LETTER

doi:10.1038/nature09819

A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression

Kevin C. Wang[1,2], Yul W. Yang[1]*, Bo Liu[3]*, Amartya Sanyal[4], Ryan Corces-Zimmerman[1], Yong Chen[5], Bryan R. Lajoie[4], Angeline Protacio[1], Ryan A. Flynn[1], Rajnish A. Gupta[1], Joanna Wysocka[6], Ming Lei[5], Job Dekker[4], Jill A. Helms[3] & Howard Y. Chang[1]

The genome is extensively transcribed into long intergenic noncoding RNAs (lincRNAs), many of which are implicated in gene silencing[1,2]. Potential roles of lincRNAs in gene activation are much less understood[3–5]. Development and homeostasis require coordinate regulation of neighbouring genes through a process termed locus control[6]. Some locus control elements and enhancers transcribe lincRNAs[7–10], hinting at possible roles in long-range control. In vertebrates, 39 *Hox* genes, encoding homeodomain transcription factors critical for positional identity, are clustered in four chromosomal loci; the *Hox* genes are expressed in nested anterior-posterior and proximal-distal patterns colinear with their genomic position from 3′ to 5′of the cluster[11]. Here we identify *HOTTIP*, a lincRNA transcribed from the 5′ tip of the *HOXA* locus that coordinates the activation of several 5′ *HOXA* genes *in vivo*. Chromosomal looping brings *HOTTIP* into close proximity to its target genes. *HOTTIP* RNA binds the adaptor protein WDR5 directly and targets WDR5/MLL complexes across *HOXA*, driving histone H3 lysine 4 trimethylation and gene transcription. Induced proximity is necessary and sufficient for *HOTTIP* RNA activation of its target genes. Thus, by serving as key intermediates that transmit information from higher order chromosomal looping into chromatin modifications, lincRNAs may organize chromatin domains to coordinate long-range gene activation.

We examined chromosome structure and histone modifications in human primary fibroblasts derived from several anatomic sites[12], and found distinctive differences in the *HOXA* locus. High throughput chromosome conformation capture (5C)[13] across *HOXA* revealed that its higher order structure is dependent on positional identity. In anatomically distal cells (for example, foreskin and foot fibroblasts), we detected abundant chromatin interactions within the transcriptionally active 5′ *HOXA* locus (with reference to the directions of transcription of constituent *Hox* genes), pointing to a compact and looped conformation. In contrast, no long-range chromatin interactions are detected within the transcriptionally silent 3′ *HOXA* which seems largely linear (Fig. 1a). Strikingly, anatomically proximal cells (for example, lung fibroblasts) have the diametrically opposite pattern. The ON and OFF states of *Hox* and other key developmental genes are maintained by the MLL/Trithorax (Trx) and polycomb group (PcG) proteins, which mediate trimethylation of histone H3 lysine 4 (H3K4me3) to activate genes or lysine 27 (H3K27me3) to repress genes[14]. The portions of *HOXA* in tight physical interaction are marked by broad domains of H3K4me3, whereas H3K27me3 marks the physically extended and transcriptional silent regions (Fig. 1a).

On the very 5′ and 3′ edges of the two respective interaction clusters are two lincRNA loci that exhibit distinct chromatin modifications. The 3′element has been previously identified as the myelopoiesis-associated lincRNA *HOTAIRM1* (ref. 15). The 5′ element, for which we suggest the name *HOTTIP* for '*HOXA* transcript at the distal tip', exhibits bivalent H3K4me3 and H3K27me3, a histone modification pattern associated with poised regulatory sequences[16]. Comparison with RNA polymerase II occupancy and RNA expression showed that the bivalent H3K4me3 and H3K27me3 modifications on *HOTTIP* gene do not require *HOTTIP* transcription, but transcription of *HOTTIP* is associated with increased H3K4me3 and decreased H3K27me3 (Fig. 1a, left). Chromatin immunoprecipitation (ChIP) analysis confirmed that the *HOTTIP* gene is occupied by both polycomb repressive complex 2 (PRC2) and MLL complex, consistent with the bivalent histone marks (Supplementary Fig. 1a).

HOTTIP transcription yields a 3,764-nucleotide, spliced and polyadenylated lincRNA that initiates ~330 bases upstream of *HOXA13*. Only the strand antisense to *HOXA* genes is transcribed (Supplementary Fig. 1b). Genes near the 5′ end of each *HOX* cluster tend to be expressed in more posterior and/or distal anatomical locations. Consistent with its genomic location 5′ to *HOXA13*, *HOTTIP* is expressed in distal and/or posterior anatomic sites (Fig. 1b). *In situ* hybridization of developing mouse and chick embryos confirmed that *HOTTIP* is expressed in posterior and distal sites *in vivo*, indicating a conserved expression pattern from development to adulthood (Fig. 1c and Supplementary Fig. 1c). Even in distal cells where *HOTTIP* is expressed, its RNA level is very low and estimated to be ~0.3 copies per cell (Supplementary Fig. 2).

We employed small interfering RNAs (siRNAs) to knock down *HOTTIP* RNA in fibroblasts from a distal anatomic site (foreskin), and examined expression of 5′ *HOXA* genes by quantitative reverse transcription PCR. Notably, *HOTTIP* RNA knockdown abrogated expression of distal *HOXA* genes across 40 kilobases with a trend dependent on the distance to *HOTTIP*. The strongest blockade was observed for *HOXA13* and *HOXA11*, with progressively less severe effects on *HOXA10*, *HOXA9* and *HOXA7* (Fig. 2a). The effect on gene transcription appeared to be unidirectional, as there were no appreciable changes in the levels of *EVX1*, located ~40 kilobases 5′ of the *HOXA* cluster (data not shown). *HOTTIP* knockdown did not affect expression of the highly homologous *HOXD* genes, other control genes, nor induce antisense transcription at its own locus (Fig. 2b, Supplementary Fig. 3a). Several independent siRNAs targeting *HOTTIP* yielded similar results (Supplementary Fig. 3b). These results indicate that *HOTTIP* RNA is necessary to coordinate activation of 5′ *HOXA* genes.

We next addressed the function of *HOTTIP* RNA *in vivo* in the developing chick limb bud (Fig. 2c). Whereas prior genetic studies of noncoding RNAs (ncRNAs) involved deletion or insertion into the gene locus[17], we wished to distinguish the functions of *HOTTIP* RNA from its corresponding DNA element. *HOTTIP* gene can nucleate H3K4 and H3K27 methylation independent of transcription (Fig. 1a),

[1]Howard Hughes Medical Institute, Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California 94305, USA. [2]Department of Dermatology, University of California San Francisco (UCSF), San Francisco, California 94115, USA. [3]Department of Surgery, Stanford University School of Medicine, Stanford, California 94305, USA. [4]Program in Gene Function and Expression, Dept. of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. [5]Howard Hughes Medical Institute, Department of Biological Chemistry, University of Michigan Medical School, Ann Arbor, Michigan 48109, USA. [6]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, California 94305.
*These authors contributed equally to this work.

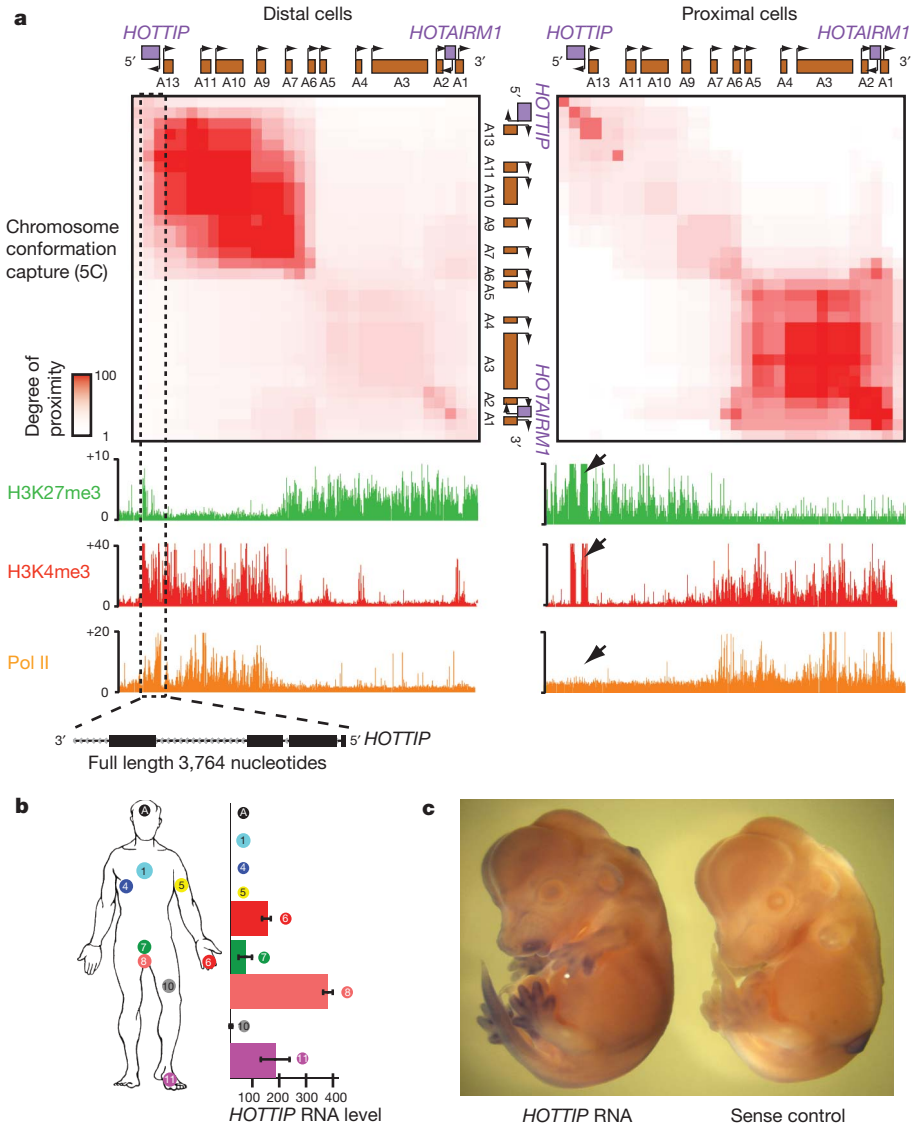

Figure 1 | *HOTTIP* is a lincRNA transcribed in distal anatomic sites. **a**, Chromatin state map of distal versus proximal cells. Top panels, chromosome conformation capture-carbon copy (5C) analysis of distal (foreskin) and proximal (lung) human fibroblasts. Heat map representations (generated by my5C, ref. 30) of 5C data (bin size 30 kb, step size 3 kb) for *HOXA* in foreskin and lung fibroblasts. Red intensity of each pixel indicates relative interaction between the two points on the genomic coordinates. The diagonal represents frequent *cis* interactions between regions located in close proximity along the linear genome. 5C signals that are away from the diagonal represent long-range looping interactions. Bottom panels, chromatin occupancy across *HOXA*. *x*-axis is genomic coordinate; *y*-axis depicts occupancy of the indicated histone marks or protein (ChIP/input). Box and arrows highlight chromatin states of *HOTTIP* gene. **b**, *HOTTIP* RNA expression in primary human fibroblasts from 11 anatomic sites. Means ± s.d. are shown ($n = 2$). **c**, *In situ* hybridization of *HOTTIP* RNA in E13.5 mouse embryo.

and the precise genomic distance between upstream enhancer elements and *Hox* genes is critical for their proper colinear activation[17]. Therefore, we used RNA interference (RNAi) in chick embryos, where replication-competent retroviruses can deliver short-hairpin RNAs (shRNAs) with high penetrance and precise spatiotemporal control[18] (Supplementary Fig. 4). In the limb bud, 5′ *HoxA* genes are transcribed in a nested pattern along the proximal-distal axis[19]. In this tissue, *HoxA* function is highly redundant with that of the *HoxD* locus, which allowed us to assess altered *HoxA* expression patterns without major changes in anatomic landmarks[20]. We injected retroviruses carrying shRNAs against chick *HOTTIP* into upper limb buds of stage 13 chicks; RT–PCR and *in situ* hybridization were performed on both control and knockdown samples after 2–4 days. Knockdown of *HOTTIP* RNA by two independent shRNAs in limb buds decreased expression of *HoxA13*, *HoxA11* and *HoxA10*—again with a graded impact depending on genomic proximity to *HOTTIP* gene. Vector control or an shRNA that fails to deplete *HOTTIP* RNA had little effect on *Hox* gene expression (Fig. 2d). *In situ* hybridization on whole embryos (Fig. 2e) and sections (Supplementary Fig. 5) revealed that *HOTTIP* RNA most strongly affects *HoxA* gene expression at the distal edge of the developing limb bud, where the 5′ *HoxA* genes are most strongly expressed. By stage 36, limbs depleted of *HOTTIP* RNA showed notable shortening and bending of distal bony elements, including the radius, ulna and third digit (~20% length reduction for each compared to contralateral and stage-matched limbs treated with control virus, $P < 0.05$, Student's *t*-test, Fig. 2f). This phenotype resembled some of the defects in mice lacking *HoxA11* and *HoxA13* (refs 21–23). Together, these data indicate that *HOTTIP* RNA controls activation of distal *Hox* genes *in vivo*.

The broad impact of *HOTTIP* RNA on gene activation across the *HOXA* locus is reminiscent of the broad domains of chromatin modifications demarcating active and silent chromosomal domains[12]. 5C analysis of control and *HOTTIP*-depleted cells showed little change in higher order chromosomal structure, indicating that the chromosomal looping is pre-configured and upstream of gene expression

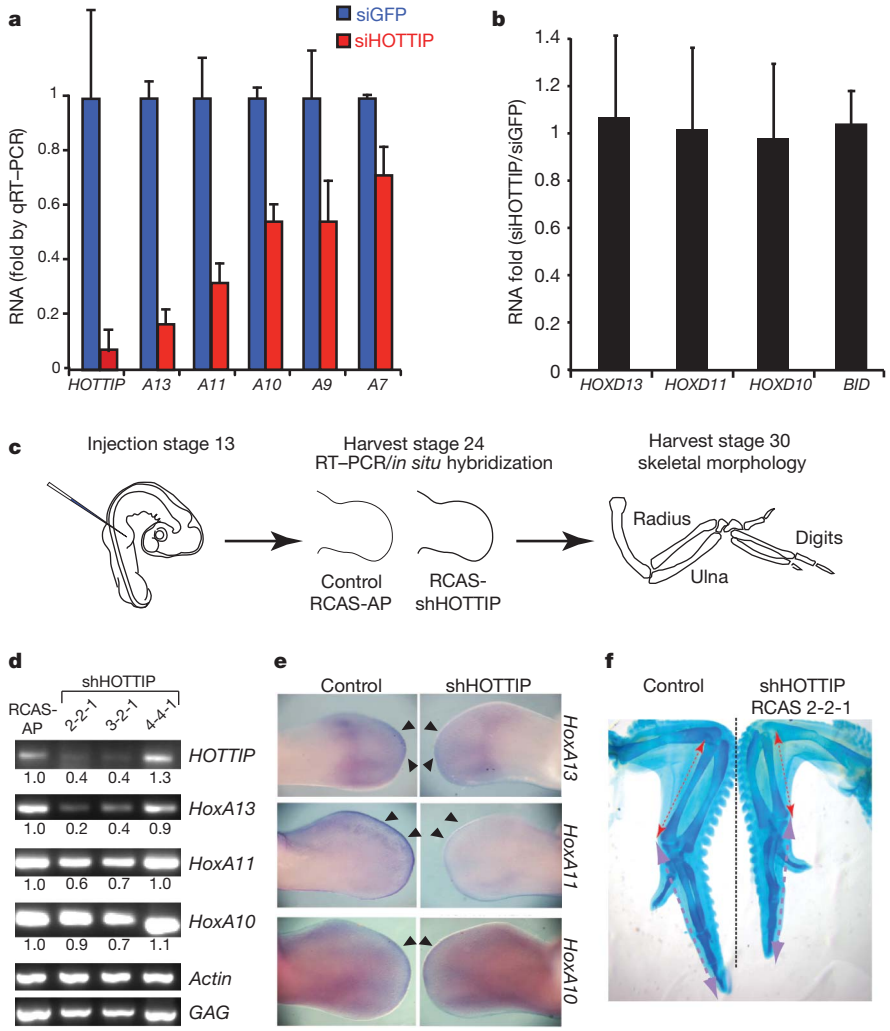

Figure 2 | *HOTTIP* is required for coordinate activation of 5′ *HOXA* genes. **a**, **b**, Knockdown of *HOTTIP* RNA abrogates expression of 5′ *HOXA* genes in foreskin fibroblasts (**a**), but not *HOXD* or *BID* genes (**b**). Means + s.d. are shown ($n = 3$). GFP, green fluorescent protein. **c**, Schematic of chick RNAi experiment. **d**, *HOTTIP* RNA is required for 5′ *HoxA* gene expression *in vivo*. RT–PCR of the indicated genes from control or *HOTTIP*-depleted distal limb bud is shown; quantification and normalization by *Actin* signal is shown below each band. *GAG* signal confirms successful retroviral transduction in all cases. **e**, *In situ* hybridization of 5′ *HoxA* genes in chick limb buds. Arrowheads highlight distal domains of high *HoxA* gene expression that are affected by *HOTTIP* knockdown. **f**, Shortening of distal bony elements in *HOTTIP*-depleted forelimbs. Alcian blue staining highlights the skeletal elements. Red and purple lines highlight radius and 3rd digit lengths, respectively.

(Supplementary Fig. 6a). In contrast, *HOTTIP* RNA knockdown led to broad loss of H3K4me3 and H3K4me2 across the *HOXA* locus, most prominently over 5′ *HOXA* and *HOTTIP* gene itself (Fig. 3a, Supplementary Figs 6b and 7). *HOTTIP* RNA knockdown also increased H3K27me3 focally over *HOTTIP* gene, but had little impact on H3K27me3 across *HOXA*. These results indicate that *HOTTIP* RNA is required for maintenance of H3K4me3 across the *HOXA*. These findings also imply that loss of 5′ *HOXA* gene transcription upon *HOTTIP* RNA knockdown is likely to be due to loss of H3K4me3 (or other changes) rather than ectopic spread of H3K27me3.

H3K4 methylation of the *HOX* loci is carried out by the MLL family of complexes[24]. In mammals, at least six MLL family members of SET-domain-containing lysine methyltransferases interact with a core complex of WDR5, ASH2L, RBBP5, as well as with other proteins, for substrate recognition and genomic targeting[24]. Genetic analyses indicate that MLL1 and 2 are most essential for *HOX* gene expression in fibroblasts[25], and MLL1 in particular is recruited to promoters of *HOX* genes to maintain their activation states[26]. In distally-derived human fibroblasts, MLL1 and WDR5 densely occupied extended region of the 5′ *HOXA* cluster, coincident with the H3K4me3 domain, with specific 'peaks' of occupancy near the transcriptional start sites (TSS) of multiple 5′ *HOXA* genes (Fig. 3b). Strikingly, *HOTTIP* RNA knockdown abrogated the peaks of MLL1 and WDR5 occupancy near TSS, resulting in diffuse and less intense binding of MLL1 and WDR5 across *HOXA* cluster, most prominently over the 5′ *HOXA* domain. *HOTTIP* RNA knockdown also led to increased accumulation of MLL1 and WDR5 on *HOTTIP* gene itself (Supplementary Fig. 8). Thus, *HOTTIP* RNA seems critical for maintaining a specific pattern of MLL complex occupancy across the *HOXA* locus to facilitate H3K4me3 and active transcription.

To define the molecular link between *HOTTIP* RNA and MLL complex, we reasoned that *HOTTIP* RNA may physically interact with one or more subunits of the MLL complex. Purified, *in-vitro*-transcribed, full-length *HOTTIP* RNA bound specifically to recombinant glutathione-*S*-transferase-conjugated WDR5 (GST–WDR5), but not to GST, RBBP5, ASH2L, or the telomeric protein TRF1 (also known as TERF1; Fig. 4a, b). The C terminus of MLL1, containing the SET domain, bound non-specifically to all RNAs, consistent with previous studies[27]. Immunoprecipitation of endogenous WDR5 from two different cell lines each specifically retrieved endogenous *HOTTIP* RNA (Fig. 4c), indicating that WDR5 and *HOTTIP* RNA interact in living

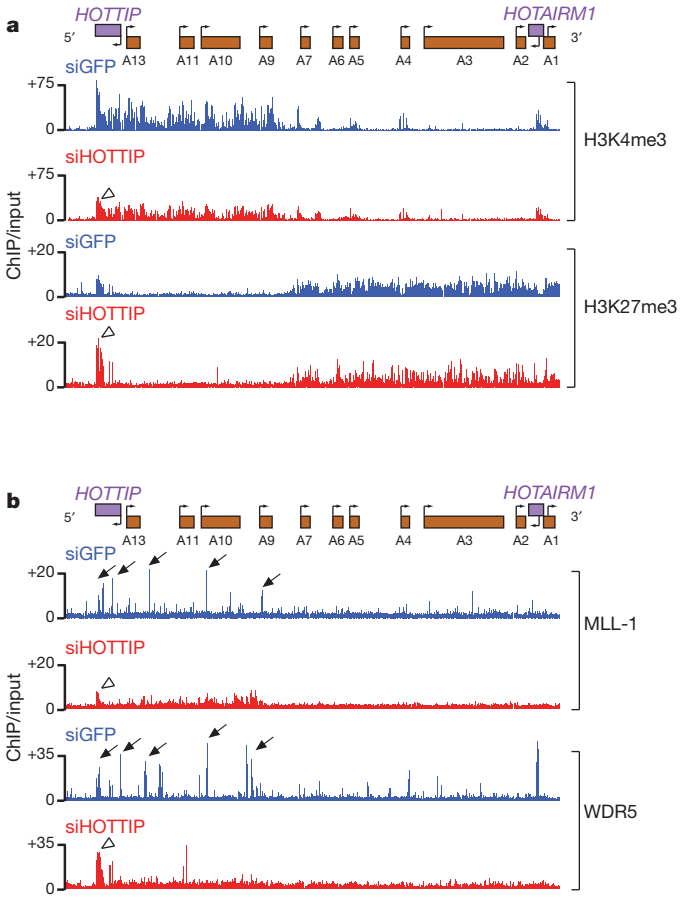

Figure 3 | *HOTTIP* RNA is required for the active chromatin state of 5′ *HOXA* cluster. **a**, Knockdown of *HOTTIP* RNA broadly decreases H3K4me3 across 5′ *HOXA* locus but focally affects H3K27me3 at *HOTTIP* gene. Display is as in Fig. 1a. **b**, Knockdown of *HOTTIP* RNA abrogates peaks of MLL1 and WDR5 occupancy near TSSs of 5′ *HOXA* genes and leads to accumulation of these proteins at *HOTTIP* gene itself. Arrows highlight peaks of MLL1 and WDR5 occupancy; open arrowheads highlight chromatin state of *HOTTIP* gene upon *HOTTIP* RNA knockdown.

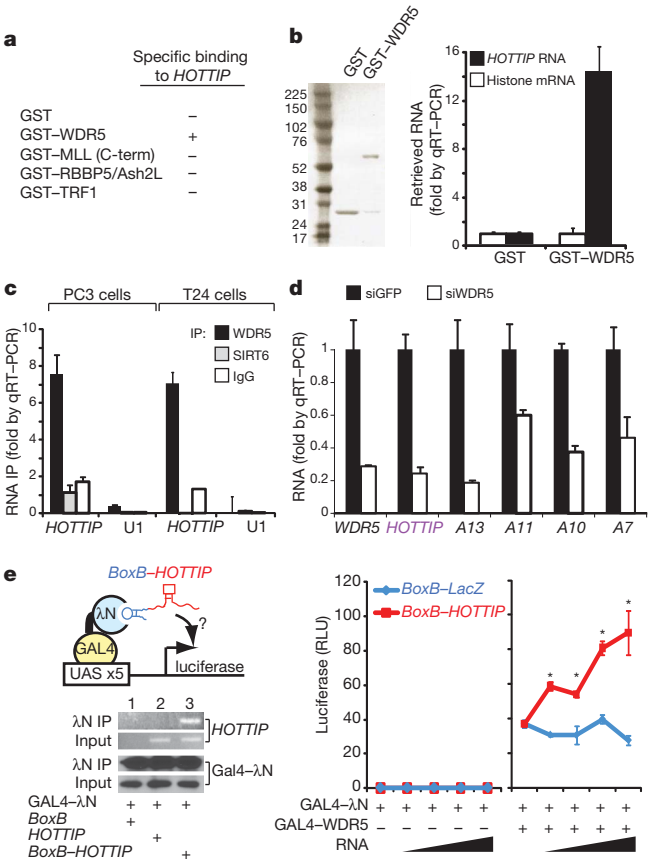

Figure 4 | *HOTTIP* RNA programs active chromatin via WDR5. **a**, Summary of RNA–protein interaction studies. Each of the indicated recombinant protein was purified and used to retrieve purified *HOTTIP* RNA or control histone RNA *in vitro*. Only GST–WDR5 specifically retrieved *HOTTIP*. **b**, *HOTTIP* RNA binds directly and specifically to WDR5. Left, purified GST and GST–WDR5 are visualized by SDS–PAGE and Coomassie Blue staining. Right, retrieved RNAs are quantified by qRT–PCR. **c**, *HOTTIP* RNA binds specifically to WDR5 in cells. Immunoprecipitation (IP) of endogenous WDR5 protein from PC3 (prostate) and T24 (bladder) carcinoma cells specifically retrieved *HOTTIP*, but not control IPs with IgG or chromatin binder SIRT6. U1 spliceosomal RNA served as negative control. **d**, WDR5 is required for 5′ *HOXA* gene expression, including *HOTTIP* RNA. **e**, *HOTTIP* RNA recruitment potentiates transcription. Left, the *BoxB* tethering system. *BoxB*–RNA specifically binds λN fused to GAL4 DNA binding domain, recruiting the complex to a UAS-luciferase reporter gene. After transient transfection, IP of GAL4- λN specifically retrieves *BoxB–HOTTIP* RNA. Right, luciferase activity after co-transfection of the indicated constructs. $*P < 0.05$ Student's *t*-test comparing *BoxB–LacZ* versus *BoxB–HOTTIP*). Sloped triangle indicates increasing input of plasmids encoding ncRNAs. Means ± s.d. ($n = 3$) are shown for all panels.

cells. Immunoprecipitation of an epitope-tagged WDR5 from a stable cell line that previously enabled stoichiometric purification of WDR5-interacting proteins[28] also specifically retrieved *HOTTIP* RNA (Supplementary Fig. 9). Knockdown of WDR5 broadly inhibited expression of 5′ *HOXA* genes, and also abrogated *HOTTIP* transcription, demonstrating mutual interdependence between *HOTTIP* RNA and WDR5 (Fig. 4d).

HOTTIP RNA seems to regulate genes *in cis*, due to its low copy number, distance dependence of *HOXA* target gene activation on endogenous *HOTTIP*, and the physical proximity of *HOTTIP* and its target genes as seen in 5C. Indeed, ectopic expression of *HOTTIP* RNA by retroviral transduction of lung fibroblasts, which do not express *HOTTIP*, failed to activate expression of distal *HOXA* genes, and did not change H3K4me3 and H3K27me3 patterns across *HOXA* (Supplementary Fig 10). Moreover, in foreskin fibroblasts that express endogenous *HOTTIP*, ectopic *HOTTIP* expression did not induce 5′ *HOXA* genes, nor rescue the effects of depleting endogenous nascent *HOTTIP* RNA (Supplementary Fig. 11). The lack of response in foreskin fibroblasts is notable because endogenous *HOTTIP* RNA is active in these cells, indicating that the protein partners of *HOTTIP* are all present and target genes are receptive. Ectopically expressed *HOTTIP* RNA, being transcribed from retroviral insertion sites scattered randomly in the genome, may not be able to find 5′*HOXA* genes. In contrast, endogenous *HOTTIP* RNA is directly positioned near the 5′ *HOXA* genes by chromosomal looping, allowing interaction and control.

To test the requirement of an exogenous targeting mechanism, we engineered an allele of *HOTTIP* RNA that can be artificially recruited to a reporter gene. Addition of five copies of the *BoxB* RNA element[29] to *HOTTIP* RNA allows the fusion transcript to be recruited to the λN RNA binding domain fused to a GAL4 DNA-binding domain (Fig. 4e). Recruitment of *HOTTIP* RNA to a silent GAL4 promoter is not sufficient to initiate transcription, but can significantly boost transcription if the promoter is also bound by WDR5 and transcriptionally active (Fig. 4e). By uncoupling the sites of *HOTTIP* transcription versus *HOTTIP* RNA function, this experiment indicates that the proximity of *HOTTIP* RNA—rather than the act of transcription—maintains target gene expression. To further support the functionality of *HOTTIP* RNA, deletion analysis identified a ~1 kb domain in the 5′ of *HOTTIP* RNA (*HOTTIP*$^{\text{Exons 1–2}}$) that retains WDR5 binding activity (Supplementary Fig. 12a). Enforced overexpression of *HOTTIP*$^{\text{Exons 1–2}}$

in foreskin fibroblasts inhibited 5′ *HOXA* gene expression in an apparently dominant negative manner (Supplementary Fig. 12b).

In summary, *HOTTIP* RNA is a key locus control element of *HOXA* genes and distal identity. Chromosomal looping brings *HOTTIP* RNA in close proximity to the 5′ *HOXA* genes. *HOTTIP* transcription acts as a switch to produce *HOTTIP* lincRNA, which binds to and targets WDR5–MLL complexes to the 5′ *HOXA* locus, yielding a broad domain of H3K4me3 and transcription activation (Supplementary Fig. 13). The mutual interdependence between *HOTTIP* RNA and WDR5 creates a positive feedback loop that maintains the ON state of the locus. These findings provide an integrated view linking three dimensional genome organization to dynamic programming of chromatin states, and ultimately to developmental pattern formation.

H3K4 methylation is a feature of almost all transcribed genes, and MLL family proteins are involved in many cell fate decisions in development and disease[24]. Our findings suggest that additional lincRNAs, especially those associated with enhancers or enhancer-like activities[8–10], may also be involved in gene activation by programming active chromatin states, and highlight WDR5 and other WD40 repeat proteins as candidate adaptors that link chromatin remodelling complexes to lincRNAs. *Cis*-restricted lincRNAs may be ideally suited to link chromosome structure and gene expression. Because such lincRNA can only act on its neighbours in space, information in higher order chromosomal looping can be faithfully transmitted to chromatin modification via RNA recruitment of enzymatic activities, and thus into gene expression.

METHODS SUMMARY

High throughput chromosome confirmation capture (5C) was performed on foreskin and lung fibroblasts, as well as foreskin fibroblasts treated with control or siRNA against *HOTTIP* RNA, as described[13]. siRNA knockdown experiments on cultured human fibroblasts and qPCR were performed as described previously[12]. ChIP-chip was performed as described[12] using ultra-high-density *HOX* tiling arrays. Full-length *HOTTIP* RNA was cloned by 5′ and 3′ rapid amplification of cloned/cDNA ends (RACE). Single-molecule RNA-fluorescent in situ hybridization (FISH) was performed using a pool of fluorescently-labelled oligonucleotides specific to *HOTTIP* RNA. *In vivo HOTTIP* RNA knockdown in chick was accomplished by microinjecting retroviruses carrying shRNA into prospective wing and leg buds and the animals were harvested at 2, 4 and 9 weeks post injection for RNA *in situ* hybridization, immunohistochemistry and whole-mount limb analysis, respectively. RNA-immunoprecipitation with WDR5 was performed as described[12]. Tethering experiments were done in 293T cells with co-transfections of various constructs containing a upstream activating sequence (UAS)-luciferase reporter, GAL4-WDR5, *BoxB* alone, and *BoxB* fused to full-length *HOTTIP* or full-length *LacZ*; cells were lysed 48 h after transfection and luciferase activity was determined.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 February 2010; accepted 12 January 2011.

Published online 20 March; corrected 7 April 2011 (see full-text HTML version for details).

1. Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nature Rev. Genet.* **10,** 155–159 (2009).
2. Ponting, C. P., Oliver, P. L. & Reik, W. Evolution and functions of long noncoding RNAs. *Cell* **136,** 629–641 (2009).
3. Sanchez-Elsner, T., Gou, D., Kremmer, E. & Sauer, F. Noncoding RNAs of trithorax response elements recruit *Drosophila* Ash1 to *Ultrabithorax*. *Science* **311,** 1118–1123 (2006).
4. Petruk, S. *et al.* Transcription of *bxd* noncoding RNAs promoted by trithorax represses *Ubx* in *cis* by transcriptional interference. *Cell* **127,** 1209–1221 (2006).
5. Dinger, M. E. *et al.* Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* **18,** 1433–1445 (2008).
6. Dean, A. On a chromosome far, far away: LCRs and gene expression. *Trends Genet.* **22,** 38–45 (2006).
7. Ashe, H. L., Monks, J., Wijgerde, M., Fraser, P. & Proudfoot, N. J. Intergenic transcription and transinduction of the human β-globin locus. *Genes Dev.* **11,** 2494–2509 (1997).
8. De Santa, F. *et al.* A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biol.* **8,** e1000384 (2010).
9. Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465,** 182–187 (2010).
10. Ørom, U. A. *et al.* Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143,** 46–58 (2010).
11. Chang, H. Y. Anatomic demarcation of cells: genes to patterns. *Science* **326,** 1206–1207 (2009).
12. 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).
13. Dostie, J. *et al.* Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16,** 1299–1309 (2006).
14. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B. & Cavalli, G. Genome regulation by polycomb and trithorax proteins. *Cell* **128,** 735–745 (2007).
15. Zhang, X. *et al.* A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. *Blood* **113,** 2526–2534 (2009).
16. Bernstein, B. E. *et al.* A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315–326 (2006).
17. Kmita, M., Fraudeau, N., Herault, Y. & Duboule, D. Serial deletions and duplications suggest a mechanism for the collinearity of *Hoxd* genes in limbs. *Nature* **420,** 145–150 (2002).
18. Harpavat, S. & Cepko, C. L. RCAS-RNAi: a loss-of-function method for the developing chick retina. *BMC Dev. Biol.* **6,** 2 (2006).
19. Nelson, C. E. *et al.* Analysis of *Hox* gene expression in the chick limb bud. *Development* **122,** 1449–1466 (1996).
20. Kmita, M. *et al.* Early developmental arrest of mammalian limbs lacking *HoxA/HoxD* gene function. *Nature* **435,** 1113–1116 (2005).
21. Small, K. M. & Potter, S. S. Homeotic transformations and limb defects in Hox A11 mutant mice. *Genes Dev.* **7,** 2318–2328 (1993).
22. Davis, A. P., Witte, D. P., Hsieh-Li, H. M., Potter, S. S. & Capecchi, M. R. Absence of radius and ulna in mice lacking *hoxa-11* and *hoxd-11*. *Nature* **375,** 791–795 (1995).
23. Fromental-Ramain, C. *et al. Hoxa-13* and *Hoxd-13* play a crucial role in the patterning of the limb autopod. *Development* **122,** 2997–3011 (1996).
24. Ruthenburg, A. J., Allis, C. D. & Wysocka, J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell* **25,** 15–30 (2007).
25. Wang, P. *et al.* Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. *Mol. Cell. Biol.* **29,** 6074–6085 (2009).
26. Guenther, M. G. *et al.* Global and Hox-specific roles for the MLL1 methyltransferase. *Proc. Natl Acad. Sci. USA* **102,** 8603–8608 (2005).
27. Krajewski, W. A., Nakamura, T., Mazo, A. & Canaani, E. A motif within SET-domain proteins binds single-stranded nucleic acids and transcribed and supercoiled DNAs and can interfere with assembly of nucleosomes. *Mol. Cell. Biol.* **25,** 1891–1899 (2005).
28. Wysocka, J. *et al.* WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell* **121,** 859–872 (2005).
29. Baron-Benhamou, J., Gehring, N. H., Kulozik, A. E. & Hentze, M. W. Using the λN peptide to tether proteins to RNAs. *Methods Mol. Biol.* **257,** 135–154 (2004).
30. Lajoie, B. R., van Berkum, N. L., Sanyal, A. & Dekker, J. My5C: web tools for chromosome conformation capture studies. *Nature Methods* **6,** 690–691 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank C. Tabin for chick *Hox* gene probes, M. Scott and members of our labs for input, and M. Lin for use of the confocal microscope and imaging expertise. Supported by grants from the California Institute for Regenerative Medicine (H.Y.C., J.W.), the National Institutes of Health (HG003143 to J.D.), and the Scleroderma Research Foundation (H.Y.C.). K.C.W. is a recipient of a Dermatology Foundation Career Development Award. J.D. is a recipient of the W. M. Keck Foundation Distinguished Young Scholar Award. H.Y.C. and M.L. are Early Career Scientists of the Howard Hughes Medical Institute.

Author Contributions K.C.W., R.A.G. and H.Y.C. initiated the project; K.C.W. and H.Y.C. designed the experiments; K.C.W., Y.W.Y., B.L., A.S., R.C.-Z., B.R.L., A.P., R.A.F., J.D. and J.A.H. conducted the experiments and analysed the data; Y.C. and M.L. purified the recombinant proteins; J.W. provided antibodies and cell lines; K.C.W. and H.Y.C. prepared the manuscript with inputs from all co-authors.

Author Information Sequence for human *HOTTIP* RNA has been deposited with GenBank under the accession number GU724873. Microarray data are deposited in Gene Expression Omnibus (GEO) under accession number GSE26540. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.Y.C. (howchang@stanford.edu).

METHODS

Cells. Primary human fibroblasts derived from different anatomic sites were as described[12,31–36]. Primary human fibroblasts in culture retain their positional identity and have been used to examine chromatin states associated with positional memory, which have been confirmed *in vivo*[33,34,37].

Chromatin immunoprecipitation followed by microarray analysis. ChIP-chip was performed using anti-H3K27me3 (Abcam), anti-H3K4me3 (Abcam), anti-H3K4me2 (Abcam), anti-histone H3 (Abcam), anti-PolII (Abcam), anti-MLL1 (gift of R. Roeder), and anti-WDR5[28] antibodies as previously described[12]. Chromatin from each indicated cell type or RNAi treatment is split into multiple tubes and subject to ChIP with different antibodies in parallel. Retrieved DNA and input chromatin were competitively hybridized to custom tiling arrays interrogating human *HOX* loci at 5-bp resolution as previously described[12].

5C analysis of the ENm010 HoxA1 region. 5C primers were designed at HindIII restriction sites using 5C primer design tools previously developed[13] and made available online at http://my5C.umassmed.edu (ref. 30). Reverse primers were designed for fragments overlapping a known transcription start site from GENCODE transcripts[38], or overlapping a start site as experimentally determined by CAGE Tag data of the ENCODE pilot project[39]. Forward primers were designed for all other HindIII restriction fragments. Primers were excluded if highly repetitive sequences prevented the design of a sufficiently unique 5C primer. Primers settings were: U-BLAST: 3; S-BLAST: 130: 15-MER: 1320; MIN_FSIZE: 40; MAX_FSIZE: 50000; OPT_TM: 65; OPT_PSIZE: 40. DNA sequence of the universal tails of forward primers was CCTCTCTATGGGCAGTCGGTGAT; DNA sequence for the universal tails of reverse primers was AGAGAATGAGGAACC CGGGGCAG. A 6-base barcode was included between the specific part of the primers and the universal tail. In total 17 reverse primers and 90 forward primers were designed in the 500 kb *HoxA1* locus (ENm010) and hence a total of 1,530 *cis* interaction were interrogated in this region. Primer sequences are available separately (Supplementary Table 1).

3C was performed with HindIII as previously described[40] separately for fetal lung and foreskin fibroblasts (FB) and also for the control and *HOTTIP* knockdown foreskin FBs. For the 5C reaction, a total of 107 forward and reverse primers of HoxA1 region were mixed with either the ENCODE random region (ENr) primer pool comprising of 2,673 forward and 523 reverse primers (covering 30 additional ENCODE regions) or the ENr313 primer pool comprising of 57 forward and 58 reverse primers (covering 1 additional ENCODE region). 5C was then performed in 10 reactions each containing an amount of 3C library that represents 200,000 genome equivalents and 1 fmol of each primer. The 5C analysis of *HoxA1* region was carried out in two biological replicates of fetal lung and foreskin FBs. 5C ligation products were amplified using a pair of universal primers that recognize the common tails of the 5C forward and reverse primers described above and pooled together. To facilitate paired-end DNA sequence analysis on the Illumina GA2 platform, paired-end adaptor oligonucleotides were ligated to the 5C library using the Illumina PE protocol and PCR amplification of the library was carried out for 18 cycles with Illumina PCR primer PE 1.0 and 2.0. The 5C library was then sequenced on the Illumina GA2 platform generating 36 base paired end reads. For fetal lung FBs we obtained 7,625,276 and 10,947,424 mapped reads for two biological replicates of which 1,339,861 and 242,301 could be specifically mapped back to interactions within ENm010 using Novoalign (http://www.novocraft.com), respectively. For two biological replicates of foreskin FBs we obtained 7,311,386 and 5,731,107 mapped reads of which 2,752,789 and 66,769 could be mapped back to the ENm010 region, respectively. In the case of the knockdown study, control green fluorescent protein (GFP) knockdown foreskin FB 5C library yielded 4,909,482 mapped reads whereas *HOTTIP* knockdown foreskin FB had 5,565,389 mapped reads of which 39,168 and 38,950 could be mapped back to ENm010 for control GFP and *HOTTIP* knockdown, respectively. In the set with fetal lung and foreskin fibroblast samples, 5C for ENm010 was multiplexed for deep sequencing with 5C of one other region, ENr313; in the set containing the knockdown samples, ENm010 was multiplexed with 5C of 30 other genomic regions. The different extent of multiplexing resulted in different number of sequencing reads mapping back to ENm010. In all instances the mappable reads were proportional to the degree of multiplexing, indicating equivalent library quality despite different read numbers. Supplementary Table 2 outlines the library composition of each experiment. The heat maps are scaled as follows—for Fig. 1a, distal (foreskin) FBs: 262–17,467, proximal (lung) FBs: 7–5,846; for Supplementary Fig. 6, siGFP: 1–100, siHOTTIP: 1–100. Raw data from the 5C experiments used to generate the binned heat maps in Fig. 1a and Supplementary Fig. 6 can be found in Supplementary File 1. Raw data are available by request.

***HOTTIP* cloning, sequence and expression analysis.** We previously identified a portion of *HOTTIP* as a non protein-coding transcribed region named ncHOXA13-96 (ref. 12). This region also overlaps expressed sequence tag (EST) clone AK093987 that was previously observed to be expressed in cancer cell lines derived from posterior anatomic sites[41]. 5′ and 3′ RACE (RLM Race kit, Applied Biosystems/Ambion) showed full-length *HOTTIP* RNA to be 3,764 nucleotides, extending the known transcribed region by more than 1,400 bases. BLAST and BLAT confirmed that portions of *HOTTIP* are well conserved in mammals and even in avians but had no protein coding potential. Full-length *HOTTIP* RNA sequence has been deposited at NCBI (accession number GU724873). qRT–PCR with SYBR Green was conducted as recommended by the manufacturer (Agilent Technologies). Primer sequences specific for *HOTTIP* were CCTAAAGCCACGC TTCTTTG (HOTTIP-F) and TGCAGGCTGGAGATCCTACT (HOTTIP-R). For Supplementary Fig. 11, endogenous nascent *HOTTIP* was distinguished from ectopic *HOTTIP* expressed from cDNA using primers that spanned intron–exon junctions.

Strand-specific RT–PCR. RNA extracted from primary foreskin fibroblasts was reverse transcribed (SuperScript III, Invitrogen) using combinations of the previously described *HOTTIP*-specific primers HOTTIP-F and/or HOTTIP-R as diagrammed in Supplementary Fig. 1b. Resulting cDNA was then PCR-amplified using both HOTTIP-F and HOTTIP-R primers to visually determine strand specificity.

***HOTTIP* transcript count per cell.** The level of *HOTTIP* transcript per cell was calculated from the level of *HOTTIP* in 500,000 cells. Full-length *HOTTIP* in pcDNA3.1+ was assayed by qPCR using primers HOTTIP-F and HOTTIP-R at predetermined concentrations in triplicate to generate a linear amplification curve dependent on the moles of template DNA (Supplementary Fig. 2). The qRT–PCR value from 500,000 foreskin fibroblasts was determined and plotted, and the corresponding total molecules of transcript was divided by 500,000 to determine the approximate number of transcripts per cell.

Single-molecule RNA fluorescence *in situ* hybridization (RNA-FISH). Single molecule RNA-FISH was performed as described in ref. 42 with the following modifications: the amount of hybridization solution per chamber was doubled to allow for proper coating of the chamber and the amount of glucose-oxidase buffer was tripled to assist in image acquisition. Images were acquired using an Olympus FV1000 confocal microscope within 2 h of the addition of the glucose-oxidase buffer.

RNA interference. Primary foreskin fibroblasts were transfected with siRNAs targeting *HOTTIP* and *WDR5* using Lipofectamine 2000 (Invitrogen) as per manufacturer's instructions. Total RNA was harvested 48–72 h later using TRIzol (Invitrogen) and RNeasy Mini Kits (Qiagen) as previously described[34]. For the intronic *HOTTIP* knockdown experiment in Supplementary Figure 11, a pool of 10 siRNAs (Supplementary Table 3) targeting intronic regions in *HOTTIP* were transfected into foreskin fibroblasts, and RNA isolated as above.

Generation of shRNAs against chicken *HOTTIP*. A reporter construct encoding a GFP–chicken *HOTTIP* fusion transcript was used in a small-scale screen to identify highly effective shRNA constructs. Eleven shRNAs targeting conserved regions of chicken *HOTTIP* were designed and inserted into the pSMP system (Thermo/Open Biosystems). The reporter construct and shRNA constructs were cotransfected into Phoenix cells, and *HOTTIP* transcript levels were analysed via reduced GFP fluorescence and by qRT–PCR. Three shRNAs that were effective *in vitro* were then cloned into RCAS vector for studies in chick embryos[18].

Chick RNAi. RCAS *HOTTIP* hairpin and RCAS AP viruses were made by transfecting DF-1 cells with viral DNA. Transfected DF-1 cells were grown and passaged, after which the virus-containing supernatant was collected, concentrated and titred. Fertilized chicken eggs were incubated in a humidified rotating incubator at 37 °C until they reached Hamilton/Hamburger stage 10. Eggs were then windowed to expose the embryos. After gently removing the vitelline membrane, chicken embryos were microinjected with RCAS-*HOTTIP* hairpins and RCAS-AP viruses at the prospective wing and leg buds. All viral stocks have titres of 1×10^8 IU ml^{-1}, and each limb was injected five times. The infected embryos were allowed to incubate at 37 °C and were harvested 2 or 4 days after injection to detect viral infection by immunohistochemistry. Total RNA was extracted from injected forelimbs, and RT–PCR analysis was performed 4 days after injection. Chicken embryos were harvested 9 days post-injection to carry out whole-mount Alcian blue staining. A total of 50 animals were injected.

Hairpin sequences for chick *HOTTIP* were TGCTGTTGACAGTGAGCGAC CCGAAGATGTGTCTGATTTGTAGTGAAGCCACAGATGTACAAATCAGA CACATCTTCGGGCTGCCTACTGCCTCGGA (2-2-1), TGCTGTTGACAGTG AGCGCCGCTCTGCTCTCCTCTCTCTCTAGTGAAGCCACAGATGTAGAG AGAGAGGAGAGCAGAGCGATGCCTACTGCCTCGGA (3-2-1), and TGCT GTTGACAGTGAGCGAATCCTTAATCGAATCTGATTTTAGTGAAGCCACA GATGTAAAATCAGATTCGATTAAGGATCTGCCTACTGCCTCGGA (4-4-1).

***HOTTIP* overexpression.** Full-length *HOTTIP* and a truncated transcript consisting of exons 1 and 2 (*HOTTIP*$^{\text{Exons 1–2}}$) were cloned into the LZRS vector (gift of P. Khavari), and then transfected into Phoenix cells (gift of G. Nolan) to generate amphotropic retroviruses. Primary human fibroblasts were infected with either LZRS-full length *HOTTIP* (lung), LZRS-truncated *HOTTIP* (foreskin), or LZRS-GFP (both lung and foreskin), then passaged over 60 days, with periodic

testing of *HOXA* and *HOTTIP* expression by qRT–PCR. These cells were used in the rescue experiments depicted in Supplementary Fig. 11.

GST pull-down. Full-length *HOTTIP*, truncated *HOTTIP* containing exons 1 and 2 ($HOTTIP^{Exons\ 1-2}$), and histone H2B1 mRNA were transcribed *in vitro* using T7 polymerase according to manufacturer's instructions (Promega), denatured, and refolded in folding buffer (100 mM KCl, 10 mM $MgCl_2$, Tris pH 7.0). GST-tagged WDR5, C-terminal MLL1, RBBP5/Ash2L and TRF1 were expressed in *Escherichia coli* and purified as described[43]. Each GST-fusion protein was bound to glutathione beads (Amersham/GE Healthcare) and blocked with excess yeast total mRNA in PB100 buffer (20 mM HEPES pH 7.6, 100 mM KCl, 0.05% NP40, 1 mM DTT, 0.5 mM PMSF) for 1 h at room temperature. Beads were then incubated with either *in-vitro*-transcribed *HOTTIP* or histone H2B1 mRNA for 45 min at room temperature. After three washes in PB200 buffer (20 mM HEPES pH 7.6, 200 mM KCl, 0.05% NP40, 1 mM DTT, 0.5 mM PMSF), bound RNAs were extracted and analysed by qRT–PCR, as previously described.

RNA immunoprecipitation. HeLa-WDR5-Flag cells: 48 h after Lipofectamine 2000-mediated transfection of *HOTTIP* into HeLa WDR5-Flag cells (approximately 10^7 cells), total protein was extracted as previously described, with modifications[44]. Briefly, cells were resuspended in Buffer A (10 mM HEPES pH 7.5, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 1.0 mM PMSF), lysed in 0.25% NP40, and fractionated by low speed centrifugation. The nuclear fraction was resuspended and lysed in Buffer C (20 mM HEPES pH 7.5, 10% glycerol, 0.42 M KCl, 4 mM $MgCl_2$, 0.5 mM DTT, 1.0 mM PMSF). Combined nuclear and cytoplasmic fractions were immunoprecipitated with mouse anti-Flag M2 monoclonal antibody (Sigma) or mouse IgG affixed to agarose beads (Sigma) for 3 to 4 h at 4 °C. Beads were washed four times with wash buffer (50 mM TrisCl pH 7.9, 10% glycerol, 100 mM KCl, 5 mM $MgCl_2$, 10 mM β-mercaptoethanol, 0.1% NP40). After elution using Flag peptide (Sigma), co-immunoprecipitated RNA was extracted and analysed by qRT–PCR.

Endogenous WDR5 and SIRT6 RIP: cellular fractions were isolated as above and incubated with the anti-WDR5 (ref. 31) or anti-Sirt6 (ab62739, Abcam) antibodies overnight at 4 °C. Samples were washed in wash buffer, and co-immunoprecipitated RNA was extracted and analysed by qRT–PCR.

RNA chromatography. Full-length *in-vitro*-transcribed *HOTTIP* RNA was conjugated to adipic acid dehydrazide agarose beads as described[45]. The complexed beads were incubated with whole cell lysates from Hela WDR5-Flag cells, washed, and bound proteins visualized by western blotting.

***BoxB* tethering assay.** 293T cells were grown to about 50% confluence in 6-well plates on the day of transfection. Using Lipofectamine 2000 (Invitrogen), a plasmid encoding a luciferase gene under the control of five tandem GAL4 UAS sites were co-transfected with plasmids encoding GAL4-WDR5, GAL4-λN (the 22 amino acid RNA-binding domain of the lambda bacteriophage antiterminator protein N) peptide fused to a C-terminal GFP tag, *BoxB* (containing five repeats of the λN-specific 19 nucleotide binding site), *BoxB* fused to full-length *LacZ*, or *BoxB* fused to full-length *HOTTIP*. Cells were lysed 48 h after transfection, and luciferase assay kit (Promega) was used to determine relative levels of the luciferase gene product, following the manufacturer's protocol.

31. Chang, H. Y. *et al.* Diversity, topographic differentiation, and positional memory in human fibroblasts. *Proc. Natl Acad. Sci. USA* **99,** 12877–12882 (2002).
32. Chang, H. Y. *et al.* Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. *PLoS Biol.* **2,** 206–214 (2004).
33. Bernstein, B. E. *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120,** 169–181 (2005).
34. Rinn, J. L., Bondre, C., Gladstone, H. B., Brown, P. O. & Chang, H. Y. Anatomic demarcation by positional variation in fibroblast gene expression programs. *PLoS Genet.* **2,** e119 (2006).
35. Rinn, J. L. *et al.* A dermal *HOX* transcriptional program regulates site-specific epidermal fate. *Genes Dev.* **22,** 303–307 (2008).
36. Rinn, J. L. *et al.* A systems biology approach to anatomic diversity of skin. *J. Invest. Dermatol.* **128,** 776–782 (2008).
37. Soshnikova, N. & Duboule, D. Epigenetic temporal control of mouse *Hox* genes in vivo. *Science* **324,** 1320–1323 (2009).
38. Harrow, J. *et al.* GENCODE: producing a reference annotation for ENCODE. *Genome Biol.* **7** (Suppl 1), S4 (2006).
39. Birney, E. *et al.* Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799–816 (2007).
40. Dostie, J. & Dekker, J. Mapping networks of physical interactions between genomic elements using 5C technology. *Nature Protocols* **2,** 988–1002 (2007).
41. Sasaki, Y. T., Sano, M., Kin, T., Asai, K. & Hirose, T. Coordinated expression of ncRNAs and *HOX* mRNAs in the human *HOXA* locus. *Biochem. Biophys. Res. Commun.* **357,** 724–730 (2007).
42. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods* **5,** 877–879 (2008).
43. Smith, D. B. & Johnson, K. S. Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione *S*-transferase. *Gene* **67,** 31–40 (1988).
44. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11,** 1475–1489 (1983).
45. Michlewski, G. & Caceres, J. F. RNase-assisted RNA chromatography. *RNA* 16, 1673–1678 (2010).