1. The correction factors for discs 2 and 3, respectively, are as follows: **Supplementary Figure 7a** (control): 1.04 and 1.17; **Supplementary Figure 7b** (reverse-strand overexpression): 0.93 and 0.76; **Supplementary Figure 7c** (forward-strand overexpression) 0.94 and 0.92.

For the plots in **Figure 3b**, the data were further normalized between genotypes. Thus, for the pKC27*mw* control line in **Figure 3b**, the distances shown are the actual observed distances, whereas those of the other two transgenic lines were multiplied by a factor representing the difference in mean distance between the anterior compartment of the line to that of the control. The data for pKC27*vg*.rev′ were multiplied by 1.38; those for pKC27*vg*.fwd′ were multiplied by 1.29.

doi:10.1038/ng.3058

Fly crosses. The *w*; *P*{*en-GAL4.M616*} (chr. II) driver line was provided by B. Dickson (Janelia Farm). Homozygous *en*GAL4 flies were crossed to homozygous transgenic lines carrying a UAS cassette linked to the *vg* PRE/TRE or to control lines with a UAS cassette without the PRE/TRE. We dissected the 3rd instar larval wing discs of the progeny and subjected them to double RNA *in situ* hybridization or three-dimensional DNA FISH, as described above. For the RNAi control shown in **Supplementary Figure 8**, VDRC line 16896 expressing an RNAi hairpin against *vg* mRNA was used instead of the UAS *vg* PRE/TRE lines described above[44]. For ChIP (**Fig. 6g**), a homozygous *da*GAL4 (chr. III) driver line provided by J.M. Dura (Institute of Human Genetics, Montpelier, France) was crossed to appropriate homozygous transgenic lines.

Radioactive histone methyltransferase assays (*Drosophila* PRC2). *Drosophila* PRC2 showed minimal activity on peptide substrates and recombinant mononucleosomes, precluding its use in the DELFIA assay, and was therefore assayed on HeLa polynucleosomes using ^{3}H-labeled S-adenosylmethionine. *Drosophila* PRC2 was combined at a final concentration of 20 or 50 nM with a final concentration of ~1.4 µM ^{3}H-labeled S-adenosylmethionine and 1 µg (0.4 nM) of HeLa polynucleosomes (Reaction Biology, HMT-35-160). Reactions were performed in assay buffer (12 mM HEPES, pH 7.9, 0.24 mM EDTA, 12% glycerol, 4 mM DTT, 2.5 mM $MgCl_2$ and 30 mM KCl) in a final volume of 20 µl. Inhibition experiments were performed with final concentrations of RNA ranging from 1 nM to 70 nM. PRE/TRE RNAs were generated by *in vitro* transcription of a 1.6-kb RNA from the PRE/TRE (**Fig. 2a**), in forward or reverse orientation. Control RNAs were generated by *in vitro* transcription of a 1.6-kb fragment amplified from the pCR-XL-TOPO plasmid (Invitrogen) using the primers ctr1.fwd and ctr1.rev (control 1) or from the pWalium20 plasmid[45] using the primers ctr2.fwd and ctr2.rev (control 2). Samples were incubated for 60 min at 30 °C, and reactions were stopped by the addition of SDS loading buffer before samples were loaded on a 4–12% NuPAGE Bis-Tris gel (Invitrogen). Protein blotting, amido black staining (Sigma) and autoradiography were performed as described[46]. Quantification of amino black signal and autoradiography films was performed using ImageJ software (US National Institutes of Health).

DELFIA histone methyltransferase assay (human PRC2). To detect human PRC2 activity *in vitro*, 2.5 µl of human PRC2 (Reaction Biology, HMT-25-114; final concentration of 155 nM) was combined with final concentrations of 16 µM S-adenosylmethionine and 500 nM biotinylated H3 peptide, spanning residues 21–40 of histone H3 (ATKAARKSAPATGGVKKPHR-NTPEG-biotin; New England Peptide). Reactions were performed in assay buffer (50 mM Tris-HCl, pH 9.0, 50 mM NaCl, 1 mM DTT) in a final volume of 25 µl. Mono-, di- or trimethylated H3K27 peptides served as controls (New England Peptide). Inhibition experiments were performed with a total of 5 µg of RNA (380 nM final concentration) or 10 µM GSK126 (Chemietek) added before the enzyme. Samples were incubated for 60 min at 30 °C and transferred onto a streptavidin-coated 384-well plate (PerkinElmer, CC11-H10). Plates were incubated for 60 min at room temperature to allow biotinylated peptides to bind and were washed three times with wash buffer II (50 mM Tris, pH 7.5, 150 mM NaCl, 0.05% Tween-20). Primary mouse antibody to H3K27me3 (Active Motif, 61017; 1:1,000 dilution) and Europium-labeled secondary antibody (PerkinElmer, AD0207; 200 ng/ml final concentration; EU-N1-anti-mouse-IgG) were diluted in FI buffer (50 mM Tris, pH7.5, 150 mM NaCl, 0.05% Tween-40, 25 mM DTPA, 0.2% BSA, 0.05% bovine γ-globulins) and added to the wells. Plates were incubated for 60 min and were washed three times with wash buffer II. After the addition of Enhancement Solution (PerkinElmer), plates were incubated for 45 min, and fluorescence was measured on an EnVision plate reader with excitation at 340 nm and emission at 615 nm (PerkinElmer).

RNA immunoprecipitation. The plasmids used for transfection were as follows: pKC27.Tub.GFP and pKC27.Tub.E(Z)::GFP (described previously[47]). Transgenic flies carrying the construct encoding Tub.*E(z)*::*GFP* were generated. These transgenic flies give full genetic rescue of lethal *E(z)* mutations[47]. pKC27*vg*.fwd and pKC27*vg*.rev (**Fig. 6a**) are described under "Transgenic flies" above.

For strand expression controls, 3×10^5 S2 cells per well were seeded in a 24-well plate. S2 cells were obtained from J. Knoblich (IMBA). After 24 h, 250 ng of pKC27.Tub.EGFP-nls, 300 ng of either pKC27*vg*.fwd or pKC27*vg*.rev, and 450 ng of carrier DNA were transfected into cells using FuGENE HD transfection reagent (Promega), with a FuGENE:DNA ratio of 1:5. After 24 h, cells were directly lysed in TRIzol (Invitrogen), and RNA extraction was performed according to the manufacturer's instructions (Invitrogen). Isolated RNA was treated with Turbo DNase (Invitrogen) according to the manufacturer's protocol. Strand-specific cDNA synthesis and qPCR analysis were performed using primers for *Tbp* and the PRE/TRE core as described above (see **Supplementary Table 1** for primer sequences). For quantification of the input RNA (**Fig. 6c**), strand-specific cDNA synthesis and RT-qPCR were performed, and RNA levels were quantified as the percentage of *Tbp* levels, after having controlled to exclude DNA contamination. The forward strand was typically transcribed at 3- to 20-fold higher levels than the reverse strand (data not shown). Data were normalized to the most highly expressed strand, typically present at between 200 and 1,000% of *Tbp* levels.

For RNA immunoprecipitation, for each transfection, we seeded 3.5×10^7 cells per dish in three 15-cm dishes. After 24 h, cells were cotransfected with 25.5 µg of either pKC27.Tub.GFP or pKC27.Tub.E(Z)::GFP and 30.6 µg of either pKC27*vg*.fwd or pKC27*vg*.rev, using FuGENE HD as described above. Cells were collected and lysed as described[48], and immunoprecipitation of GFP or E(Z)-GFP was performed using GFP Trap agarose beads (Chromotek). Samples were processed according to the manufacturer's protocol for the GFP Trap with the following modifications: beads were equilibrated using NT2 buffer[48] and washing steps were performed as described[48].

For protein blotting, input, unbound and immunoprecipitate fractions were run on NuPAGE Mini Gels (Invitrogen) and blotted on Immobilon-P PVDF membranes (Millipore) according to standard protocols. Primary antibodies included antibody to α-tubulin (Sigma-Aldrich, T5168; 1:20,000 dilution) and antibody to GFP (Clontech, JL-8; 1:2,000 dilution). Signal was detected on film with HRP-coupled secondary antibodies and ECL (GE Healthcare). For RNA analysis, RNA extraction from beads was performed using TRIzol. Randomly primed (**Fig. 6c**,**e**) or strand-specific (**Fig. 6f**) cDNA synthesis and qPCR were performed as described above.

Levels of immunoprecipitated RNA (**Fig. 6e**) were quantified by randomly primed cDNA synthesis and RT-qPCR. Data were normalized to input and *GFP* mRNA levels and were controlled to exclude DNA contamination. Data are shown normalized to RNA immunoprecipitate levels in the immunoprecipitation for GFP. Despite the fact that the forward strand was typically transcribed at 3- to 20-fold higher levels than the reverse strand after transfection with equivalent amounts of each plasmid (see "Quantification of input RNA" above), RNA immunoprecipitation with E(Z)-GFP consistently resulted in the pulldown of RNA from the reverse- but not the forward-strand plasmid that was transfected into cells. Quantification of the forward and reverse strands after immunoprecipitation of E(Z)-GFP upon transfection with the reverse-strand plasmid (**Fig. 6f**) was performed by strand-specific cDNA synthesis and RT-qPCR. Data were controlled to exclude DNA contamination and were normalized to the percentage of *Tbp* levels in input.

Chromatin immunoprecipitation. Chromatin fixation and immunoprecipitation were performed essentially as described[49]. Embryos (0.2 g) were dechorionated with 1.4% sodium hypochloride and cross-linked with 1.8% formaldehyde for 15 min at room temperature. Cross-linked embryos were sonicated to produce chromatin fragments with an average size of 200–500 bp. Soluble chromatin was separated from insoluble material by centrifugation. The supernatant containing chromatin corresponding to 50–60 µg of DNA was used for immunoprecipitation. Antibodies used were antibody to PC (2 µg per immunoprecipitation; d-220, sc-25762, Santa Cruz Biotechnology), antibody to E(Z) (10 µl per immunoprecipitation; from R. Jones (Dedman College, Southern Methodist University); ref. 50) and antibody to H3K27me3 (4 µg per immunoprecipitation; CS200603, Millipore).

Preparation of ChIP-seq libraries. Sequencing libraries were prepared with the KAPA Library Preparation kit Illumina series (KK8201) according to the manufacturer's instructions. After adaptor ligation, library fragments of 200–800 bp were isolated using Ampure XP Beads (Beckman Coulter, A63881). DNA was PCR amplified with Illumina primers for 15 cycles, purified and loaded onto an Illumina flow cell for cluster generation. Libraries were

sequenced on the Illumina HiSeq 2000. Reads were aligned to the *Drosophila* genome (BDGP R5/dm3) using Bowtie (0.12.5)[51].

Bioinformatics analysis of bidirectional transcription at PcG targets. For *Drosophila* analysis, ChIP data from the modENCODE Project[21] and from ref. 20 were both compared to CAGE and RACE data from ref. 31 and to Massive Analysis of cDNA Ends (MACE) data from ref. 20. ChIP data sets are specified in **Supplementary Table 2**. modENCODE data for the PcG proteins E(Z), PHO, PC and PSC and for the transcription factors Hairy and Caudal from several developmental stages were downloaded from the modENCODE website. Peaks from the same protein (Caudal or Hairy) or group of proteins (PcG) were merged using MultOvl[52]. For overlapping peaks, the region of overlap was merged to a single peak. Data files from ref. 20 were downloaded from the GEO database (accession GSE24521).

For mouse analysis, ChIP data from ref. 33 (Sox2) and ref. 34 (Jarid1, Jarid2 and Suz12) were each compared to CAGE tag data from the FANTOM3 CAGE database[32]. Mouse embryonic stem cell (ESC) PcG binding data from ref. 34 were downloaded from the GEO database (accession GSE18776). The peaks for Jarid2 and Suz12 as defined by the authors were used for further analyses. Mouse ESC Sox2 binding data were downloaded from the GEO database (accession GSE11431), and the same peak finding method and parameters were applied to the data as in ref. 34, using the QuEST peak finder[53]. The FANTOM3 CAGE summary data set was downloaded from the FANTOM3 website. The original mm5 coordinates were transformed into mm9 coordinates using the liftOver tool from the UCSC Genome Browser.

For promoter analyses in both mouse and fly, promoters were defined as stretching from −2 kb to +500 bp with respect to the annotated transcription start site. Overlapping promoters were merged and defined as being bound by PcG if one of the above-listed PcG ChIP peaks overlapped (by at least one base) with the promoter.

To compare CAGE, RACE or MACE tags with ChIP peaks or promoters, the following classes of transcript tags were defined: 'mono', transcription peaks occur in only one direction (one or more MACE or CAGE peaks allowed); 'divergent' transcription peaks occur in both directions, where the first start coordinate for plus-strand transcription is greater than the first start coordinate for the minus strand; and 'convergent' transcription peaks occur in both directions, where the first start coordinate on the plus strand is less than the first start coordinate of the minus strand. Custom R/Bioconductor scripts were used to find the overlap between CAGE and/or MACE peaks with ChIP peaks or promoters and to perform statistical analysis of the contingencies[54]. For both mouse and fly, the identified regions with both PcG or transcription factor binding and MACE or CAGE tags in both directions (mono, divergent and convergent) were annotated using 'Fuge' (CSF Vienna) to the dm3 RefSeq genes (downloaded in October 2011 from the UCSC Genome Browser) or the mm9 RefSeq genes (downloaded in April 2010 from the UCSC Genome Browser).

40. Gibson, D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
41. Gemkow, M.J., Buchenau, P. & Arndt-Jovin, D.J. FISH in whole-mount *Drosophila* embryos. RNA: activation of a transcriptional locus, DNA: gene architecture and expression. *Bioimaging* **4**, 107–120 (1996).
42. Milán, M., Campuzano, S. & Garcia-Bellido, A. Cell cycling and patterned cell proliferation in the wing primordium of *Drosophila*. *Proc. Natl. Acad. Sci. USA* **93**, 640–645 (1996).
43. Aegerter-Wilmsen, T., Aegerter, C.M., Hafen, E. & Basler, K. Model for the regulation of size in the wing imaginal disc of *Drosophila*. *Mech. Dev.* **124**, 318–326 (2007).
44. Dietzl, G. *et al.* A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* **448**, 151–156 (2007).
45. Wang, J.W., Beck, E.S. & McCabe, B.D. A modular toolset for recombination transgenesis and neurogenetic analysis of *Drosophila*. *PLoS ONE* **7**, e42102 (2012).
46. Rai, A.N. *et al.* Elements of the polycomb repressor SU(Z)12 needed for histone H3-K27 methylation, the interface with E(Z), and *in vivo* function. *Mol. Cell. Biol.* **33**, 4844–4856 (2013).
47. Steffen, P.A. *et al.* Quantitative *in vivo* analysis of chromatin binding of Polycomb and Trithorax group proteins reveals retention of ASH1 on mitotic chromatin. *Nucleic Acids Res.* **41**, 5235–5250 (2013).
48. Keene, J.D., Komisarow, J.M. & Friedersdorf, M.B. RIP-Chip: the isolation and identification of mRNAs, microRNAs and protein components of ribonucleoprotein complexes from cell extracts. *Nat. Protoc.* **1**, 302–307 (2006).
49. Orlando, V., Jane, E.P., Chinwalla, V., Harte, P.J. & Paro, R. Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early *Drosophila* embryogenesis. *EMBO J.* **17**, 5141–5150 (1998).
50. Carrington, E.A. & Jones, R.S. The *Drosophila Enhancer of zeste* gene encodes a chromosomal protein: examination of wild-type and mutant protein distribution. *Development* **122**, 4073–4083 (1996).
51. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
52. Aszódi, A. MULTOVL: fast multiple overlaps of genomic regions. *Bioinformatics* **28**, 3318–3319 (2012).
53. Valouev, A. *et al.* Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat. Methods* **5**, 829–834 (2008).
54. Gentleman, R.C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).

doi:10.1038/ng.3058