Phase separation drives X-chromosome inactivation: a hypothesis

The long non-coding RNA *Xist* induces heterochromatinization of the X chromosome by recruiting repressive protein complexes to chromatin. Here we gather evidence, from the literature and from computational analyses, showing that *Xist* assemblies are similar in size, shape and composition to phase-separated condensates, such as paraspeckles and stress granules. Given the progressive sequestration of *Xist*'s binding partners during X-chromosome inactivation, we formulate the hypothesis that *Xist* uses phase separation to perform its function.

Andrea Cerase, Alexandros Armaos, Christoph Neumayer, Philip Avner, Mitchell Guttman and Gian Gaetano Tartaglia

The ability of protein and RNA to interact affects the formation of membrane-less organelles, such as paraspeckles and stress granules, which are involved in various biological functions[1], including RNA processing and responses to environmental changes[2–4]. As studied by confocal microscopy[5] and super-resolution approaches[6], these assemblies appear as round foci that are in dynamic exchange with their environment. They form when locally saturated molecules separate into two phases, a process known as 'liquid – liquid phase separation' (LLPS)[7].

LLPS is promoted by proteins carrying intrinsically disordered regions (IDRs) that lack a specific three-dimensional (3D) structure. IDRs are enriched for amino acid repetitions[7] characterized by polar residues that favor protein – protein interactions[8] and isolated hydrophobic regions that, exposed to the solvent, promote aggregation[9]. In RNAs, the presence of nucleotide repetitions can induce the formation of structures that attract proteins[2,4]. For example, the long non-coding RNA *Neat1* uses repeats to sequester specific proteins and drives LLPS in the paraspeckle[3].

Here we provide several lines of evidence in support of a model in which the long

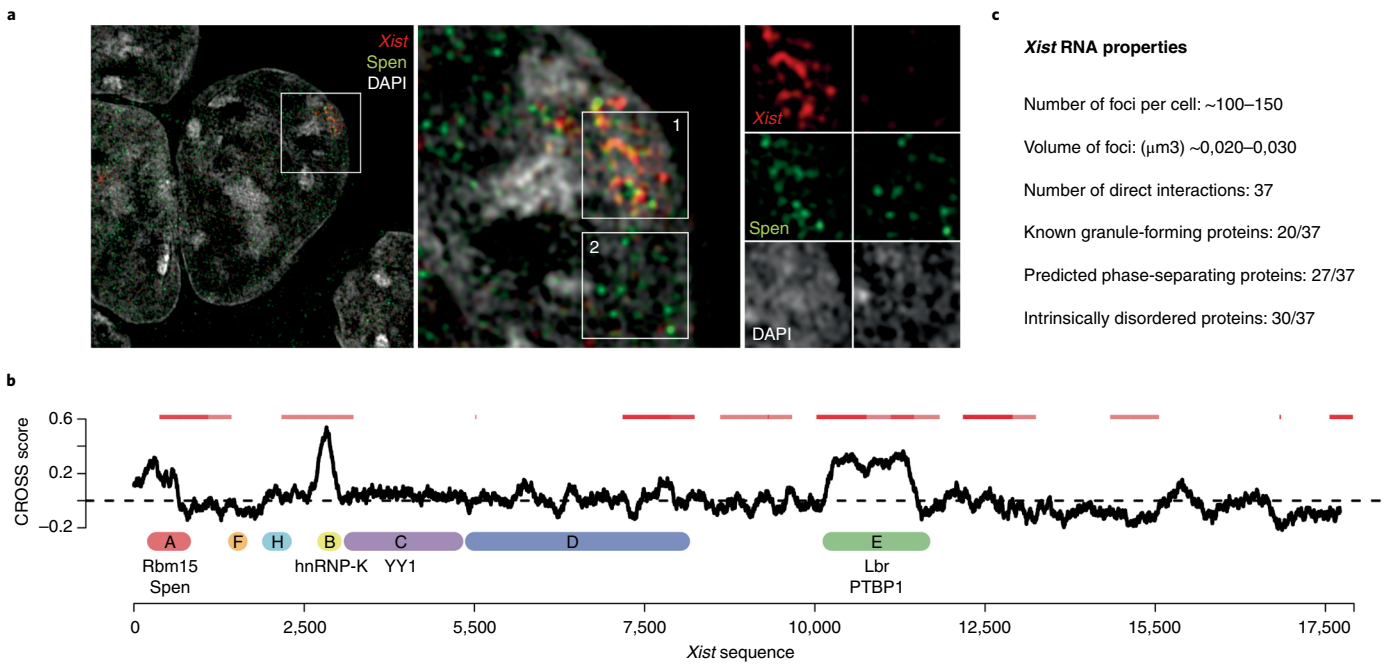

Fig. 1 | Supporting evidence that *Xist* might form a phase-separated compartment. a, Example of *Xist* assemblies reproduced from ref. [30], published under a CC-BY license, showing *Xist* RNA (red) and one representative interacting protein, Spen (green), with DNA in gray (DAPI staining). Outlined area in left image is enlarged at right; stacked outlined areas (middle) indicate regions with (1) or without (2) the inactive X chromosome; right, single-color images of those outlined areas (1, near right; 2, far right). **b**, *Xist* secondary structure prediction, performed with the CROSS (Computational Recognition of Secondary Structure) algorithm[15]: *y*-axis, double-stranded regions (positive values) and single-stranded regions (negative values); *x*-axis, linear sequence of *Xist*, from mouse mm10 (MGI:98974)[15]. Red horizontal lines (top) indicate *Xist* regions at which proteins with strong phase-separation propensity are predicted to bind[12]; horizontal bars in plot (various colors) indicate *Xist* repeat regions (A–F and H), with *Xist* binding partners validated by crosslinking and immunoprecipitation (below horizontal bars). **c**, The properties and cellular distribution of *Xist*[8,10,11] (denominator in each fraction indicates total number of proteins analyzed).

Table 1 | The *Xist* direct interactome is enriched for phase-separating proteins

Name	Ensembl ID	UniProt	IDR	DC	LLPS	SG	PSP	I
Safb	ENSMUSG00000071054	D3YXK2	Y	100%	1.77			0.61
HnrnpA2b1	ENSMUSG00000004980	O88569	Y	100%	4.64	Y		0.77
Rbm3	ENSMUSG00000031167	O89086		95%	3.43	Y	Y	1.00
Ptbp1	ENSMUSG00000006498	P17225		20%	0.70	Y		1.00
Celf1	ENSMUSG00000005506	P28659		25%	0.58	Y		0.75
HnrnpK	ENSMUSG00000021546	P61979		95%	1.60	Y	Y*	1.00
Srsf3	ENSMUSG00000071172	P84104		95%	0.91			0.97
Rbm15	ENSMUSG00000048109	Q0VBL3	Y	100%	1.74			0.89
Fubp3	ENSMUSG00000026843	Q3TIX6		65%	1.31	Y		0.98
Khsrp	ENSMUSG00000007670	Q3U0V1	Y	100%	2.87	Y		0.96
Lbr	ENSMUSG00000004880	Q3U9G9		5%	0.31			0.79
Spen	ENSMUSG00000040761	Q3UV27	Y	100%	1.33			0.59
HnrnpD	ENSMUSG00000000568	Q60668	Y	100%	2.92	Y		0.98
Khdrbs1	ENSMUSG00000028790	Q60749	Y	100%	1.14	Y		0.95
Pcbp2	ENSMUSG00000056851	Q61990		30%	0.53	Y		0.99
Raly	ENSMUSG00000027593	Q64012	Y	100%	1.55			1.00
HnrnpQ	ENSMUSG00000032423	Q7TMK9		85%	2.28			0.94
Srsf7	ENSMUSG00000024097	Q8BL97		80%	1.08			1.00
Rbfox2	ENSMUSG00000033565	Q8BP71	Y	100%	1.32			0.93
Rbm14	ENSMUSG00000006456	Q8C2Q3		95%	0.81		Y*	1.00
Myef2	ENSMUSG00000027201	Q8C854		90%	2.67			0.99
Matr3	ENSMUSG00000037236	Q8K310	Y	100%	1.52	Y		0.99
HnrnpL	ENSMUSG00000015165	Q8R081		95%	1.86	Y		0.88
Fbxw7	ENSMUSG00000028086	Q8VBV4		60%	0.93			0.99
HnrnpU	ENSMUSG00000039630	Q8VEK3		95%	2.57	Y		0.66
HnrnpR	ENSMUSG00000066037	Q8VHM5		90%	2.22	Y	Y*	0.69
Tardbp	ENSMUSG00000041459	Q921F2		75%	2.02	Y	Y*	0.88
HnrnpAb	ENSMUSG00000020358	Q99020	Y	100%	2.20	Y		1.00
Nxf1	ENSMUSG00000010097	Q99JX7		65%	0.83	Y		0.89
HnrnpA0	ENSMUSG00000007836	Q9CX86	Y	100%	4.78	Y		0.98
Srsf9	ENSMUSG00000029538	Q9D0B0		45%	1.69			0.94
HnrnpM	ENSMUSG00000059208	Q9D0E1		50%	2.52	Y		0.75
Srsf10	ENSMUSG00000028676	Q9R0U0	Y	100%	1.10		Y*	1.00
Preb	ENSMUSG00000045302	Q9WUQ2		30%	0.55			0.89
Tcf7l1	ENSMUSG00000055799	Q9Z1J1	Y	100%	0.38			0.70
HnrnpC	ENSMUSG00000060373	Q9Z204	Y	100%	1.32	Y		1.00
HnrnpF	ENSMUSG00000042079	Q9Z2X1		15%	1.48		Y*	0.96

Proteins predicted to contain IDRs[21], the degree of disorder confidence (DC)[21], the propensity for LLPS (all values are >0, which indicates that the proteins are prone to phase separate; values above 1 indicate a strong propensity)[8], known components of stress granules (SG)[13,18] and paraspeckles (PSP) (*, essential for paraspeckle formation; 'Y', assignment to the classification)[6,17]. 'I' represents the score of *Xist*-interacting proteins predicted by the catRAPID Global Score (a score of >0.5 indicate strong binding, and all scores were verified by at least two independent experiments)[12]. Bold indicates proteins with a known role in XCI. HnrpK, which directly interacts with Pcgf3 and Pcgf5 of the PRC1 complex, is intrinsically disordered (Supplementary Table 1).

non-coding RNA *Xist* likewise triggers phase separation to ensure efficient and persistent X-chromosome inactivation (XCI). To start, stress granules and the assemblies formed by *Neat1* and those formed by *Xist* show considerable similarity in size and shape. High-resolution microscopy indicates that the inactive X chromosome contains about 100 *Xist* foci identifiable by RNA-FISH. Each focus is spheroid in shape, measures around 300 – 400 nm in diameter[10,11] and includes several *Xist*-interacting proteins, such as Spen[12]. This is comparable to the morphologies and dimensions of stress granules[3,13] and paraspeckles[3] (Fig. 1a).

Neat1 adopts a hairpin-like structure in the paraspeckle[3]. As shown by RNA-FISH and immunohistochemistry, proteins such as Nono, Fus, Sfpq and Ppsp1 bind to tandem repeats in the central region of *Neat1*, while other domains in the 5′ and 3′ regions contact fewer proteins[6]. *Xist* also contains nucleotide repeats (A – F) that are conserved in mammals and have a crucial role in XCI[14]. Both dimethyl-sulfate experiments and computational models indicate that repeats A, B and E are structured and repeat D is partially structured, while the 3′ region of *Xist* is mostly single-stranded[15]. Computational analyses suggest that the

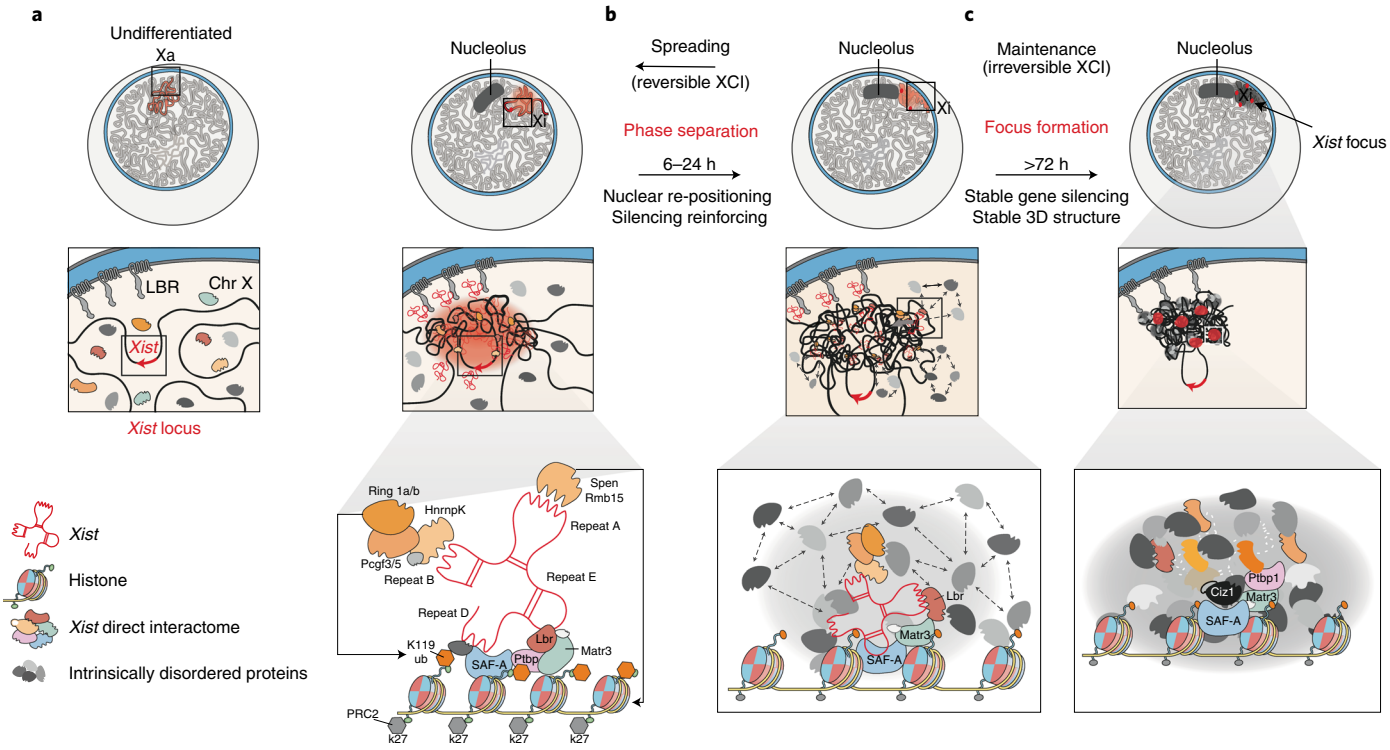

Fig. 2 | XCI model. a, Initiation of XCI. Prior to XCI (left), the X chromosome (Chr X) is in an active state (Xa) and shows a classic euchromatic nuclear localization and basal *Xist* transcription. When *Xist* is upregulated (right), triggered by cell differentiation or by drug induction, it spreads locally via 3D proximity and the process of XCI initiates (1–3 h)[24,25]. Early stages are mediated through the direct recruitment of proteins by *Xist* (bottom). *Xist* repeat B recruits the partially disordered protein HnrnpK (light orange), which in turn recruits the PRC1 complex (Ring1a or Ring1b, and Pcgf3 or Pcgf5; light orange)[20]. PRC1 leads to the accumulation of H2AK119ub (K119ub; orange hexagons) and indirectly to the accumulation of H3K27me2 and H3K27me3 (K27; gray hexagons)[23]. *Xist* repeat A recruits Spen and Rmb15 (dark orange), which mediate histone deacetylation and early gene silencing. *Xist* repeat E attracts Matr3 (green) and Ptbp1 (pink) and possibly indirectly interacting proteins, such as Ciz1 (also partially disordered; black in **b**); these drive the interaction of *Xist* with the nuclear matrix and the Lamin-B receptor (Lbr; ruby), which mediates localization of the inactive X chromosome (Xi) to the periphery of the nucleus[14,19,27]. Saf-A binds *Xist* broadly and may help focus formation via its intrinsically disordered domains[19] (Table 1). **b**, Spreading of XCI. *Xist* spreads along the whole X chromosome, and the silencing process is reinforced by histone modifications (6–24 h)[24,25]. Phase separation is suggested to start as a consequence of a local increase in the concentration of IDR-containing proteins (gray) interacting with *Xist* and possibly by peripheral targeting of the X chromosome. At this stage, XCI is still reversible[28]. **c**, Maintenance of XCI. A high concentration of disordered proteins leads to further recruitment of additional disordered proteins and focal accumulation. Maintenance of XCI is independent of *Xist* and is irreversible at this stage, due to the multi-layered accumulation of repressive marks (>72 h of differentiation)[19,28]. The time reported for protein recruitment are based mostly on data from *Xist* inducible systems[24,25] and may differ in other systems depending on cellular conditions such as differentiation or endogenous *Xist* regulation.

structured region E (as well as regions A and B, to a much lesser extent) is (are) crucial for the interactions of *Xist* with proteins[12] (Fig. 1b).

Xist directly binds to several proteins, and hundreds more have been found to associate through indirect, yet potentially relevant, interactions. The *Xist* protein network significantly overlaps that of *Neat1* (14 of 37 direct interactors; details in Table 1, with further information (including details of statistical analysis) in Supplementary Table 1)[16], including seven paraspeckle components, of which six have been shown to be essential for paraspeckle formation[17]. In addition, there is overlap with stress-granule components (20 of 37; Table 1)[13,18]. For example, among the shared interacting proteins are Rbm14 and Tardbp1[17], which are present in the paraspeckle[6], and Hnrnp proteins. These are constituents of stress granules and also have a role in *Xist*-mediated gene silencing (HnrnpU; also called Saf-A)[19] and recruitment of Polycomb transcription-repressive complex (PRC) proteins (HnrnpK)[20]. Taking advantage of published calculations[8], we found that proteins that directly interact with *Xist* and 54% of the entire *Xist* interactome are predicted to be prone to phase separation (Table 1 and Fig. 1c). The potential for *Xist* to drive phase separation is further supported by the significant enrichment for IDR-containing proteins in the *Xist* interactome (Fig. 1c). Indeed, 15 of 37 directly interacting proteins are predicted to be highly disordered, with a confidence of 100% (details of statistical analysis, Supplementary Table 1), as indicated by calculations carried out in previous work[8,21]. For comparison, the fraction of proteins predicted to be structurally disordered in the *Mus musculus* genome is only 18.7% (i.e., 3,175 of 16,964 proteins with sequence similarity of <75%; Uniprot database).

LLPS is typically characterized by fast recovery times of the molecular constituents, as measured by fluorescence recovery after photobleaching. This is due to rapid exchange between the assembly and the environment. Comparison of such experiments indicates that paraspeckle components interacting with *Neat1* in myoblast cells have diffusion kinetics similar to those of *Xist*'s partners in embryonic stem cells[5,22]. The recovery times (half-life ($t_{1/2}$)) of paraspeckle core components Psp1,

P54nrb and Psf lie in the range of 5 – 10 s, which coincides with those reported in experiments with processing bodies and stress granules[8]. Similarly, Pcgf3 and Pcgf5, part of the Polycomb silencing complex PRC1, directly interact with *Xist* and recover within seconds[23]. However, while the recovery time ($t_{1/2}$) of *Neat1* is ~90 s, *Xist* requires a longer period for recovery ($t_{1/2}$ = ~600 s)[5,22]. This is not surprising, since *Xist* is firmly anchored to the nuclear matrix and thus its diffusion is limited by physical constrains[22].

Below, we put our hypothesis into the context of XCI. During the initiation phase, *Xist* acts as an indispensable RNA scaffold that recruits repressive proteins and induces histone deacetylation and early gene silencing (1 – 3 h)[19,24,25] (Fig. 2a). We suggest that at the onset of *Xist*'s spreading from its transcription locus on the X chromosome into the 3D proximal neighborhood, phase separation is initiated by direct interactions involving IDR-containing proteins (6 – 24 h; Fig. 2b)[24,25]. At this stage, proteins that directly interact with *Xist*, such as Spen, Ptbp1, HnrnpK, PRC1 and PRC2, multimerize[23], which increases their local concentration. This recruitment leads to histone modifications such as further deacetylation and gain of repressive chromatin marks (H3K27me3 and H2AK119ub)[19]. As *Xist*-interacting proteins bind to other IDR-containing proteins, further compaction and focus formation take place[26] (Fig. 2b). As XCI proceeds, the inactive X chromosome is sequestered to the periphery of the nucleus by specific interactions between *Xist* and the Lamin-B receptor[12,27]. It is possible that peripheral localization of the inactive X chromosome aids in sustaining phase separation. Indeed, the lamina has been shown to be a largely heterochromatic region that serves as a hub for inactive genes. Heterochromatic proteins are known to be enriched for IDRs (45 of 102). Although heterochromatic proteins show little overlap with proteins that directly interact with *Xist*, we found that 21 of 37 *Xist* partners contact heterochromatin proteins. Finally, a shift from reversible XCI to irreversible XCI occurs after 72 h of differentiation[24,25,28]. From then on, *Xist* is not required for the maintenance of XCI[29] (Fig. 2c). This event is compatible with the nonlinearity of phase separation, which indicates that when a critical concentration is reached, the scaffold provided by *Xist* is no longer necessary and the protein assembly is stable on its own.

In summary, on the basis of the points reiterated below, we suggest that *Xist*, together with its binding partners, promotes phase separation: (1) *Xist* foci are similar in size and morphology to paraspeckles and stress granules; (2) *Xist* contains nucleotide repeats that are present in scaffold RNAs and promote protein sequestration; (3) the *Xist* interactome contains components of paraspeckles and stress granules and is significantly enriched for structurally disordered proteins with a strong propensity for phase separation; and (4) binding partners of *Xist* and *Neat1* diffuse in a liquid-like manner. It will be pivotal to better understand if and how *Xist* undergoes phase separation with its protein partners, and we hope that our hypothesis will stimulate work in this direction. ❐

Andrea Cerase[1,7,8*], Alexandros Armaos[2,8], Christoph Neumayer[3], Philip Avner[1], Mitchell Guttman[3*] and Gian Gaetano Tartaglia[2,4,5,6*]

*[1]EMBL-Rome, Monterotondo, Italy. [2]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain. [3]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [4]ICREA and UPF, Barcelona, Spain. [5]Department of Biology 'Charles Darwin', Sapienza University of Rome, Rome, Italy. [6]Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, Genoa, Italy. [7]Present address: Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [8]These authors contributed equally: Andrea Cerase, Alexandros Armaos. *e-mail: a.cerase@qmul.ac.uk; mguttman@lncrna.caltech.edu; gian.tartaglia@crg.eu*

Published online: 6 May 2019
https://doi.org/10.1038/s41594-019-0223-0

References

1. Chujo, T. et al. *EMBO J.* **36**, 1447 – 1462 (2017).
2. Maharana, S. et al. *Science* **360**, 918 – 921 (2018).
3. Yamazaki, T. et al. *Mol. Cell* **70**, 1038 – 1053.e1037 (2018).
4. Cid-Samper, F. et al. *Cell Rep.* **25**, 3422 – 3434.e3427 (2018).
5. Mao, Y. S., Sunwoo, H., Zhang, B. & Spector, D. L. *Nat. Cell Biol.* **13**, 95 – 101 (2011).
6. West, J. A. et al. *J. Cell Biol.* **214**, 817 – 830 (2016).
7. Shin, Y. & Brangwynne, C. P. *Science* **357**, eaaf4382 (2017).
8. Bolognesi, B. et al. *Cell Rep.* **16**, 222 – 231 (2016).
9. Tartaglia, G. G. et al. *J. Mol. Biol.* **380**, 425 – 436 (2008).
10. Cerase, A. et al. *Proc. Natl Acad. Sci. USA* **111**, 2235 – 2240 (2014).
11. Smeets, D. et al. *Epigenetics Chromatin* **7**, 8 (2014).
12. Cirillo, D. et al. *Nat. Methods* **14**, 5 – 6 (2016).
13. Markmiller, S. et al. *Cell* **172**, 590 – 604.e513 (2018).
14. Pintacuda, G., Young, A. N. & Cerase, A. *Front. Mol. Biosci.* **4**, 90 (2017).
15. Delli Ponti, R., Marti, S., Armaos, A. & Tartaglia, G. G. *Nucleic Acids Res.* **45**, e35 (2017).
16. Van Nostrand, E. L. et al. *Nat. Methods* **13**, 508 – 514 (2016).
17. Naganuma, T. et al. *EMBO J.* **31**, 4020 – 4034 (2012).
18. Jain, S. et al. *Cell* **164**, 487 – 498 (2016).
19. Cerase, A., Pintacuda, G., Tattermusch, A. & Avner, P. *Genome Biol.* **16**, 166 (2015).
20. Pintacuda, G. et al. *Mol. Cell* **68**, 955 – 969.e910 (2017).
21. Klus, P. et al. *Bioinformatics* **30**, 1601 – 1608 (2014).
22. Ng, K. et al. *Mol. Biol. Cell* **22**, 2634 – 2645 (2011).
23. Almeida, M. et al. *Science* **356**, 1081 – 1084 (2017).
24. Engreitz, J. M. et al. *Science* **341**, 1237973 (2013).
25. Zylicz, J. J. et al. *Cell* **176**, 182 – 197.e123 (2019).
26. Isono, K. et al. *Dev. Cell* **26**, 565 – 577 (2013).
27. Chen, C. K. et al. *Science* **354**, 468 – 472 (2016).
28. Wutz, A. & Jaenisch, R. *Mol. Cell* **5**, 695 – 705 (2000).
29. Csankovszki, G., Nagy, A. & Jaenisch, R. *J. Cell Biol.* **153**, 773 – 784 (2001).
30. Moindrot, B. et al. *Cell Rep.* **12**, 562–572 (2015).

Acknowledgements

We thank all members of the Avner, Tartaglia and Guttman groups, as well as Greta Pintacuda and Kathrin Plath for critical reading of the manuscript. P.A. was funded by an EMBL grant (50800), and A.C. was funded by an EMBL fellowship and a Rett Syndrome Research Trust (RSRT) grant. The research leading to these results has been supported by the European Research Council (RIBOMYLOME_309545), the Spanish Ministry of Economy, Industry, and Competitiveness (BFU2014-55054-P and BFU2017-86970-P) and "Fundació La Marató de TV3" (PI043296). M.G. was funded by a Caltech grant. We acknowledge support from the Spanish Ministry of Economy, Industry and Competitiveness (MEIC) to the EMBL partnership, the Centro de Excelencia Severo Ochoa and the CERCA Programme/Generalitat de Catalunya.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41594-019-0223-0.