Supplemental Methods

Read filtering, alignment and weighting

Low-complexity reads were filtered out based on their dinucleotide entropy (removing <1% of the reads). Alignments to the *D. melanogaster* genome were performed by the software bowtie (version 0.9.9.1) (Langmead et al. 2009) with parameters -v 2 -a -m 100, tracking up to 100 best alignment positions per query and allowing at most two mismatches. To track genomically untemplated hits (e.g., exon-exon junctions or missing parts in the current assembly), the reads were also mapped to an annotation database containing known *D. melanogaster* sequences (miRNA from ftp://ftp.sanger.ac.uk/pub/mirbase/sequences/11.0, rRNA,snRNA, snoRNA and RefSeq mRNA from GenBank http://www.ncbi.nlm.nih.gov/sites/entrez, downloaded on May 27, 2008, tRNA from http://lowelab.ucsc.edu/GtRNAdb/ and piRNA from GEO accession GSE9138). In that case, all best hits with at most two mismatches were tracked. Each alignment was weighted by the inverse of the number of hits. In the cases where a read had more hits to an individual sequence from the annotation database than to the whole genome, the former number of hits was selected to ensure that the total weight of a read does not exceed one. All quantifications were based on weighted alignments.

Calculation of read and IP enrichments

Enrichment of reads in genomic regions (see above for definition) were calculated as the ratio of observed over expected number of reads in a region, where the observed number is the sum of weights from overlapping read alignments, and the expected number of reads is the fraction of genomic bases in that region type, multiplied with the total number of aligned reads in the sample.

IP enrichments of a genomic region (TSS windows or peak region) were calculated as e = log2((n_fg /N_fg *min(N_fg,N_bg) + p) / (n_bg /N_bg *min(N_fg,Nbg) + p)), where n_fg and n_bg are the summed weights of overlapping foreground and background (input chromatin) read alignments, respectively. N_fg and N_bg are the total number of aligned reads in foreground and background samples, and p is a pseudocount constant (p=8) used to regularize enrichments based on low counts that would otherwise be dominated by sampling noise.

Cut-offs for TSS window enrichments were manually defined for each sample using a scatter plot of the two biological replicates against one another to separate correlated higher enrichments (positives) from the presumably negative and uncorrelated lower enrichment values (Supplementary Figure 1). The resulting cut-off values used for in this study were: Pc=0.4, Ph=0.1, Psc=0.4, Trx=0.3.

Calculation of RNA expression levels

Expression levels of RefSeq transcripts were calculated as x = log2(n /len *avg_len + 1), where n is the weighted sum of read alignments to a RefSeq transcript, len is the length of the transcript and avg_len is the mean length over all RefSeq transcripts. Expressed transcripts were defined based on the overall distribution of expression levels as transcripts with expression levels of at least 5.0.

Peak finding

Genomic regions of increased ChIP-seq read alignment densities were identified using MACS (version 1.3.7.1) (Zhang et al. 2008), using a pool of read alignments from all biological replicates (weights rounded to integers) as input, parameters --mfold=4 --gsize=150000000 --tsize=36 and default values for all other parameters. IP enrichments (see above) of resulting peak candidates were calculated and peak candidates with

enrichments lower than 1.5-fold above background (e < 0.585) were removed, resulting in 2826, 4402, 2108 and 5240 peaks for Pc, Ph, Psc and Trx proteins, respectively. Genomic regions that were overlapping peaks from any three of the four analyzed proteins were defined as PcG binding sites (n=2,274).

Similarly, TSS clusters were identified from pooled 5'MACE read alignments using MACS with parameters --bw=100 --futurefdr --nomodel --shiftsize=1 --gsize=150000000 --tsize=36 --pvalue=1e-3 and default values for all other parameters (n=9626).

Distances between peaks or between peaks and TSSs were measured from and to middles of peak regions.

Computational analysis of RNA polymerase stalling in S2 cells

For re-analysis of data from ((Nechaev et al. 2010), GSM463298, GSM463299, GSM463297) fastq-files were downloaded from the short read archive under accession numbers (SRX015/SRX015109, SRX015/SRX015110, and SRX015/SRX015111). Reads were aligned with bwa (Li and Durbin 2009) and default options to the R5/dm3 assembly of the *D. melanogaster* genome. The alignments were further processed with the samtools library (Li et al. 2009) and pysam (http://pysam.googlecode.com) to compute TSS profiles at annotated, non-overlapping transcription start sites.

TSS profiles of the 3' and 5' short promoter-associated RNAs were obtained by counting the first nucleotide and relative position of reads mapping in the correct orientation in a 200nt window around known TSS. For replicate experiments TSS profiles were added.

The mean 3'-abundance for PRC1-regulated promoters was calculated as the average over all TSS that were grouped as PRC1+ or PRC1- (see above; Figure 3A). The total of all short RNA libraries (3' + 5') was computed by summing over all positions in the TSS

profiles and experiments for PRC1+ and PRC1- TSSs, respectively, and recording the number of TSS with zero short RNAs (Figure 3B).

RNA Pol II ChIP-Seq Watson and Crick tags were shifted by +/- 65 nt, respectively, before the TSS profiles were computed and the stalling index was calculated as described in (Nechaev et al. 2010). Differences in the stalling index between PRC1+/- groups were tested with a Kolmogorov-Smirnov test (Figure 3C).

Goterm enrichment analysis

Functional classifications of PcG target genes were assessed using the DAVID tool (DAVID v6.7, http://david.abcc.ncifcrf.gov/home.jsp) (Huang et al. 2009) with standard settings. P-values were corrected according to multiple hypothesis testing using the method of Benjamini-Hochberg. The set of PcG target genes contained a total of 1126 Refseq mRNAs with TSSs less than 500 bp distant from the middle of a PcG target site (see Supplemental Table 3) recognized as 1046 DAVID IDs.

In situ hybridization and handling of *Drosophila* Fly strains

The the Pc[3] allele was balanced with TM3, kr>GFP (FBst0005195), (Casso et al. 2000). Homozygous embryos could be distinguished using a T7 RNA Polymerase transcribed, fluorescin-labelled RNA probe against GFP-S65T together with a DIG-labelled probe against full-length *pri-mir-iab8* in double *in situ* hybridizations. Hybridizations were performed as described (Hauptmann and Gerster 2000) with following modifications: ~30 µl embryos were hybridized over night at 65˚C in 50µl hybridization buffer (50% (v/v) Formamide, 5xSSC, 50ug/mL heparin, herring sperm DNA 10ug/mL, 0.1% (w/v) Tween 20, 5x Denhardt's, 5mg/mL yeast torula RNA), containing RNA-probe from *in vitro* transcription of 1 µg linearized plasmid template in a final dilution of 1:5000. Probes were detected with anti-digoxigenin-AP coupled

antibodies plus NBT/BCIP and anti-fluorescein-AP plus INT/BCIP (Roche), respectively.

AP-buffer contained 1mM Levamisol hypochloride (SIGMA).

Lab protocol for preparation of a 5'-MACE library

In order to prepare a genome-wide library of 5' cDNA ends suitable for next generation sequencing this method uses a combined approach of RNAoligo-ligation and tagged random hexamer amplification. The end product is a size distributed 1st strand cDNA library ready for amplification, size selection and directional massive parallel sequencing.

Steps included in protocol:

1) Dephosphorylating RNA
2) Removing the mRNA Cap Structure
3) Ligating the RNA oligo to the decapped RNA
4) 1st strand cDNA synthesis using tagged random-hexamers
5) 2nd strand cDNA synthesis by limited PCR amplification
6) Size Selection of amplimers via agarose gel electrophoresis
7) Final amplification of 350 bp library
8) Massive Parallel Sequencing with Illumina's Genome Analyzer

Oligonucleotides:

> RiboAdapter 1.4
acacucuuucccuacacgacgcucuuccgaucuaaa

> Adapter 1.1revtag_hexamer
caagcagaagacggcatacgagctcttccgatctnnnnnn

>Primer 1.2 (incl. RiboAdapter 1.4 sequence)
aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatct

>Primer 1.1_rev
caagcagaagacggcatacgagctcttccgatct

>Illumina Genomic DNA PCR Primer 2.1
aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatct

>Illumina Genomic DNA PCR Primer 1.1
caagcagaagacggcatacgagctcttccgatct

1) Dephosphorylating RNA

In this step, you treat total RNA or mRNA with calf intestinal phosphatase (CIAP) to dephosphorylate non-mRNA or truncated mRNA

Dephosphorylation Reaction

1. Set up on ice the following 20 µl dephosphorylation reaction in a 1.5 ml sterile micro centrifuge tube

Use 10 µg of total RNA

(Trizol extracted, TURBO DNAse treated, verified integrity by Nanodrop and Agilent Bioanalyzer)

RNA	x µl
10X CIAP Buffer	2 µl
RNaseOut™ (40 U/µl)	2 µl
CIAP (20 U/µl)	1 µl
DEPC water	y µl
Total Volume	20 µl

2. Mix gently by pipetting and vortex briefly. Centrifuge to collect fluid.

3. Incubate at 50°C for 1 hour.

4. After incubation, centrifuge briefly and place on ice.

RNA cleanup (1)

// avoid to use cooled centrifuge, work quickly.

1. Adjust to a volume of 100 µl by adding 80 µl RNase free water.

2. Add 350 µl buffer RLT and mix well.

3. Add 250 µl 100% Ethanol to the RNA, vortex briefly and immediately proceed to 4 without centrifuging.

4. Transfer the sample (700 µl) to an RNeasy MinElute spin column placed in a 2 ml collection tube. Close the lid gently, and centrifuge for 15 s at full speed. To avoid precipitation in subsequent steps keep centrifuge and columns at room temperature and work swiftly. Discard flow through and tube.

5. Place the column in *a new* 2 ml tube. Add 500 µl Buffer RPE to the spin column. Close the lid gently, and centrifuge for 15 s at full speed. Discard the flow-through.

6. Add 500 µl of 80% ethanol to the spin column. Close the lid gently, and centrifuge for 2 min at full speed to wash the membrane. Carefully remove column and discard collection tube.

7. Place the spin column in a new 2 ml collection tube for a drying spin. Open the lid of the spin column, and centrifuge at full speed for 5 min. Discard the flow-through and collection tube.

8. Place the spin column in a new 1.5 ml tube. Add 18 µl RNase-free water directly to the center of the spin column membrane. Incubate 30 sec.

9. Close the lid gently, and centrifuge for 1 min at full speed to elute the RNA in a total volume of 16 µl

2) Removing the mRNA Cap Structure

Decapping Reaction

1. Set up on ice the 20 µl decapping reaction in a 1.5 ml sterile micro centrifuge tube using the reagents in the kit.

Dephosphorylated RNA	16	µl
10X TAP Buffer	2	µl
RNaseOut™ (40 U/µl)	2	µl
TAP (5 U/µl)	0.2	µl
Total Volume	20	µl

2. Mix gently by pipetting and vortex briefly. Centrifuge briefly to collect fluid.

3. Incubate at 37°C for 1 hour.

4. After incubation, centrifuge briefly and place on ice.

RNA cleanup (2)

1. Adjust to a volume of 100 µl by adding 80 µl RNase free water.

2. Add 350 µl buffer RLT and mix well.

3. Add 250 µl 100% Ethanol to the RNA, vortex briefly and immediately proceed to 4 without centrifuging.

4. Transfer the sample (700 µl) to an RNeasy MinElute spin column placed in a 2 ml collection tube. Close the lid gently, and centrifuge for 15 s at full speed. To avoid precipitation in subsequent steps keep centrifuge and columns at room temperature and work swiftly. Discard flow through and tube

5. Place the column in *a new* 2 ml tube. Add 500 µl Buffer RPE to the spin column. Close the lid gently, and centrifuge for 15 s at full speed. Discard the flow-through.

6. Add 500 µl of 80% ethanol to the spin column. Close the lid gently, and centrifuge for 2 min at full speed to wash the membrane. Carefully remove column and discard collection tube.

7. Place the spin column in a new 2 ml collection tube for a drying spin. Open the lid of the spin column, and centrifuge at full speed for 5 min. Discard the flow-through and collection tube.

8. Place the spin column in a new 1.5 ml tube. Add 14 µl RNase-free water directly to the center of the spin column membrane. Incubate 30 sec.

9. Close the lid gently, and centrifuge for 1 min at full speed to elute the RNA in a total volume of 12 µl

3) Ligating the RNA Oligo to Decapped mRNA

Once you have decapped the mRNA, you are ready to ligate the Solexa Sequencing compatible Adapter1.4 RNAoligo to the 5' end of your mRNA. This primer contains the Illumina Genomic Primer Sequence and additional adenines at the 3'-end to allow for more efficient RNA ligation (Uhlenbeck and Gumport, 1982 the enzymes).

> RiboAdapter 1.4
acacucuuucccuacacgacgcucuuccgaucuaaa

Example calculation for RNA oligo concentrations

For an RNAoligo concentration 17.755 pMol per µl (convenient 2 µl volume to add to reaction):Shortly spin tube with dried RNAoligo. Add 56.3 µl Nuclease-Free water to one aliquot containing 1000 pMol of RNAoligo. Vortex shortly and spin down again

Ligation Reaction

1. To the tube containing 12 µl of dephosphorylated, decapped RNA add 2 µl of RNA Oligo (35.51 pmol or 0.42 µg in total). Pipet up and down several times to mix RNA-Oligo thoroughly. Centrifuge briefly to collect the fluid in the bottom of the tube.

2. Incubate at 65°C for 5 minutes to relax the RNA secondary structure.

3. Place on ice to chill (~2 minutes) and centrifuge briefly.

4. Add the following reagents to the tube, mix gently by pipetting, and centrifuge briefly (evaporation due to 65°C incubation already considered).

10X Ligase Buffer	2 µl
10 mM ATP	2 µl
RNaseOut™ (40 U/µl)	2 µl
T4 RNA ligase (5 U/µl)	2 µl
Total Volume	~20 µl

5. Incubate at 37°C for 75 minutes.

6. Centrifuge briefly and place on ice.

RNA cleanup (3)

1. Adjust to a volume of 100 µl by adding 80 µl RNase free water.

2. Add 350 µl buffer RLT and mix well.

3. Add 250 µl 100% Ethanol to the RNA, vortex briefly and immediately proceed to 4 without centrifuging.

4. Transfer the sample (700 µl) to an RNeasy MinElute spin column placed in a 2 ml collection tube. Close the lid gently, and centrifuge for 15 s at full speed. To avoid precipitation in subsequent steps keep centrifuge and columns at room temperature and work swiftly. Discard flow through and tube

5. Place the column in *a new* 2 ml tube. Add 500 µl Buffer RPE to the spin column. Close the lid gently, and centrifuge for 15 s at full speed. Discard the flow-through.
6. Add 500 µl of 80% ethanol to the spin column. Close the lid gently, and centrifuge for 2 min at full speed to wash the membrane. Carefully remove column and discard collection tube.

7. Place the spin column in a new 2 ml collection tube for a drying spin. Open the lid of the spin column, and centrifuge at full speed for 5 min. Discard the flow-through and collection tube.

8. Place the spin column in a new 1.5 ml tube. Add 18 µl RNase-free water directly to the center of the spin column membrane. Incubate 30 sec.

9. Close the lid gently, and centrifuge for 1 min at full speed to elute the RNA in a total volume of 16 µl

At this point 8 µl RACE-ready RNA can be e.g. stored as a backup at -80°C while the other sample is directly processed for cDNA synthesis.

4) 1st strand cDNA synthesis using random-hexamers tagged with Adaper 1.1

Once you have the 5'-Adapter1.4 RNA-ligation product you are ready to produce a 1st strand cDNA containing both adapters by using random-hexamers tagged with the second adapter sequence. This will create a pool of 1st strand cDNAs with a suitable size distribution.

> Adapter 1.1revtag_hexamer
caagcagaagacggcatacgagctcttccgatctnnnnnn

First-Strand cDNA Synthesis using the Superscript III Kit

1. Mix and briefly centrifuge each component before use.

2. Combine the following in a 0.2 ml tube:

Riboadapter 1.4-ligated RNA		*8* µl
Adapter.1.1	(2µM)	1 µl
10 mM dNTP mix		1 µl
Total is		10 µl

3. Incubate at 65°C for 5 min, then place on ice for 2 min.

4. Prepare the following cDNA Synthesis Mix, adding each component in the indicated order.

10X RT buffer		2 µl
25 mM MgCl2		4 µl
0.1 M DTT		2 µl
RNaseOUT	(40 U/µl)	1 µl
SuperScript III RT	(200 U/µl)	1 µl
Total is		10 µl

5. Add 10 µl of cDNA Synthesis Mix to each RNA/primer mixture, mix gently, and collect by brief centrifugation. Incubate as follows in a Thermocycler:

10 min at 25°C
50 min at 50°C
5 min at 85°C

7. Collect the reactions by brief centrifugation. Add 1 µl of RNase H and incubate for 20 min at 37°C.

8. 20 µl cDNA synthesis reaction can be stored at -20°C for several months or used for PCR immediately.

5) 2nd strand cDNA synthesis by limited PCR amplification

From the double-adapter-tagged cDNA library you can now amplify suitable amount of 2nd strand cDNA for size selection on an agarose gel.

>Primer 1.2 (incl adapter 1.2 sequence)
aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatct

>Primer 1.1_rev
caagcagaagacggcatacgagctcttccgatct

5x HF-Buffer		5	µl
dNTP mix	(25 mM)	0.25	µl
Primer 1.1 rev	(5 µM)	1.5	µl
Primer 1.2	(5 µM)	1.5	µl
1st strand cDNA		10	µl
water		6.5	µl

mix thoroughly and perform hot start

Phusion Pol	0.25	µl
Total volume	25	µl

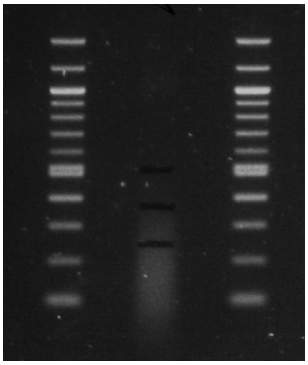
Thermocycle as follows:	1 x	98°C -	30 sec
	8 x	98°C	10 sec
		65°C	30 sec
		72°C	30 sec
		72°C	5 min

6) Size Selection of amplimers via gel electrophoresis

1. Prepare a 1.5% agarose gel in a final volume of 50 mL 1x TAE buffer.
2. Load 30µL (25 µl of 2nd strand cDNA + 5µl 10x loading buffer) of sample into one well and 1µL of 100 bp DNA ladder (Invitrogen 15628-019) into another well.
3. Run gel at 120V for 45~60min until sufficient separation of the of the DNA ladder.
4. Cut a gel slice at 350 bp (+/- 25 bp) and purify the DNA with QIAquick gel extraction kit (Qiagen #28706). Elute the DNA into 30 µL of EB solution.

The cDNA Library will be visible as a faint smear up to 500 bp and the gel can be cut at several points for evaluation. Gel slices can be stored at -20°C for several weeks. If only a strong 92 bp amplicon is visible the library amplification failed and the only product left is a Riboadapter/Adapter 1.1 dimer.

An example gel with a 5'MACE library is shown below (cut at 250, 350 and 500 bp)



7) PCR enrichment of cDNA templates

>Illumina Genomic DNA PCR Primer 2.1
aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttccgatct

>Illumina Genomic DNA PCR Primer 1.1
caagcagaagacggcatacgagctcttccgatct

1. set up PCR mastermix and aliquot 20 µl into PCR tube

5 x cloned Phusion Buffer (NEB, #F-530)	10	µL
PCR primer 1.1	1	µL
PCR primer 2.1	1	µL
25 mM dNTP mix	0.5	µL
Phusion polymerase (NEB, #F-530)	0.5	µL
H2O	22	µL
purified 2nd strand cDNA	15	µl
Total volume	50	µl

Thermocycle as follows:	1 x	98°C	30 sec
	18 x	98°C	10 sec
		65°C	30 sec
		72°C	30 sec
		72°C	5 min

2. Purify the DNA with a QIAquick column (Qiagen, #28106) and elute in 30µL of EB solution.

4. Run 1 µL of the purified product from step 4 on an Agilent Bioanalyzer DNA 1000 chip (Agilent, #5067-1504) to QC the final product and quantify the DNA concentration. Successful preparation should yield a distinct band at ~350 bp.

8) Massive Parallel Sequencing with Illumina's Genome Analyzer

This material is suitable for cluster generation on the Illumina Cluster Station using standard protocols. The quality of the final library pool can be accessed by transforming 1 µl into e.g. TOPO-pCR blunt vectors and sanger sequencing 5-10 clones.

Using the Genomic DNA Sequencing Primer will exactly reproduce the (strand specific) RNA sequence (with an addition of the sequence AAA at the 5' end due to Riboadapter1.4)

>Genomic DNA sequencing Primer
acactctttccctacacgacgctcttccgatct

9) Additional Information

References:

Several references were used for construction of the protocol (Maruyama and Sugano 1994; Wong et al. 1996) including the manual of Invitrogen`s Genracer Kit and Illumina's mRNA-seq protocol.

Material:

Calf Intestinal Alkaline Phosphatase (CIAP)(20 U/µl)	1,000 units	SKU# 18009-027	Invitrogen
Tobacco Acid Pyrophosphatase (TAP)(5U/µl)	50 Units	T81050	Epicentre
T4RNA ligase (5 U/µl)	1,000 Units	LR5010	Epicentre
RNeasy MinElute Cleanup Kit (50)	50 rxns	74204	Quiagen
Superscript III First strand Synthesis System	50 rxns	18080-051	Invitrogen
QIAquick gel extraction kit (Qiagen #28706)	50 rxns	#28706	Qiagen
Phusion High Fidelity DNA Polymerase	100 U	F-530S	Finnzymes

Supplemental References

Casso D, Ramírez-Weber F, Kornberg TB. 2000. GFP-tagged balancer chromosomes for Drosophila melanogaster. *Mech Dev* **91**(1-2): 451-454.

Hauptmann G, Gerster T. 2000. Multicolor whole-mount in situ hybridization. *Methods Mol Biol* **137**: 139-148.

Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**(1): 44-57.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**(3): R25.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**(14): 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**(16): 2078-2079.

Maruyama K, Sugano S. 1994. Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene* **138**(1-2): 171-174.

Nechaev S, Fargo DC, dos Santos G, Liu L, Gao Y, Adelman K. 2010. Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. *Science* **327**(5963): 335-338.

Wong KK, Stillwell LC, Dockery CA, Saffer JD. 1996. Use of tagged random hexamer amplification (TRHA) to clone and sequence minute quantities of DNA--application to a 180 kb plasmid isolated from Sphingomonas F199. *Nucleic Acids Res* **24**(19): 3778-3783.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**(9): R137.