

XACT, a long noncoding transcript coating the active X chromosome in human pluripotent cells

Céline Vallot^{1,2}, Christophe Huret^{1,2}, Yann Lesecque³,
Alissa Resch⁴, Noufissa Oudrhiri⁵, Annelise Bennaceur-Griscelli⁵,
Laurent Duret³ & Claire Rougeulle^{1,2}

X-chromosome inactivation (XCI) in mammals relies on *XIST*, a long noncoding transcript that coats and silences the X chromosome in cis. Here we report the discovery of a long noncoding RNA, *XACT*, that is expressed from and coats the active X chromosome specifically in human pluripotent cells. In the absence of *XIST*, *XACT* is expressed from both X chromosomes in humans but not in mice, suggesting a unique role for *XACT* in the control of human XCI initiation.

A major recent breakthrough in the conception of eukaryotic gene regulation has been the identification of a multitude of long noncoding RNAs (lncRNAs) in mammals that are scattered throughout the genome and are located, in particular, in intergenic regions (lincRNAs)¹. Although the number of lncRNAs is estimated to be more than 1,000, only 12 have been functionally characterized. These lncRNAs were shown to function in diverse cellular processes, most predominantly in the regulation of gene expression. In this context, many lncRNAs participate in gene silencing pathways, some of which involve the recruitment of repressive chromatin complexes².

XCI is one of the most studied processes involving lncRNA-mediated repression. *XIST* is a noteworthy transcript in that in addition to silencing an (approximate) entire chromosome, it is the only lncRNA described thus far to widely coat the chromosome from which it is expressed. This coating by *XIST* induces substantial nuclear reorganization and recruitment of histone-modifying complexes that are important for the initiation and maintenance, respectively, of X-chromosome silencing³.

XCI is established early during embryonic development, and embryonic stem cells can be used to decipher the kinetics and molecular actors of the process. In female human embryonic stem cells (hESCs), one X chromosome is, in most cases, already inactivated⁴. Using an RNA sequencing (RNA-seq) analysis of female H9 hESCs (C.V., Ouimette, J.-F., Makhoulouf, M., Féraud, O., N.O., A.B.-G., Côme, J., Martinat, C., A.R., Lalande, M. & C.R., unpublished data), we identified a large unannotated region (~252 kb) on

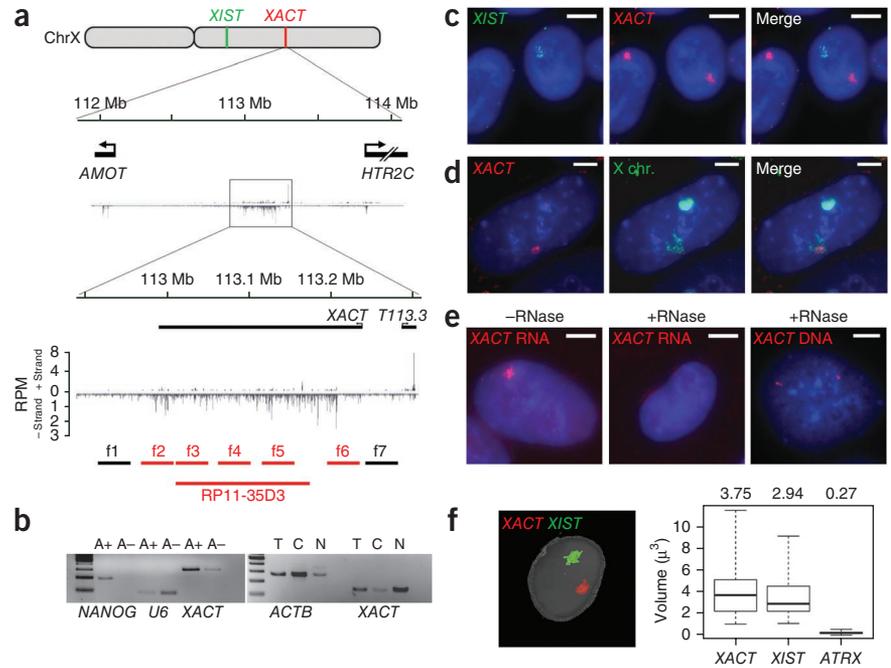
the X chromosome for which we could detect a substantial amount of expression originating from the minus strand (Fig. 1a). We called this transcript *XACT*. The *XACT* region is located on chromosome Xq23 between the protein-coding genes *AMOT* and *HTR2C* in an unusually large intergenic domain of 1.7 Mb (only 1% of intergenic regions in humans are >1.5 Mb). This domain is GC poor (37%) and rich in repeat sequences (68% compared to 62% for the X chromosome and 45% for the overall human genome), notably long interspersed nuclear elements and long terminal repeats (LTRs) (Supplementary Fig. 1). All RNA-seq reads were co-linear with the genomic sequence, suggesting that the *XACT* transcript is unspliced (Supplementary Methods). We detected similar transcription in male H1 hESCs, and alignment with published chromatin immunoprecipitation sequencing data⁵ revealed histone H3 Lys36 trimethylation (H3K36me3) and histone H3 Lys79 dimethylation (H3K79me2) blocks along the region, as well as peaks of H3K4me3 and RNA polymerase II (PolII) at the telomeric extremity of the transcribed region (Supplementary Fig. 2a). These peaks correspond to LTR elements, which might serve as promoters for this transcribed region⁶. We further characterized the 5' end of *XACT* by RT-PCR and 5' RACE (Supplementary Fig. 2b). This analysis revealed multiple transcription start sites (TSSs) clustered in a 126-bp region characteristic of broad-type promoters⁷, which coincide with peaks of H3K4me3 and RNA PolII (Supplementary Fig. 2). Together these data suggest that *XACT* corresponds to a single 251.8-kb transcription unit (112,983,323–113,235,148 bp). The *XACT* transcript is polyadenylated and mostly nuclear (Fig. 1b), further suggesting that it acts, at least in part, as a noncoding RNA. We also identified a distinct transcript 5' of *XACT* on the plus strand called *T113.3* (Fig. 1a and Supplementary Fig. 3). Unlike *XACT*, *T113.3* is spliced and mostly cytoplasmic, and its TSS, mapped by 5' RACE, lies within a peak of H3K4me3 located 48 kb upstream of the TSS of *XACT*.

We next investigated transcription of *XACT* at the cellular level by RNA FISH. Remarkably, a BAC probe covering 151 kb of the transcribed region detected a unique, large signal in female H9 hESCs that was reminiscent of the *XIST* RNA cloud. However, this signal corresponded to the active X chromosome, as determined by combined RNA FISH with a *XIST* probe, which labels the inactive X chromosome (Fig. 1c), and simultaneous RNA and DNA FISH with an X-paint probe (Fig. 1d). The *XACT* RNA cloud signal was partially resistant to the stringent denaturation steps used in the DNA FISH experiments (Fig. 1d; note that the *XACT* signal appears smaller than it does in classical RNA FISH experiments) but not to RNase treatment (Fig. 1e), indicating a strong association of the RNA with the active X chromatin. The BAC probe detected two pinpoints in the DNA FISH

¹Université Paris Diderot, Sorbonne Paris Cité, Epigenetics and Cell Fate, Paris, France. ²Centre National de la Recherche Scientifique (CNRS), Unité Mixte de Recherche (UMR) 7216 Epigenetics and Cell Fate, Paris, France. ³Laboratoire de Biométrie et Biologie Evolutive, UMR CNRS 5558, Université de Lyon, Université Lyon 1, Villeurbanne, France. ⁴Stem Cell Institute, University of Connecticut Health Center, Farmington, Connecticut, USA. ⁵ESTeam Paris Sud, Institut National de la Santé et de la Recherche Médicale (INSERM) U935, Université Paris Sud 11, Assistance Publique-Hôpitaux de Paris (AP-HP), Villejuif, France. Correspondence should be addressed to C.R. (claire.rougeulle@univ-paris-diderot.fr).

Received 18 July 2012; accepted 20 December 2012; published online 20 January 2013; doi:10.1038/ng.2530

Figure 1 Characterization of *XACT*, a long intergenic RNA that coats the active X chromosome in hESCs. (a) Schematic map showing the localization of *XACT* and *XIST* on the human X chromosome. Bottom, strand-specific RNA-seq showing >250 kb of continuous transcription originating from the minus strand in female H9 cells. The boxed area is a magnification of the locus. The locations of the BAC probe (RP11-35D3) and fosmid probes (f1–f7) used for further FISH analysis are indicated below (**Supplementary Fig. 4**). Probes giving an RNA cloud signal for *XACT* are in red, and probes not giving a signal are in black. RPM, reads per million mapped reads. (b) Left, semiquantitative RT-PCR analysis of poly(A)-positive (A+) and poly(A)-negative (A-) enriched RNA fractions of H9 hESCs. *NANOG* and *U6* were used as positive controls for the poly(A)-positive and poly(A)-negative fractions, respectively. Right, RT-PCR analysis of total (T), cytoplasmic (C) and nuclear (N) RNAs. (c) RNA FISH analysis of H9 cells with a *XIST* probe (green) and a *XACT* probe (red). All *XACT* FISH experiments were performed using the BAC RP11-35D3 probe, except where otherwise indicated. (d) Simultaneous RNA and DNA FISH of H9 cells with the *XACT* probe (red) detecting *XACT* RNA and DNA and an X-paint probe (green) detecting X chromosomes. (e) Left and middle, *XACT* RNA FISH with and without initial treatment with RNase. Right, *XACT* DNA FISH with initial treatment with RNase. Scale bars (c–e), 5 μm . (f) Three-dimensional model of the nuclear volumes occupied by *XACT* and *XIST* RNAs. The boxplot represents the distribution of the volumes ($n = 50$ nuclei). The distribution of the volumes corresponding to the transcription signal of an X-linked gene (*ATRX*) is shown for comparison. The median value of each distribution is indicated above the corresponding boxplot.



experiments (after RNase treatment of the slides), which confirms that the RNA cloud signal corresponds to actual coating of the chromosome by *XACT* RNA and not to the genomic organization of the locus. We called this RNA *XACT* (X active coating transcript). The nuclear volume occupied by *XACT* in the nucleus was similar to that occupied by *XIST* (median volumes of 3.75 μm^3 and 2.94 μm^3 , respectively; **Fig. 1f**). RNA FISH analysis of *XACT* expression using a series of fosmid probes covering the region (**Fig. 1a**) further confirmed the extent and expression profile of *XACT* (**Supplementary Fig. 4**).

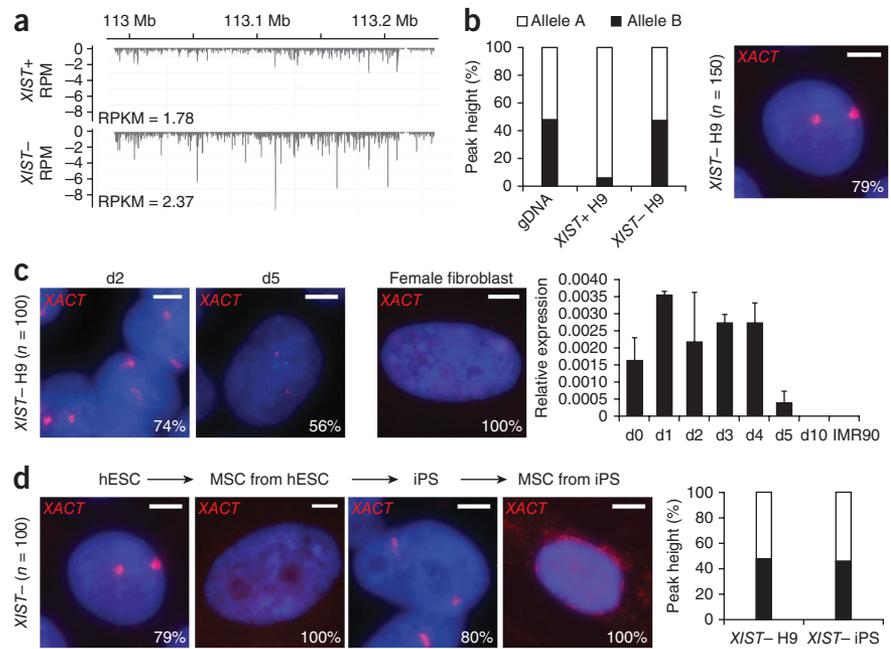
X inactivation is highly unstable in female hESCs, and *XIST* downregulation tends to occur spontaneously in culture⁸. Loss of *XIST* expression leads to substantial but incomplete gene reactivation from the originally inactive X chromosome (Xi^*)⁹ (C.V., Ouimette, J.-F., Makhoul, M., Féraud, O., N.O., A.B.-G., Côme, J., Martinat, C., A.R., Lalande, M. & C.R., unpublished data). RNA-seq performed in H9 hESCs not expressing *XIST* revealed a slight increase in *XACT* expression compared to H9 cells that did express *XIST* (**Fig. 2a**; reads per kilobase per million reads (RPKM) = 1.78 in cells expressing *XIST*, and RPKM = 2.37 in cells not expressing *XIST*). We took advantage of a SNP (rs5929175) within *XACT* and the clonal X-inactivation pattern shown by H9 cells^{10,11} to address the allelic expression of *XACT* in cells not expressing *XIST* compared to *XIST*-expressing cells. Whereas *XACT* was monoallelically transcribed in cells expressing *XIST*, it became biallelically expressed in cells not expressing *XIST* (**Fig. 2b**). RNA FISH analysis further revealed that *XACT* is not only re-expressed from but also coats the Xi^* in cells not expressing *XIST*, leading to two *XACT* clouds in these cells (**Fig. 2b**). Whether *XACT* re-expression from and coating of the Xi^* is a cause or a consequence of *XIST* repression and subsequent partial reactivation of the chromosome remains to be established.

XACT is expressed from active X chromosome(s) in hESCs. In contrast, we were not able to detect *XACT* reads from female fibroblast RNA-seq libraries. Similarly, RT-PCR analysis revealed no or weak

expression of *XACT* in various tissues, including brain, muscle and placenta (**Supplementary Fig. 5**), suggesting that *XACT* is down-regulated after differentiation. To investigate the kinetics of *XACT* silencing, we induced H9 hESCs to differentiate and found substantial downregulation of *XACT* at day 5 and full silencing at day 10 (**Fig. 2c**). To further probe the link between *XACT* and pluripotency, we used a model system in which H9 hESCs not expressing *XIST* were subjected to several rounds of differentiation and reprogramming¹² (**Fig. 2d**). As detected with undirected differentiation, the differentiation of these H9 hESCs into mesenchymal stem cells (MSCs) was accompanied by biallelic silencing of *XACT*. Remarkably, we found strong re-expression of *XACT* from both X chromosomes in MSC-derived induced pluripotent stem (iPS) cells, whereas silencing of *XACT* was re-established when these iPS cells were differentiated into MSCs (iPS-MSCs). Together these results indicate that, in this context, *XACT* expression and coating of the X chromosomes is restricted to pluripotent and early differentiating cells in humans. However, we cannot exclude the possibility that *XACT* is expressed in some differentiated cell types.

We then investigated the conservation of *XACT* by combining *in silico* analyses and experimental studies (**Supplementary Fig. 6**). The organization of the genomic region encompassing *AMOT* and *HTR2C* is well conserved in placental mammals and marsupials. In contrast, the sequence between these two genes shows moderate conservation, with several conserved blocks present in placental mammals but not marsupials, which are interspaced with divergent regions. The LTR that corresponds to the 5' end of *XACT* is conserved in chimpanzees but not macaques or more distally related species (**Supplementary Fig. 6a**). This suggests that the insertion of these LTR elements is a very recent event. To further probe *XACT* conservation in mice, we undertook systematic DNA and RNA FISH studies in mouse ESCs (mESCs) using a series of six BAC probes spanning the region between *Amot* and *Htr2c*. Although all these probes detected

Figure 2 *XACT* coats both X chromosomes in hESCs not expressing *XIST*, and the expression of *XACT* is restricted to pluripotent cells. (a) RNA-seq data for H9 hESCs expressing (*XIST*⁺) or not expressing (*XIST*⁻) *XIST* with RPKM values shown. (b) Left, pyrosequencing analysis of rs5929175 for H9 genomic DNA (gDNA) and *XIST*⁺ and *XIST*⁻ complementary DNA. The bar chart indicates the percentage of the peak height corresponding to each allele. Right, RNA FISH using the *XACT* probe in *XIST*⁻ H9 hESCs revealing that 79% of the nuclei (*n* = 150) have two *XACT* RNA clouds. (c) RNA FISH and quantitative RT-PCR analysis of *XACT* expression during differentiation of *XIST*⁻ hESCs and in female fetal (IMR90) and adult (Coriell, AG09603) fibroblasts. d0–d10, days 0–10. The images are representative of the major population. The bar charts correspond to the average of two independent differentiation experiments. Errors bars indicate the s.d. calculated for two samples. (d) RNA FISH and pyrosequencing analysis of *XACT* expression during successive rounds of differentiation and reprogramming of *XIST*⁻ H9 cells into MSCs. Scale bars (b–d), 5 μ m. The percentages shown in b–d indicate the number of nuclei showing similar expression patterns of *XACT*.



the two X chromosomes by DNA FISH, none of them generated a signal by RNA FISH (**Supplementary Fig. 6b**), suggesting that the region syntenic to *XACT* is not expressed in mESCs. In agreement with this result, analysis of mESC RNA-seq data did not reveal broad, ‘*XACT*-like’ transcription in the syntenic region. However, we were able to identify a few discrete peaks in mESCs and other cell types (**Supplementary Fig. 6b**). Whether these peaks correspond to real transcripts remains to be investigated. Together our data suggest that there is no *XACT*-like transcript expressed in mESCs.

We have identified *XACT* as the first lncRNA that coats the active X chromosome in humans. The identification of a lncRNA that coats an active chromosome (whereas most lncRNAs studied so far are involved in gene silencing) underlies the multifaceted nature of lncRNAs. *XACT* might not be conserved in mice, and its function might be related to the specific kinetics of XCI that were recently described in the human. Indeed, in human preimplantation embryos, *XIST* is expressed from the paternal and maternal X chromosomes, but this does not lead to chromosome-wide silencing¹³. In contrast, paternal *Xist* expression and XCI characterize mouse preimplantation development¹⁴. Given its expression profile, it is tempting to speculate that *XACT* is involved in the control of XCI initiation in humans. More generally, rapidly evolving lncRNAs such as *XACT* in the human and *Tsix* in the mouse¹⁵ could be involved in species-specific regulation of XCI, thus underlying the important plasticity that characterizes this process among mammals.

Accession codes. Sequence data and alignments have been submitted to the Gene Expression Omnibus (GEO) database under accession code GSE39757.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank members of our laboratory and M. Lalande for stimulating discussion and critical reading of the manuscript. The research leading to these results has received funding from the European Research Council (ERC) under the European Community’s Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement 206875 and the INSERM (Avenir Program R0721HS).

AUTHOR CONTRIBUTIONS

C.V. and C.H. performed the experiments. Y.L. and L.D. did the bioinformatic analysis of *XACT* conservation. N.O. and A.B.-G. provided the H9-MSC-iPS system. A.R. contributed to the bioinformatics analysis of the RNA-seq data. C.V. and C.R. conceived and designed the experiments, and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/doi/10.1038/ng.2530>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Guttman, M. *et al. Nature* **458**, 223–227 (2009).
- Khalil, A.M. *et al. Proc. Natl. Acad. Sci. USA* **106**, 11667–11672 (2009).
- Augui, S., Nora, E.P. & Heard, E. *Nat. Rev. Genet.* **12**, 429–442 (2011).
- Dvash, T. & Fan, G. *Epigenetics* **4**, 19–22 (2009).
- The ENCODE Project Consortium. *PLoS Biol.* **9**, e1001046 (2011).
- Romanish, M.T., Lock, W.M., van de Lagemaat, L.N., Dunn, C.A. & Mager, D.L. *PLoS Genet.* **3**, e10 (2007).
- Sandelin, A. *et al. Nat. Rev. Genet.* **8**, 424–436 (2007).
- Silva, S.S., Rowntree, R.K., Mekhoubad, S. & Lee, J.T. *Proc. Natl. Acad. Sci. USA* **105**, 4820–4825 (2008).
- Mekhoubad, S. *et al. Cell Stem Cell* **10**, 595–609 (2012).
- Mitjavila-Garcia, M.T. *et al. J. Mol. Cell Biol.* **2**, 291–298 (2010).
- Shen, Y. *et al. Proc. Natl. Acad. Sci. USA* **105**, 4709–4714 (2008).
- Giuliani, M. *et al. Blood* **118**, 3254–3262 (2011).
- Okamoto, I. *et al. Nature* **472**, 370–374 (2011).
- Okamoto, I., Otte, A.P., Allis, C.D., Reinberg, D. & Heard, E. *Science* **303**, 644–649 (2004).
- Rougeulle, C. & Avner, P. *Semin. Cell Dev. Biol.* **14**, 331–340 (2003).