CAT7 and cat7l Long Non-coding RNAs Tune Polycomb Repressive Complex 1 Function during Human and Zebrafish Development

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

| Citation       | Ray, Mridula K., Ole Wiskow, Matthew J. King, Nidha Ismail, Ayla Ergun, Yanqun Wang, Aaron J. Plys, et al. 2016. “CAT7 and cat7l Long Non-Coding RNAs Tune Polycomb Repressive Complex 1 Function during Human and Zebrafish Development.” Journal of Biological Chemistry 291 (37): 19558–72. https://doi.org/10.1074/jbc.m116.730853. |
|----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Citable link   | http://nrs.harvard.edu/urn-3:HUL.InstRepos:41461214                                                                                                                                                                                                                     |
| Terms of Use   | This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA                                                                                                                                 |
CAT7 and cat7l Long Non-coding RNAs Tune Polycomb Repressive Complex 1 Function during Human and Zebrafish Development*§

Received for publication, April 3, 2016, and in revised form, July 1, 2016 Published, JBC Papers in Press, July 12, 2016, DOI 10.1074/jbc.M116.730853

Mridula K. Ray‡1, Ole Wiskow‡1, Matthew J. King‡‡, Nidha Ismail‡2, Ayla Ergun‡3, Yanqun Wang‡3, Aaron J. Plys‡‡4, Christopher P. Davis‡‡5, Katie Kathrein‡6, Ruslan Sadreyev‡7, Mark L. Borowsky‡‡8, Kevin Eggan‡‡4, Leonard Zon‡‡9, Jenna L. Galloway‡, and Robert E. Kingston‡8

From the ‡Department of Molecular Biology, Massachusetts General Hospital, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114, §Harvard Stem Cell Institute, Department of Stem Cell and Regenerative Biology, Harvard University and the Stanley Center for Psychiatric Research, Broad Institute, Cambridge, Massachusetts 02138, ¶Center for Regenerative Medicine, Department of Orthopaedic Surgery, Massachusetts General Hospital, Harvard Stem Cell Institute, Harvard Medical School, Boston, Massachusetts 02114, ‡‡Department of Molecular Biology, Massachusetts General Hospital, and Department of Pathology, Harvard Medical School, Boston, Massachusetts 02114, **Division of Hematology/Oncology, Boston Children’s Hospital, Dana-Farber Cancer Institute, Harvard Stem Cell Institute, Boston, Massachusetts, 02115, and ††‡The Howard Hughes Medical Institute, Cambridge, MA 02138

The essential functions of polycomb repressive complex 1 (PRC1) in development and gene silencing are thought to involve long non-coding RNAs (lncRNAs), but few specific lncRNAs that guide PRC1 activity are known. We screened for lncRNAs, which co-precipitate with PRC1 from chromatin and found candidates that impact polycomb group protein (PcG)-regulated gene expression in vivo. A novel lncRNA from this screen, CAT7, regulates expression and polycomb group binding at the MNX1 locus during early neuronal differentiation. CAT7 contains a unique tandem repeat domain that shares high sequence similarity to a non-ontic zebrafish analog, cat7l. Defects caused by interference of cat7l RNA during zebrafish embryogenesis were rescued by human expression and polycomb group binding at the locus during development.

Cellular specification during development is an intricate process involving many layers of genomic regulation, including gene repression and chromatin modification by the PcG families. Recent evidence suggests that a diverse class of lncRNAs may tune repression in a cell-type specific manner by initiating or stabilizing binding of PcG complexes to individual gene loci. Most of this work focuses on the polycomb repressive complex 2 (PRC2) family of PcG complexes, which methylates lysine 27 of histone H3 (H3K27me3) but whose activity is not sufficient to silence chromatin. The PRC2 family of complexes interacts with thousands of coding and non-coding transcripts in vivo, with varying degrees of specificity in vitro (3–6), making it difficult to discern which specific lncRNAs functionally impact PcG biology via interaction with PRC2.

In contrast, the potential for lncRNAs to influence PRC1 function or targeting has been less studied despite the clear mechanistic importance of PRC1 to stable gene silencing. PRC1 family complexes are functionally and compositionally diverse (7, 8) (Fig. 1A) and either ubiquitylate histone H2A or compact nucleosomes in vitro. This activity is necessary for stable repression in vivo, presumably limiting access of the transcriptional machinery to DNA (9–11). The complexes are guided to chromatin by various mechanisms, including binding to H3K27me3 deposited by PRC2 (12) and PRC2-independent mechanisms (13). Recent studies have reported a handful of RNAs that functionally affect gene silencing and might directly impact PRC1 targeting to specific loci (14–16). Defining additional lncRNAs that affect recruitment or silencing by PRC1 may provide novel insights into the regulation of PcG repression during development.

* This work was supported, in whole or in part, by National Institutes of Health Grant RO1-GM043901 (to R. E. K.). The authors declare that they have no conflicts of interest with the contents of this article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

† This article contains supplemental Tables 1–4. All sequencing data can be found under GEO submission GSE61549 and can be accessed at www.ncbi.nlm.nih.gov.

1 Supported by the ALS Association.
2 Supported by funding from the Eunice Kennedy Shriver NICHD, National Institutes of Health Grant R00HD069533 and the Harvard Stem Cell Institute (HSCI).
3 Present address: Novartis Institutes for BioMedical Research, Cambridge, MA 02142.
4 Supported by a National Research Service Award fellowship from the National Institutes of Health (F32GM109693).
5 Supported by a graduate research fellowship from the National Science Foundation.
6 A Howard Hughes Medical Investigator early career scientist.
7 To whom correspondence may be addressed. E-mail: jenna_galloway@hms.harvard.edu.
8 To whom correspondence may be addressed. E-mail: kingston@molbio.mgh.harvard.edu.
9 The abbreviations used are: PcG, polycomb group; lncRNA, long non-coding RNA; PRC1 and PRC2, polycomb repressive complex 1 and 2, respectively; CAT7, chromatin-associated transcript; TRD, tandem repeat domain; GRIP, gradient RNA immunoprecipitation; qPCR, quantitative PCR; FDR, false discovery rate; hESC, human embryonic stem cell; hpf, hours post fertilization; MO, morpholino; RACE, rapid amplification of cDNA ends; IP, immunoprecipitate; RPKM, reads per kilobase per million.
Identifying lncRNAs that impact PRC1 biology is challenging because it is not known which specific component(s) of PRC1 (or ancillary factors) might interact with RNA, limiting the utility of in vitro approaches or UV-cross-linked cross linked IP approaches. To address this challenge, we developed a method to systematically identify lncRNAs, which are both in close physical proximity to PRC1 family complexes on chromatin and which affect PRC1 function regardless of the exact components that interact with RNA. We isolated candidate lncRNAs via an immunoprecipitation from cross-linked, purified chromatin with antisera specific for the BMI1 subunit of PRC1. We identified a number of co-precipitating lncRNAs, referred to as CATs (chromatin-associated transcripts), and tested the ability of these lncRNAs to influence regulation of PRC1 function. We focused on a lncRNA from this screen called CAT7, which contains a tandem repeat domain (TRD) found in both humans and zebrafish. Using RNA depletion experiments, we showed that human CAT7 expression is necessary for full repression of the nearby PcG-target gene, MNX1, during neuronal differentiation, and that cat7l, the independently evolved zebrafish analog of CAT7, interacts genetically with zebrafish bmi1a/b to affect brain development. Deficiency of cat7l during zebrafish embryogenesis can be functionally rescued by human CAT7 expression of PcG-regulated genes, we individually knocked down 17 of the 43 candidates and sequenced total RNA after 48 h (Tables 3 and 4). We identified differentially expressed genes in cells treated with pooled siRNAs targeting the lncRNA versus cells treated with a scrambled control. Knockdown of at least 5 novel candidates distinctly and reproducibly altered PcG-target gene expression (p-value < 0.001, hypergeometric probability). For these 5 candidates, PcG-target genes were among the top 10 overrepresented categories of genes from 6176 previously curated gene sets from HeLa and other cell types (20) (FDR < 0.001 hypergeometric probability, Fig. 2, C and D), a feature not observed after knockdown of several mRNA/lncRNA controls or between scrambled knockdown controls (supplemental Table 4). We conclude that multiple candidates from our screen cause distinct, widespread molecular phenotypes upon knockdown, characteristic of perturbations to PcG-regulated gene networks (21, 22). We note that this screen cannot discriminate between direct and indirect changes to PcG target gene expression, and we anticipate that some observed changes in expression are indirect/cascading effects typical of lncRNAs that perturb PcG-silenced master regulators (6, 23).

Description of Purified lncRNAs—We anticipated that lncRNAs that impact PcG biology would be primarily localized to the nucleus and likely expressed in a tissue-specific manner in cells from across the body. We characterized the 64 candidates selected as above with respect to cellular localization and expression in various tissues and also examined transcript abundance and possible sequence motifs. To query nuclear
CAT7 and cat7l IncRNAs Tune PRC1 Function

localization, we examined cell-fractionated RNA-seq data from ENCODE and found that 53/59 of the candidates (for which there was sufficient signal) were primarily or exclusively observed in the nuclear fraction (supplemental Table 2, Fig. 2B). This set included all 43 candidates that were validated in the biological replicate of the BMI1 gradient RNA IP. Nuclear localization might be expected from RNAs identified via a chromatin fractionation and co-precipitated by a nuclear factor. However, in contrast, PRC2 binds many (possibly nascent) mRNA exons (4, 19), which are primarily transported to the cytoplasm to enact their known functions. The nuclear localization supports the hypothesis that the primary functional roles of the candidates identified in this screen likely occur within the nucleus. We also analyzed the sequences of the candidates but were not able to find compelling specific sequence motifs. Many candidates (22 of 43; supplemental Table 2), such as CAT7 and XIST REP-E (Fig. 1, F and G, supplemental Fig. 2C), contained unique tandem repeats, which are much less likely to be contained in mRNA exons. Although the different characteristics of the RNAs that interact with PRC1 and PRC2 may partially be attributed to the technical implementation of GRIP versus prior IP techniques, it is possible that PRC1 and PRC2, known to have distinct activities, also have distinct functional interactions with RNAs.

We further tested RNA expression of several candidates across a variety of adult and embryonic tissues by RT-qPCR and found that the IncRNA candidates were expressed in a highly tissue-specific manner (data not shown). We found that expression levels of the 64 candidates varied widely in the input fraction (supplemental Table 1). This demonstrates that the candidates we identified are expressed at a large range of abundances (>100-fold range of expression in input, measured in RPKM).

**CAT7 Regulates MNXI Expression in Neuronal Lineages—**

To further explore possible regulatory relationships between the candidate IncRNAs and PRC1, we focused on the 1.6-kb, polyadenylated, single-exon IncRNA, CAT7 (RACE-PCR results in Fig. 1J). We selected CAT7 from the five candidates above because its knockdown reproducibly impacted many PcG-regulated genes (supplemental Table 4) and because it contains a TRD that is unique to a single locus in the genome yet is present in several organisms (see below). We hypothesized that CAT7 may positively influence PcG binding at gene(s) whose expression increased after CAT7 knockdown. We examined the relationship between CAT7 and the up-regulated PcG-target, MNXI, which stood out because of its proximity to the CAT7 gene (~400 kb) and the precedent for IncRNAs to regulate gene expression over a long range in cis (24). MNXI is essential for formation of both pancreatic beta islets and motor neurons, and its dysfunction is implicated in neonatal diabetes (25) and the dominant developmental disorder, Currarino syndrome (26). Moreover, CAT7 is located inside a gene desert region between SHH and RNF32. This region, extending beyond RNF32 toward MNXI (a target of SHH) is highly cross-regulatory (27), supporting the hypothesis that there could exist a regulatory relationship between CAT7 and the MNXI locus.

During early motor neuron differentiation from human embryonic stem cells (hESCs), MNXI is initially silenced (28). CAT7 expression commences at day 2 (estimated ~5 copies/cell), whereas the rapid up-regulation of MNXI did not commence until day 6 (Fig. 3A). We hypothesized that CAT7 impacts the temporal regulation of the MNXI locus. Specifically, as with many genes silenced in hESCs, PcG proteins are associated with MNXI, but further mechanisms are required to modulate PcG silencing as cells traverse through differentiation (28). Thus, CAT7 might help to maintain PcG repression before cells become fully committed to the motor neuron lineage. To test this, we knocked down CAT7 in hESCs (day 0) and exposed the cells to posteriorizing neural differentiation conditions (29). Subsequently, we harvested the cells after 72 h of differentiation (before de-repression of MNXI under normal conditions).

After CAT7 knockdown, expression profiling by RNA sequencing showed an increase in expression of MNXI in the posteriorized neural culture relative to the scrambled control (supplemental Table 4). We confirmed this result by quantifying MNXI up-regulation by RT-qPCR alongside two other SHH/PcG-regulated genes, ISL1 and IRX3, which were up- or down-regulated after CAT7 knockdown, respectively (Fig. 3B), and which show analogous expression patterns in primary mature motor neurons (28). Additionally, we saw similar expression changes across the genes in hESC-derived motor neuron progenitors generated from an alternative differentiation protocol, in neuronal cells treated with a variety of anti-sense oligos, and in HeLa cells after CAT7 knockdown (Fig. 3, B–D). CAT7 knockdown did not significantly alter expression of SHH, a known regulator of MNXI (28), suggesting that the effect on MNXI is not a consequence of SHH up-regulation (supplemental Table 4) and that CAT7 might regulate the MNXI locus.

If the observed up-regulation of MNXI was caused by impaired PRC1-mediated repression after CAT7 knockdown, we might expect to see lowered PcG binding at the MNXI pro-

---

*FIGURE 1. PRC1 co-precipitates with RNAs on chromatin.* A, PRC1 family complexes (adapted from Ref. 13). B–H, HeLa nuclear lysate was applied to a CsCl density gradient. Fractions were probed by Western blot (C), Bradford assay (D), absorbance at 260 nm (E), DNA purification and agarose gel electrophoresis (F), RNA purification and electrophoresis (via Agilent Bioanalyzer) (G), and RT-qPCR of various long non-coding RNAs or controls (H). In A, SUZ12 is a subunit of the related complex; PRC2; PHC1 is another PRC1 component; H3 is representative of bulk histones; CTCF is a widely bound chromatin protein that is not known to interact with PRC1. I–K, RNAs co-precipitating with BMI1 and FLAG-Bmi1 (using fractions 4–6 as input (Inp)) were sequenced. Here, I and J represent positive, chromatin-bound hits from unannotated and annotated IncRNA regions, and K is a negative IncRNA region. These findings are replicated via qPCR in R, and XIST expression and localization in these cells is validated by RNA-FISH (L). Enrichment (reads in IP/reads in input) was compared across the following regions of the genome: de novo peaks found in both BMI1 and FLAG-Bmi1 (M), mRNA exons (N), de novo peaks intersecting annotated IncRNAs (O), or annotated IncRNAs (P) (see “Experimental Procedures”). The number of significantly co-enriched RNAs is shown in red versus totally in black. Correlation coefficients (Pearson’s) between BMI1 and FLAG-Bmi1 IPs across the plotted regions are noted. 17 peaks (blue) were selected for downstream validation. Q, GRIP eluate was analyzed to show that BMI1 or FLAG-Bmi1 is selectively bound at known peaks from a published BMI1 positive control region from HeLa cells (see supplemental Table 2; Ref. 46). Here, FLAG-Bmi1 composes 20% of the total BMI1/Bmi1 in the cell.
To test this model, we performed ChIP-qPCR at the MNX1 locus with antisera targeting a PRC1 component, a PRC2 component, and H3K27me3 (HeLa only) in CAT7-knockdown or scrambled-control cells. We performed these experiments in both HeLa and hESC-derived posteriorized neural culture. In each case we observed a ~30% loss in PRC1/K11011

**FIGURE 2. Candidate lncRNAs impact distinct PcG-regulated transcriptional networks.** A and B, selected candidate lncRNAs were knocked down by siRNAs in the nucleus, as assayed by RT-qPCR (n = 2, error = S.D.) (A) and Northern blot (B). C, genes that are up-regulated after lncRNA knockdowns compared with the scrambled control show significant overlap with PcG-target genes. PcG-targets are among the most overrepresented from >6000 gene sets in the Molecular Signatures (MolSig) Database (hypergeometric FDR). The statistics shown reflect the most significantly enriched PcG-related gene subset in the MolSig database. D, chromosomal locations of lncRNAs and the up-regulated PcG-target genes (from MolSig database) are mapped; there is no significant overlap of up-regulated genes caused by different siRNA knockdowns. **, FDR < 0.001.
PRC2/H3K27me3 signal at the MNX1 promoter after CAT7 knockdown (Fig. 3, E and F), a level of effect similar to those caused by knockdown of numerous known PcG-interacting lncRNAs (17, 31). Various PRC1 family complexes and PRC2 positively cross-regulate each other (12, 32, 33), so it is not surprising that both PRC1 and PRC2 subunit binding are impacted by CAT7 knockdown. Conversely, we saw no significant changes in ChIP signal at a HOX positive control locus and minimal signal in all IPs at an intergenic, negative-control locus. We conclude that knockdown of CAT7 de-repressed MNX1 and decreased PcG binding at the MNX1 promoter during human neuronal differentiation as well as in HeLa cells.
**CAT7 and cat7l IncRNAs Tune PRC1 Function**

This supports the hypothesis that a class of IncRNAs can tune PcG binding and function at individual loci.

A CAT7 Analog Interacts Genetically with bmi1a/b in Zebrafish—CAT7 has broad sequence similarity to genomic regions across many species (supplemental Table 3), raising the possibility that CAT7-like RNAs are part of a generalized mechanism for regulating PRC1 function. We focused on a unique ~1.2-kb TRD of CAT7 (Fig. 1G) that has 57% sequence identity with a single genomic region in zebrafish. This region is expressed in zebrafish embryos as evaluated by RT-PCR at 24 h post fertilization (hpf). Because the regions of similarity are not syntenic between primates and other orders (supplemental Table 3) and might, therefore, have arisen independently, we would not expect the RNAs to affect expression of the same target genes (such as MNX1) in human and zebrafish. However, the zebrafish CAT7-like RNA (cat7l) might functionally interact with PRC1 based on its high sequence identity with human CAT7. We, therefore, disrupted zcat7l RNA in zebrafish embryos to study its role in a developing organism and its genetic interactions with bmi1a/b.

We tested the functional role of cat7l during zebrafish embryogenesis by morpholino interference of cat7l. Morpholino-mediated disruption was chosen over CRISPR/Cas9-mediated genomic mutations because small deletions are unlikely to affect IncRNA function; an individual key nucleotide(s) must be identified for deletion to impair function and may not exist in a repetitive domain. Moreover, phenotypes caused by deletion of DNA regulatory elements would be difficult to distinguish from those caused by impaired IncRNA function, confounding interpretation of large deletions. We, therefore, designed morpholinos that bind to cat7l with total complementarity at up to 11 locations across the zebrafish RNA, potentially blocking sites of protein interaction or interrupting RNA secondary structure (see Fig. 5A). We found that morpholinos targeting cat7l (cat7l MO), but not a control sense morpholino (cat7l sense MO control), caused microcephaly by 48 hpf (Fig. 4, A–E) and death starting on day 5. The microcephaly phenotype of cat7l knockdown was verified in p53-null fish (34, 35) using cat7l morpholinos suppress the microcephalic phenotype (36). If cat7l acts in the bmi1a/b pathway, interfering with the function of a known bmi1 target gene involved in microcephaly may also suppress the cat7l phenotype. Consistent with this reasoning, we found that co-injection of the ink4a morpholino with any of the three individual cat7l morpholinos suppress the microcephaly phenotype in both wild-type and p53 null fish (Fig. 4, E and F; Fig. 5, B and D). Furthermore, ink4a morpholinos could also suppress microcephaly caused by the synergistic effects of a low dose of bmi1a/b and cat7l morpholinos in wild-type and p53 null fish (Fig. 4, E and F; Fig. 5, B and D). Together, the ink4a suppression experiments, the synergistic cat7l/bmi1 disruptions, and the human RNA rescue experiments present orthogonal lines of evidence for a shared cat7l/bmi1a/b pathway. We conclude that cat7l genetically interacts with bmi1a/b during zebrafish development and functions upstream of the direct bmi1a/b target, ink4a. Our findings strengthen a functional association between CAT7/cat7l IncRNAs and PRC1 in humans and zebrafish.

**Discussion**

Our findings originate with a biochemical screen for RNAs, which co-precipitate with PRC1, and culminate with the identification of IncRNAs, which impact PcG biology in human neuronal development and vertebrate embryogenesis. We used a chromatin-based biochemical screen (GRIP) to isolate similar to its zebrafish analog, cat7l, in development. Although IncRNAs in zebrafish have only recently been annotated and are not well functionally or mechanistically described, our findings suggest that they may play a role in PcG biology as seen in plants and mammals.

The microcephalic phenotype of cat7l morphants is reminiscent of previously characterized bmi1a/b morphants (Fig. 4, D and E), although onset of the phenotype is earlier for bmi1a/b depletion, at 28 hpf (36). To test if cat7l and bmi1a/b genetically interact in zebrafish, we performed a triple knockdown experiment targeting bmi1a, bmi1b, and cat7 using morpholino doses titrated below the effective range for microcephaly (0.32 ng each for bmi1a/b and 0.2 ng for cat7l). By administering all three morpholinos at low dosages, we expected that a synergistic effect would arise if bmi1a/b and cat7l genetically interacted but would not arise if the pathways were independent. We saw that when low morpholino dosages were administered singly, a majority of the embryos appeared normal (24 and 12% show microcephaly in low cat7l and bmi1a/b, respectively). When low dosages were administered in combination, 80% of the zebrafish exhibited microcephaly by 48 hpf (Fig. 4, D and E). This synergistic effect emerged at a lower cumulative morpholino concentration than that required to produce a phenotype for each individual morpholino. This effect was also seen in p53 null fish (Fig. 5, B and D) but was not observed when cat7l sense control morpholinos were used in combination with the bmi1a/b morpholinos (data not shown). Together, these results suggest a genetic interaction exists between cat7l and bmi1a/b.

To characterize the intersection of the bmi1a/b and cat7l pathways, we disrupted the function of the classic bmi1a/b target, ink4a, shown previously to suppress the bmi1a/b morphant phenotype (36). If cat7l acts in the bmi1a/b pathway, interfering with the function of a known bmi1 target gene involved in microcephaly may also suppress the cat7l phenotype. Consistent with this reasoning, we found that co-injection of the ink4a morpholino with any of the three individual cat7l morpholinos suppresses the microcephaly phenotype in both wild-type and p53 null fish (Fig. 4, E and F; Fig. 5, B and D). Furthermore, ink4a morpholinos could also suppress microcephaly caused by the synergistic effects of a low dose of bmi1a/b and cat7l morpholinos in wild-type and p53 null fish (Fig. 4, E and F; Fig. 5, B and D). Together, the ink4a suppression experiments, the synergistic cat7l/bmi1 disruptions, and the human RNA rescue experiments present orthogonal lines of evidence for a shared cat7l/bmi1a/b pathway. We conclude that cat7l genetically interacts with bmi1a/b during zebrafish development and functions upstream of the direct bmi1a/b target, ink4a. Our findings strengthen a functional association between CAT7/cat7l IncRNAs and PRC1 in humans and zebrafish.
CAT7 and cat7l lncRNAs Tune PRC1 Function

lncRNAs that co-localize on chromatin with PRC1. The imperative for designing a screening method tailored for chromatin is based on PRC1/lncRNA biology; high subcomplex variability and the confounding issues generated by RNA binding to DNA binding domains have limited the utility of CLIP and direct binding assays. GRIP employs a cross-linked, CsCl gradient to biochemically enrich for interactions occurring between intact complexes and RNA as they occur on chromatin. The CsCl density gradient separates chromatin from free protein and free RNA (Fig. 1, B–H), and its high ionic strength is also able to ablate many nonspecific RNA/protein and protein-protein interactions. GRIP is generally applicable to a range of chromatin complexes but is particularly amenable to PRC1 family complexes because it is not known which PRC1 subunits and/or binding partners (if any) confer RNA binding specificity in vivo or in vitro. The cross-linked approach captures RNAs in the proximity of both compacting and ubiquitylating PRC1 family complexes. Our findings demonstrate the relevance of these RNAs, first identified via biochemical and molecular purification, to PcG biology.

A novel candidate from this screen, CAT7, impacts repression and PRC1 binding at the MNX1 locus during early neuronal differentiation. CAT7 was identified via its proximity and functional relationship to BMI1, and cat7l, a CAT7 analog, was identified because it spans a TRD with sequence similarity to BMI1a/b during zebrafish development. Although CAT7/cat7l are not syntenic and, therefore, likely evolved their TRDs separately, human CAT7 RNA can rescue phenotypes caused by knockdown of cat7l during zebrafish development, indicating convergence of function in addition to convergence of the TRD sequence. We do not presume that the independently evolving CAT7/cat7l lncRNAs play identical roles in human/zebrafish development as an lncRNA may exhibit multiple cellular functions, and furthermore, those functions may depend on the presence of a specific cellular state. Rather, we propose that in some cases, an excess of a similar PRC1-associated TRD can compensate for the loss of another. TRDs are thought to be important modulators in PcG/lncRNA biology; the unique TRDs REP-A and REP-C of the mammalian lncRNA XIST have been previously implicated in PRC2 binding and spreading (4, 37), and we also observed a distinct XIST repeat, REP-E, co-precipitating with PRC1 in our screen. The functional convergence of CAT7/cat7l suggests a model whereby analogous TRDs evolve rapidly and independently as a mechanism to regulate PRC1 activity during development.

Many mechanistic questions remain surrounding how lncRNAs may influence PRC1 function. First, it is not clear whether lncRNAs directly bind to PRC1 core components and/or to ancillary proteins that modulate PRC1 function, nor is it clear which component(s) or binding partners, if any, confer specificity. Our data show that PRC1 primarily co-precipitates with lncRNAs rather than with mRNA exons (as seen with previous PRC2 pulldown assays; Refs. 3, 4, 6, and 19), suggesting that a distinct RNA binding mechanism could exist at sites of PRC1-bound chromatin. It is possible that structural or sequence motifs, such as those found within the TRDs of CAT7/cat7l and XIST REP-E, directly bind PRC1 to modulate subunit selection, recruitment to a specific gene locus, or other PRC1 functions. The identification of CAT7/cat7l and other candidates in this screen demonstrates the ability of lncRNAs to modulate PRC1 binding or function across disparate organisms, underscores the translation from biochemical screens to live-animal biology, and provides a foundation for future studies characterizing the molecular mechanisms by which lncRNAs might impact PcG regulation.

Experimental Procedures

Primers, lncRNA Coordinates, and siRNA/Morpholino Sequences are Found in supplemental Tables 2 and 3.

HeLa Cell line, Culture, and Cross-linking—HeLa cells stably transduced with a copy of murine FLAG-tagged BMI1 (18) were cultured in DMEM supplemented with 1-glutamine (Sigma), 10% FBS (Sigma), NaHCO3 (3.1g/liter media), and gentamycin sulfate (7.5 pg/liter media). The cells express the ectopic protein at ~25% of wild-type levels, and all classical PRC1 components have been shown previously to co-precipitate with FLAG-BMI1 in this cell line (18). Cells were grown in suspension in a spinner flask (Matrical) to a density of ~3 × 10⁶ cells per liter. The cells were cross-linked at 1% formaldehyde for 20 hours in media at room temperature with light rocking.

Gradient and RNA Immunoprecipitation (Nuclei Prep, Sonication, and CsCl Gradient)—All steps were performed in the presence of RNAsIn plus (Promega) RNase inhibitor, DTT, 0.05 mM spermine, 0.05 mM spermidine, and Complete EDTA-free Protease inhibitor tablets (Roche Applied Science) according to the manufacturer’s instructions.

3 × 10⁶ HeLa Bmi1 F17 cells (18) were cross-linked in media as above; this provides sufficient input material for at least 10 immunoprecipitations. Cross-linked HeLa cells were resuspended in LB1 (50 mM Tris, pH 7.4, 140 mM KOAc, 1 mM EDTA, 0.5 mM EGTA, 10% glycerol, 0.5% Nonidet P-40, 0.25% Triton X-100, 0.1% digitonin) and spun to isolate nuclei and porate the nuclear membrane. The porated nuclei were then Dounce-homogenized in LB1 using a Wheaton Dounce homogenizer type A and applied over a glycerol pad (25% glycerol, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris, pH 7.4) to further separate cytoplasmic debris and expelled nucleoplasm. The nuclei were rinsed in LB2 (Tris, pH 7.4, 10 mM EDTA, 5 mM EGTA, and 200 mM KOAc), resuspended 1:1 in LB2, and sheared using the Covaris S2 machine (30 min at 20% Duty cycle, power 7, and 200 cycles/burst in 30-s intervals; 700-µl aliquots). The sheared nuclei were then resuspended (dropwise) in LB2 with 3.37 M CsCl and 0.1% sodium sarcosyl and

FIGURE 4. cat7l genetically interacts with bmi1a/b and cat7l interference is rescued by human CAT7 in zebrafish embryogenesis. cat7l is expressed in 24 hpf embryos (4) cat7l morpholino 1 (MO1) (8), but not the sense control, causes microcephaly at 48 hpf in zebrafish, and can be rescued with human CAT7 RNA (hCAT7). C, microcephaly resulting from cat7l MOs is rescued by hCAT7 RNA in p53-null fish at 48hpf. D and E, cat7l morphants phenocopy microcephaly caused by bmi1a/b depletion. Although low dosages of cat7l or bmi1a/b MOs do not exhibit a phenotype individually, in combination, they cause microcephaly. Microcephaly caused by cat7l MO1 and cat7l/bmi1a/b MOs can be rescued by the addition of MO targeting ink4aΔ116. * indicates affected phenotype.
FIGURE 5. cat7l morpholinos in p53 null and wild-type fish. A, MOs match poorly to human CAT7 (hCAT7) RNA. B and D, multiple cat7l MOs lead to microcephaly in p53-null fish and can be rescued with hCAT7 RNA or ink4ap16 MO. Some images are repeated from Fig. 4B to facilitate comparison. C, dose dependence of hCAT7 rescue in wild-type fish. B and D, low doses of cat7l MOs and bmi1a/b MO do not cause microcephaly individually. When administered together, the phenotype emerges, and can be rescued by ink4a/p16 MO. * indicates affected phenotype.
Gradient and RNA Immunoprecipitation (Sample Prep and Immunoprecipitation)—Fractions of the gradient containing chromatin were selected for IP based on $A_{260}$ and Western blot. Selected fractions were pooled and dialyzed (molecular weight cutoff 3500 kDa) for $>5$ h in LB2 containing 5% glycerol at 4 °C. In later experiments (e.g. gradient RNA IP and RT-qPCR experiments, but not the initial IP-seq experiments), 4 units of rDNase Turbo (Ambion) was added for 10 h at room temperature to reduce DNA shear size to 700 base pairs (and thereby increase resolution; data not shown). The rDNase was removed from the sample via its streptavidin tag according to the manufacturer’s instructions. After dialysis or rDNase treatment, 0.05% Nonidet P-40 and 0.5% Triton X-100 were added as well as 100 mM urea (or 1 M urea for immunoprecipitations against FLAG). The material was then spun out and moved to a new tube to ensure removal of any aggregates. <5% loss was checked by spectrophotometry ($A_{260}$), and samples were normalized in the same buffer to 800 ng/µl. The material was then precleared for 45 min with IgG-agarose beads (Sigma). Dynal protein A beads (Invitrogen) were prebound to antibody to avoid RNAses in the serum, and beads were blocked with protein A beads (Invitrogen) were prebound to antibody to avoid RNAses in the serum, and beads were blocked with protein A beads (Invitrogen). The prebound beads or covalently conjugated FLAG-agarose beads (Sigma) rinsed in LB2 were added to the input and incubated at 4 °C overnight with light rocking. The IP was carried out as published previously (21) with additional wash steps at room temperature in low salt, for increased RNA specificity. Specifically, beads were washed at room temperature 4 times in IP buffer and twice in IP buffer with reduced salt (25 mM KAc). The immunoprecipitated material was eluted from Dynal beads in 1% SDS and 1 mM EDTA or from M2-agarose beads with 3× FLAG peptide as described by Sigma. Cross-links were reversed for 1.5 h at 65 °C in the presence (RNA isolation) or absence (protein and DNA isolation) of 40 units of proteinase K (Roche Applied Science). For every IP, protein specificity was checked by Western and/or silver stain, and DNA binding specificity was also validated by ChIP-qPCR (data not shown).

Gradient RNA Immunoprecipitation (Analysis)—Single-read 50-nucleotide sequencing was performed on IP and input samples without multiplexing on Illumina’s HiSeq 2000. We removed possible contamination of Illumina sequencing adapters with cutadapt version 0.9.5 (38). We aligned uniquely mapping reads to the hg19 build of the genome using Bowtie 0.12.9. However, we noticed a NuGEN adapter sequence present in many read sequences, as has been seen previously when working with low-input samples (39). As a result, we trimmed 15 bases at the 5’ end of unaligned reads and attempted recovery by realignment. In total, we were able to uniquely align 49 million, 46.7 million, and 13.7 million reads for the input control, BMI1, and BMI1 FLAG-tagged IP samples, reflecting a unique alignment rate of 47%, 46.7%, and 13.1%. Low alignment was confirmed to be largely due to NuGEN adapter contamination. We then further only considered genomic regions that do not have a repetitive sequence by applying Repeatmasker (40). We were able to uniquely align 31.0 million, 29.6 million, and 7.6 million reads for the input, BMI1 IP, and FLAG-BMI1 IP RNAs. We then ran MACS v1.4 (41) on each IP sample without using input as reference respectively, with the $p$ value of the peak detection at $10^{-15}$ with the rest of the parameters taking default values. We found 39,796 intersecting MACS peaks between the two IPs before excluding peaks below background. We counted reads across 1) individual MACS peaks, 2) individual RefSeq mRNA exons (42), and 3) Ensembl lincRNAs (43). From each of these three sets of genomic regions, we normalized for transcript abundance in the input by using “intensity ratios”; the ratio of normalized read coverage of the IP/input. This was done in R-2.12.0, in which GenomicFeatures_1.2.3 was used for read counting, and edgeR_1.8.1 was used for comparing read coverage in genomic regions between samples. Total read counts were used to normalize the samples for edgeR, and we defined a background threshold for signal in the IP of 50 reads and 1 read in the input. Pearson correlations of intensity ratios were then run on intersecting regions of MACS peaks, defined as having $>1$ bp overlap, mRNA exons, or annotated lincRNAs (correlation coefficients calculated at intersecting regions of MACS peaks were very similar to those calculated using the entire corresponding BMI1 MACS peak coordinates instead). Significantly co-enriched peaks had a FDR < 0.01 and an intensity ratio > 4 in both IPs. Due to the low alignment rate caused by NuGEN oligo adapter contamination in low-nanogram samples, we could not add in an IgG control. We, therefore, validated 65 individual peaks/lincRNAs from this screen by large scale RT-qPCR in future experiments.

RNA Purification, cDNA Generation, and Library Prep—Either input RNA or RNA from the immunoprecipitation was stored in Trizol LS (Invitrogen) after cross-link reversal and proteinase K treatment. For native RNA prep, whole cells, nuclei, or cytosolic RNAs were stored in TRIzol (Invitrogen). In both cases, chloroform was added, and the sample was spun out according to the manufacturer’s instructions. The aqueous phase was applied to Zymo Clean-and-concentrator 5 columns, and DNase was applied “in tube” as per the manufacturer’s instructions for RNAs larger than 200 nucleotides. For RT-qPCR, cDNA was generated using SUPERscript VILO (Invitrogen). For total RNAseq, isolated RNA was ribo-depleted using the Ribo-Zero Magnetic Gold kit (Epicenter/Illumina) according to manufacturer’s instructions, and cDNA was generated with the TrueSeq kit (Illumina). For gradient RNA IP samples, cDNA was generated using the NuGEN RNAOvation FFPE kit (7150-08). Libraries for Illumina next-generation sequencing were constructed by blunt-end ligation of sequencing adapters, and PCR with indexed sequencing primers, as previously described (44).

DNA Isolation, ChIP, and qPCR—ChIP was performed as described previously (45, 46), or 10% of gradient RNA IP eluate was saved for DNA extraction was reverse-cross-linked as above. Gradient IP or ChIP eluate was then treated with DNase-free RNase (Roche Applied Science) followed by proteinase K (Roche Applied Science) and subjected to phenol chloroform isolation. The aqueous phase was then ethanol-precipitated with glycogen and NaCl. All qPCRs (from DNA or cDNA) were performed using Bio-Rad iTaq with SYBR and ROX. A list of all
CAT7 and cat7l IncRNAs Tune PRC1 Function

Sequencing was performed using Illumina HiSeq 2000 Illumina HiSeq 2500 or Illumina MiSeq. We aligned RNA-Seq data to the hg19 transcriptome using TopHat version 2.4. Gene level expression was characterized using the Genomic Features (Carlson R package version 1.12.4) and edgeR package in R 3.0.1 (49). For all “replicate 2” samples, CAT7 probe contamination during library construction was observed but was never seen in any other library batches and did not affect downstream analysis. RPKM value for a gene was calculated as the average RPKM of all annotated transcripts for this gene. For enrichment analysis, we computed probabilities and expected values for intersections with gene sets from the Molecular Signatures database sets c2 and c5 (6176 total gene sets; 28 PcG or H3K27me3 related gene sets) (20). The PcG-related gene lists are a compendium of localization data of PcG components/marks (ChIP data) and PcG mutant or knock-out data. Only mRNAs with RPKMs > 0.5, total counts ≥5 in all replicates/scrambles, and that were found in the Molsig database were considered in this analysis. To find the list of “up-regulated” genes for each knockdown, we considered genes with 2-fold or higher expression as compared with scramble-treated controls as assessed by edgeR. Initially, siRNAs targeting either BMI1 or SUZ12 were employed as assumed positive controls for perturbations to PcG-regulatory networks. However, only the mRNAs, but not the proteins, were depleted at 48 hours after knockdown (supplemental Table 3, Western blot data are not shown), and consequently very few target genes showed changed expression (supplemental Table 4). Instead, a hypergeometric distribution was used to calculate significance of the overlap of changed genes with 6176 curated gene sets from the Molecular signatures database considering no more than one isoform per mRNA. Twenty-eight of these gene sets are PcG component/H3K27me3 ChIP data sets or genes affected by PcG component knock-out. Corrections for multiple hypothesis testing are reflected in the FDR values for the 6176 gene sets. Replicated IncRNA knockdowns with an FDR <0.001 are considered strong candidates. “Overlap expected by chance” reports the mathematically defined expectation value. We considered only genes expressed above background (as described above), which were found in the Molecular Signatures database, and calculated the expected value as: # of changed genes/# of total genes × # of genes in PcG-list. The expectation is conceptually similar to a weighted average, and it represents the number of changed PcG targets expected to change by random chance. The RNAseq is considered a screen; any specific gene targets (particularly lowly expressed PcG targets) explored in the paper are precisely quantified using RT-qPCR with exon-spanning primers, often in >4 replicates and across several cell systems.

Zebrafish Injections and Rescue—Morpholinos (Gene Tools, supplemental Table 3) targeting cat7l, ink4a<sup>e16</sup>, bmi1a, and bmi1b were injected into one-cell-stage embryos of Tu wild-type or p53 null zebrafish (34) (supplemental Table 3). The bmi1a/b and ink4a<sup>e16</sup> morpholinos were described previously (36). In all bmi1 experiments, bmi1a and bmi1b morpholinos were co-injected. A morpholino specific for cat7l in the sense direction was used as a control. Embryos were screened at either 48 or 72 hpf and scored for microcephaly and classified as “affected” or “normal.” Scoring was performed independently by at least two investigators. The previously published phenotype from the bmi1a/b fish was used as a reference for categorizing affected and unaffected fish. A few fish in each experiment (including uninjected controls) showed severe developmental abnormalities commonly arising from variations in embryo quality or the injection technique (generally 0–10% in wild-type fish and p53 null fish, with the exception of...
CAT7 and cat7l IncRNAs Tune PRC1 Function

cat7l MO2 in p53 fish, which showed >10%). The penetrance of the microcephaly was consistent and in-line with previously seen experiments with MOs targeting PcG-pathway components (36). The optimal dose for producing a phenotype by single injection of cat7l MO was 0.4–0.5 ng. For the rescue experiment, in vitro transcribed RNA was produced from human CAT7 sequence cloned into the pcS2+ vector using the mMESSAGE mMACHINE kit (Ambion). The human RNA has six well spaced mismatches to the zebrafish morpholino in a single locus, which do not allow >5 bases of complementarity and which exceed the typical 4–5 mismatch of “negative control” morpholinos. One morpholino differed from the human CAT7 RNA sequence by 12 well spaced mismatches, representing a nearly 50% mismatch of the morpholino sequence. The optimal dose of mRNA for rescue was 100 pg. In the co-injection experiment, cat7l, bml1a, and bml1b morpholinos were injected separately and combined at 0.2 ng and 0.09 ng, respectively. inka4p16 morpholino was added for rescue at 4 ng, as described previously (36). Zebrafish were maintained as described (50), and all work was approved by the Massachusetts General Hospital Institutional Animal Care and Use Committee. Images were taken using Nikon software, and image brightness was adjusted in photoshop.

Immunobots and Silver Stain—Protein was either isolated from CsCl fractions by TCA precipitation and reverse-cross-linked in SDS as above or loaded from the cross-link reversal stage from the immunoprecipitations. A final concentration of 0.5 units/µl Benzonase (Novagen) was added to each sample with Laemmli buffer and incubated at room temperature for 10 min followed by 10 min of boiling before SDS-PAGE. Antibodies were used at 1/4000 (Bmi1, Pan-H3) or 1/1000 (CTCF, Suz12). Silver staining was performed with the SilverQuest kit (Invitrogen) according to the manufacturer’s instructions.

Native Cell Fractionation—Nuclear isolation was performed in native cells, as previously published (51), and checked by a hemocytometer (>97%) with trypan blue. Cytosolic extract was also preserved alongside matched whole cell extract. All extracts were immediately stored in TRIZol.

Northern Blot and RACE-PCR—Nuclei were isolated from native cells to >97% purity by trypan blue staining on a hemocytometer. Northern blot analysis was performed using RNA probes as in previously published protocols (52), with the exception that a Hybond-N+ membrane was used instead of nitrocellulose. 20 µg of RNA was used in each lane, and 18S (present in both the cytoplasm and as newly transcribed RNA in the nucleus) was also visualized in all lanes. RACE was performed using the Ambion RLM-RACE kit (AM1700) as per the manufacturer’s instructions except using a higher reaction temperature of 55 °C rather than 42 °C. The Northern probe sequence was AACAAGCGGGUGAAGGCGAAGACAGCAAAGG-AAGAUGUGCUGAAGGCAAGGGGAGUAAUUG-UUGCCGUUCAAAUCAAGAGUCGCCUUUC-GAAAGGGAGGAGGUGCGACUGACAGCGCCUUG-CGUGUUUCAGACGCCGUGUUUCAGGAA.

Analysis of CAT7 Sequence Analogs and Coding Potential—Analogos were identified by Lifterover (University of California, Santa Cruz) in selected species of the tandem repeat locus (~1.2 kb of a total 1.6 kb in CAT7). Sequence identity was defined by Clustal Omega using standard parameters. Flanking genes were provided for each species if annotated by RefSeq within 1 MB of the locus. Both cat7l and CAT7 RNAs were screened for low protein coding potential by PhyloCSF (53) (threshold = 0), CPC (54) (threshold = 0.364), and CPAT (30) (Threshold = 0.38) using standard thresholds.

Quantification of CAT7—We estimated CAT7 expression from 100K cells using RT-qPCR relative to a standard curve. We made the standard curve of 10-fold dilutions of full-length CAT7 dsDNA. Multiple melt curves appeared in low copy number dilutions; therefore, we had to extrapolate rather than interpolate the copy number. Using smaller/alternate amplicons as a standard also resulted in (different sets of) multiple melt-curves.

Author Contributions—M. K. R., R. E. K., J. L. G., O. W., and L. Z. designed the experiments. M. K. R. developed and executed GRIP/RNAi screens and performed all the cell culture and experimental work in HeLa cells and all ChIP and molecular RNA work in neural culture. O. W. cultured and differentiated hESC and performed RNAi in hESCs. M. J. K. and N. I. performed all zebrafish injections and rescues. A. E., M. K. R., and R. S. analyzed the RNAseq data and ChIPseq data. Y. W., M. L. B., and M. K. R. analyzed the gradient RNA IP-seq data. A. J. P., C. P. D., K. K., and K. E. contributed to discussions shaping the manuscript. C. P. D. screened CAT7 for coding potential. M. K. R. and R. E. K. wrote the manuscript.

Acknowledgments—We gratefully acknowledge Jesse Engreitz, Sara Miller, Aimee Deaton, Yevgeniya Li, and the Massachusetts General Hospital sequencing core.

References

1. Orlando, V. (2003) Polycomb, epigenomes, and control of cell identity. Cell 112, 599–606
2. Brockdorff, N. (2013) Noncoding RNA and polycomb recruitment. RNA 19, 429–442
3. Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Brugmann, S. A., Goodnough, L. H., Helms, J. A., Farnham, P. J., Segal, E., and Chang, H. Y. (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129, 1311–1323
4. Zhao, J., Sun, B. K., Erwin, J. A., Song, J.-J., and Lee, J. T. (2008) Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science 322, 750–756
5. Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K. J., Gooding, A. R., Lee, J. T., and Cech, T. R. (2015) Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. Mol. Cell 57, 552–558
6. Klattenhoff, C. A., Scheuermann, J. C., Surface, L. E., Bradley, R. K., Fields, P. A., Steinhauer, M. L., Ding, H., Baty, V. L., Torrey, L., Haas, S., Abo, R., Tabeboardar, M., Lee, R. T., Burge, C. B., and Boyer, L. A. (2013) Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. Cell 152, 570–583
7. Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., Kluger, Y., and Reinberg, D. (2012) PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol. Cell 45, 344–356
8. Maertens, G. N., El Messaoudi-Aubert, S., Racek, T., Stock, J. K., Nicholls, J., Rodriguez-Niedenführ, M., Gil, J., and Peters, G. (2009) Several distinct polycomb complexes regulate and co-localize on the INK4a tumor suppressor locus. PloS ONE 4, e6380
9. Simon, J. A., and Kingston, R. E. (2009) Mechanisms of polycomb gene silencing: knowns and unknowns. Nat. Rev. Mol. Cell Biol. 10, 697–708
10. Francis, N. J., Kingston, R. E., and Woodcock, C. L. (2004) Chromatin
compaction by a polycomb group protein complex. Science 306, 1574–1577.
11. Grau, D. J., Chapman, B. A., Garlick, J. D., Borowsky, M., Francis, N. J., and
Kingston, R. E. (2011) Compaction of chromatin by diverse Polycomb
group proteins requires localized regions of high charge. Genes Dev 25,
2210–2221.
12. Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E., and Allis, C. D.
(2006) Mouse polycomb proteins bind differentially to methylated histone
H3 and RNA and are enriched in facultative heterochromatin. Mol. Cell.
26, 2560–2569.
13. Simon, J. A., and Kingston, R. E. (2013) Occupying chromatin: polycomb
mechanisms for getting to genomic targets, stopping transcriptional traf-
fc, and staying put. Mol. Cell 49, 808–824.
14. Yap, K. L., Li, S., Muñoz-Cabello, A. M., Ragu, S., and Zeng, L. (2010)
Molecular interplay of the noncoding RNA ANRIL and methylated histi-
cone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of
INK4A. Mol. Cell 38, 662–674.
15. Yang, L., Lin, C., Liu, W., Zhang, J., Ohgi, K. A., Grinstein, J. D., Dorrestein,
P. C., and Rosenfeld, M. G. (2011) ncRNA- and Pc2 methylation-depen-
dent gene relocation between nuclear structures mediates gene activation
programs. Cell 147, 773–788.
16. Hu, X., Feng, Y., Zhang, D., Zhao, S. D., Hu, Z., and Greshock, J. (2014) A
functional genomic approach identifies FAL1 as an oncogenic long non-
coding RNA that associates with BMI1 and represses p21 expression in
cancer. Cancer Cell 26, 344–357.
17. Kaneko, S., Bonasio, R., Saldaña-Meyer, R., Yoshida, T., Son, J., Nishino,
K., Umezawa, A., and Reinberg, D. (2014) Interactions between JARID2
and noncoding RNAs regulate PRC2 recruitment to chromatin. Mol.
Cell 53, 290–300.
18. Levine, S. S, Weiss, A., Erdjument-Bromage, H., Shao, Z., Tempst, P., and
Kingston, R. E. (2002) The core of the polycomb repressive complex is
compositionally and functionally conserved in flies and humans. Mol.
Cell. Biol. 22, 6070–6078.
19. Kaneko, S., Son, J., Bonasio, R., Shen, S. S., and Reinberg, D. (2014) Nascent
RNA interaction keeps PRC2 activity poised and in check. Genes Dev.
28, 1983–1988.
20. Liberon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo,
P., and Mesirov, J. (2011) Molecular signatures database (MsigDB) 3.0. Bio-
informatics 27, 1739–1740.
21. Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee,
T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte,
A. P., Vidal, M., Gifford, D. K., Young, R. A., and Jaenisch, R. (2006) Poly-
comb complexes repress developmental regulators in murine embryonic
stem cells. Nature 441, 349–353.
22. Aziz, A., Liu, Q. C., and Dilworth, F. J. (2010) Regulating a master regula-
tor: establishing tissue-specific gene expression in skeletal muscle. Epige-
nesis 5, 691–695.
23. Nie, L., Wu, H. J., Hsu, J. M., Chang, S. S., Labaff, A. M., and Li, C. W.,
Wang, Y., Hsu, J. L., and Hung, M. C. (2012) Long non-coding RNAs:
versatile master regulators of gene expression and crucial players in can-
niferous xenografts. Mol. Cancer Ther. 11, 661–673.
24. Nagano, T., and Fraser, P. (2009) Emerging similarities in epigenetic gene
silencing by long noncoding RNAs. Mamm. Genome 20, 557–562.
25. Bonnefond, A., Vaillant, E., Philippe, J., and Skrobek, B. (2013) Transcrip-
tion factor gene MNX1 is a novel cause of permanent neonatal diabetes in
a consanguineous family. Eur. J. Hum. Genet. 19, 999–1005.
26. Anderson, E., Devenney, P. S., Hill, R. E., and Lettice, L. A. (2014) Mapping
the Ssh long-range regulatory domain. Development 141, 3934–3943.
27. van Aerssen, J., García-Hurtado, J., Moran, I., Maestro, M. A., Xu, X.,
Van de Casteele, M., Skoudy, A. L., Palassini, M., Heimberg, H., and
Ferrer, J. (2010) Derepression of polycomb targets during pancreatic organogen-
esis allows insulin-producing beta-cells to adopt a neural gene activity
program. Genome Res. 20, 722–732.
28. Chambers, S. M., Mica, Y., Lee, G., Studer, L., and Tomishima, M. I. (2016)
Dual-SMAD Inhibition/WNT activation-based methods to induce neural
crest and derivatives from human pluripotent stem cells. Methods Mol.
Biol. 1307, 329–343.
29. Wang, L., Park, H. J., Dasari, S., Wang, S., Koche, J. P., and Li, W. (2013)
CPAT: coding-potential assessment tool using an alignment-free logistic
regression model. Nucleic Acids Res. 41, e74.
30. He, W., Cai, Q., Sun, F., Zhong, G., Wang, P., Liu, H., Luo, J., Yu, H., Huang,
J., and Lin, T. (2013) linc-UBC1 physically associates with polycomb re-
pressive complex 2 (PRC2) and acts as a negative prognostic factor for
lymph node metastasis and survival in bladder cancer. Biochem. Biophys.
Acta 1832, 1528–1537.
31. Blackledge, N. P., Farcas, A. M., Kondo, T., and King, H. W. (2014) Variant
PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment
and polycomb domain formation. Cell 157, 1445–1459.
32. Pengelly, A. R., Kalb, R., Finkl, K., and Müller, J. (2015) Transcriptional
regression by PRC1 in the absence of H2A monoubiquitylation. Genes
Dev. 29, 1487–1492.
33. Berghmans, S., Murphey, R. D., Wienholds, E., Neuberg, D., Kutok, J. L.,
Fletcher, C. D., Morris, J. P., Liu, T. X., Schulte-Merker, S., Kanki, J. P.,
Plasterk, R., Zon, L. I., and Look, A. T. (2005) tp53 mutant zebrafish de-
velop malignant peripheral nerve sheath tumors. Proc. Natl. Acad.
Sci. U.S.A. 102, 407–412.
34. Berghmans, S., Veldhoen, M., Ploegh, H. L., and Neuberg, D. (2011) High-
resolution Xist binding maps reveal two-step spreading during
X-chromosome inactivation. Nature 504, 465–469.
35. Martin, M. (2011) Cutadapt removes adapter sequences from high-
throughput sequencing reads. EMBNet.journal, 17, 10–12.
36. Head, S. R., Komori, H. K., Hart, G. T., Shimashita, J., Schaffer, L., Salo-
mon, D. R., and Ordoukhian, P. T. (2011) Method for improved Illu-
mina sequencing library preparation using NuGEN Ovation RNA-Seq
System. Biotechniques 50, 177–180.
37. Yang, L., Lin, C., Liu, W., Zhang, J., Ohgi, K. A., Grinstein, J. D., Dorrestein,
P. C., and Rosenfeld, M. G. (2011) ncRNA- and Pc2 methylation-depen-
dent gene relocation between nuclear structures mediates gene activation
programs. Cell 147, 773–788.
38. Xu, X., Lu, X., Zhang, Q., Zhou, R., Ju, X., Tao, W., Liu, D., Deng, H., and
Lu, Z. (2010) Identification and characterization of Bmi-1-responding element within the human p16 promoter. J. Biol. Chem. 285, 33219–33229.
39. Wang, L., Park, H. J., Dasari, S., Wang, S., Koche, J. P., and Li, W. (2013)
CPAT: coding-potential assessment tool using an alignment-free logistic
regression model. Nucleic Acids Res. 41, e74.
40. Meng, S., Luo, M., Sun, H., Yu, X., Shen, M., Zhang, Q., Zhou, R., Ju, X.,
Tao, W., Liu, D., Deng, H., and Lu, Z. (2010) Identification and character-
ization of Bmi-1-responding element within the human p16 promoter.
J. Biol. Chem. 285, 33219–33229.
41. Meng, S., Luo, M., Sun, H., Yu, X., Shen, M., Zhang, Q., Zhou, R., Ju, X.,
Tao, W., Liu, D., Deng, H., and Lu, Z. (2010) Identification and character-
ization of Bmi-1-responding element within the human p16 promoter.
J. Biol. Chem. 285, 33219–33229.
42. Wichterle, H., Lieberam, I., Porter, J. A., and Jessell, T. M. (2002) Directed
differentiation of embryonic stem cells into motor neurons. Cell 110,
385–397.
CAT7 and cat7l IncRNAs Tune PRC1 Function

49. Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140

50. Westerfield, M. (2000) *The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio)*. University of Oregon Press, 4th Ed., Eugene, Oregon

51. Tolstorukov, M. Y., Goldman, J. A., Gilbert, C., Ogryzko, V., Kingston, R. E., and Park, P. I. (2012) Histone variant H2A.Bbd is associated with active transcription and mRNA processing in human cells. *Mol. Cell* **47**, 596–607

52. Ardehali, M. B., Yao, J., Adelman, K., Fuda, N. J., Petesch, S. J., Webb, W. W., and Lis, J. T. (2009) Spt6 enhances the elongation rate of RNA polymerase II in vivo. *EMBO J.* **28**, 1067–1077

53. Lin, M. F., Jungreis, L., and Kelis, M. (2011) PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* **27**, i275–i282

54. Kong, L., Zhang, Y., Ye, Z.-Q. Q., Liu, X.-Q. Q., Zhao, S.-Q. Q., Wei, L., and Gao, G. (2007) CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35**, W345–W349