TITLE: cGAS-mediated induction of type I interferon due to inborn errors of histone pre-mRNA processing

Carolina Uggenti[1], Alice Lepelley[2†], Marine Depp[1†], Andrew P Badrock[1†], Mathieu P Rodero[2], Marie-Thérèse el-Daher[1], Gillian I Rice[3], Somdutta Dhir[1], Ann P Wheeler[4], Ashish Dhir[1], Waad Albawardi[4], Marie-Louise Frémond[2], Luis Seabra[2], Jennifer Doig[1], Natalie Blair[1], Maria José Martin-Niclos[2], Erika Della Mina[1], Alejandro Rubio-Roldan[5], Jose L Garcia-Perez[4,5], Duncan Sproul[4,6], Jan Rehwinkel[7], Jonny Hertzog[7], Anne Boland-Auge[8], Robert Olaso[8], Jean-François Deleuze[8], Julien Baruteau[9], Karine Brochard[10], Jonathan Buckley[11], Vanessa Cavallera[12], Cristina Cereda[13], Liesbeth M De Waele[14], Angus Dobbie[15], Diane Doummar[16], Frances Elmslie[17], Margarete Koch-Hogrebe[18], Ram Kumar[19], Kate Lamb[20], John H Livingston[21], Anirban Majumdar[22], Charles Marques Lorenço[23], Simona Orcesi[12,24], Sylviane Peudenier[25], Kevin Rostasy[18], Caroline A Salmon[26], Christiaan Scott[27], Davide Tonduti[28], Guy Touati[29], Marialuisa Valente[13], Hélio van der Linden Jr[30], Hilde Van Esch[31], Marie Vermelle[32], Kate Webb[27], Andrew P Jackson[4], Martin AM Reijns[4], Nick Gilbert[4], Yanick J Crow[1,2,33*]

[†] These authors contributed equally to this work.

[*] Corresponding author. Email: yanickcrow@mac.com

1. Centre for Genomic and Experimental Medicine, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK.

2. Laboratory of Neurogenetics and Neuroinflammation, Institut Imagine, Paris, France.

3. Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK.

4. MRC-Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK.

5. GENYO. Centre for Genomics and Oncological Research: Pfizer/University of Granada/Andalusian Regional Government, PTS Granada, Spain.

6. Edinburgh Cancer Research Centre, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh, UK.

7. MRC Human Immunology Unit, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK.

8. Université Paris-Saclay, CEA, Centre National de Recherche en Génomique Humaine, 91057, Evry, France.

9. Metabolic Department, Great Ormond Street Hospital for Children, London, UK.

10. Service de Médecine interne Néphrologie pédiatrique, Hôpital des enfants, Toulouse, France.

11. Department of Paediatric Nephrology, University of Cape Town, Red Cross War Memorial Children's Hospital, Cape Town, Republic of South Africa.

12. Child Neurology and Psychiatry Unit, IRCCS Mondino Foundation, Pavia, Italy.

13. Genomic and post-Genomic Center, IRCCS Mondino Foundation, Pavia, Italy.

14. Department of Development and Regeneration, KU Leuven, Paediatric Neurology, University Hospitals Leuven, Leuven, Belgium.

15. Yorkshire Clinical Genetics Service, Chapel Allerton Hospital, Leeds, UK.

16. AP-HP, GHUEP, Département de neuropédiatrie, Centre de référence de neurogénétique et mouvements anormaux de l'enfant, Hôpital Armand Trousseau, Paris, France.

17. South West Thames Regional Genetics Service, St George's, University of London, London, UK.

18. Department of Paediatric Neurology, Children`s Hospital Datteln, University Witten/Herdecke, Dattein, Germany.

19. Department of Paediatric Neurology, Alder Hey Children’s NHS Foundation Trust, Liverpool, UK.

20. Department of Paediatrics, Gloucestershire Royal Hospital, Gloucester, UK.

21. Department of Paediatric Neurology, Leeds Teaching Hospitals NHS Trust, Leeds, UK.

22. Department of Paediatric Neurology, Bristol Children’s Hospital, Bristol, UK.

23. Faculdade de Medicina - Centro Universitário Estácio de Ribeirão Preto, Brazil.

24. Department of Brain and Behavioural Sciences, University of Pavia, Pavia, Italy.

25. Centre de Référence des Déficiences intellectuelles de causes rares et polyhandicap, CHRU de Brest, Brest, France.

26. Department of Paediatrics, Royal Surrey County Hospital, Guildford, UK.

27. University of Cape Town, Red Cross War Memorial Children's Hospital, Republic of South Africa.

28. COALA (Center for diagnosis and treatment of Leukodystrophies), Pediatric Neurology Unit, V. Buzzi Children’s Hospital, Milano

29. Reference Center for Inborn errors of Metabolism- Department of Pediatrics, Hôpital des Enfants - CHU de Toulouse, Toulouse, France.

30. Department of Paediatric Neurology, Neurologic Institute of Goiania, Goiania, Brazil.

31. Center for Human Genetics, University Hospitals Leuven, KU Leuven, Leuven, Belgium.

32. Department of Paediatrics, Centre hospitalier de Dunkerque, Dunkerque, France.

33. Université de Paris, Paris, France.

ABSTRACT

Inappropriate stimulation or defective negative regulation of the type I interferon response can lead to autoinflammation. In genetically uncharacterized cases of the type I interferonopathy Aicardi-Goutières syndrome, we identified biallelic mutations in *LSM11* and *RNU7-1*, encoding components of the replication-dependent histone pre-mRNA processing complex. Mutations were associated with the misprocessing of canonical histone transcripts, and a disturbance of linker histone stoichiometry. Additionally, we observed an altered distribution of nuclear cyclic GMP-AMP synthase (cGAS), and enhanced interferon signaling mediated by the cGAS-stimulator of interferon genes (STING) pathway in patient fibroblasts. Finally, we established that chromatin without linker histone more efficiently stimulates cGAS production *in vitro*. We conclude that nuclear histones, as key constituents of chromatin, are essential in suppressing the immunogenicity of self-DNA.

The reliable discrimination of self from non-self is fundamental to immunological homeostasis. Given that viral nucleic acids represent primary drivers of the host antiviral response, a challenge exists to ensure that self-derived DNA and RNA is not misrepresented as foreign[1]. Molecular definition of the type I interferonopathies, Mendelian autoinflammatory disorders associated with a persistent enhancement of type I interferon signaling, has highlighted the importance of such discrimination through the effective processing and regulated sensing of endogenous nucleic acids[2].

The recent observation that damaged DNA within micronuclei is immunostimulatory emphasizes the nucleus *per se* as a source of nucleic acid capable of inducing an innate immune response[3-5]. Although shielded from cytoplasmic nucleic acid sensors by the nuclear membrane, the double stranded (ds) DNA sensor cyclic GMP-AMP synthase (cGAS) associates with chromatin when the nuclear membrane breaks down during mitosis[6-8]. Furthermore, cGAS enters the nucleus after nuclear membrane rupture[9,10], and can localize in the nucleus outwith mitosis[11-14]. These observations beg the question as to how the immunogenicity of nuclear DNA, potentially in direct contact with cGAS, is regulated, so as to limit the risk of auto-inflammation while maintaining the responsiveness of cGAS to DNA in the context of an anti-viral response.

We have previously shown that mutations in any of the genes encoding the DNA exonuclease TREX1 (AGS1)[15], the three components of the RNase H2 endonuclease complex (RNase H2A (AGS4); RNase H2B (AGS2); RNase H2C (AGS3)[16], the deoxynucleoside triphosphate triphosphohydrolase SAMHD1 (AGS5)[17], the double-stranded RNA (dsRNA) editing enzyme ADAR (AGS6)[18], and the cytosolic dsRNA sensor MDA5 (AGS7)[19] result in a spectrum of type I interferonopathy phenotypes including Aicardi-Goutières syndrome (AGS)[20]. We now describe mutations in components of the replication-dependent histone (RDH) pre-mRNA processing complex as two further AGS-related genotypes. These mutations lead to a dysregulation of histone mRNA transcripts, a disturbance of histone protein composition, and the induction of an interferon response through a cGAS-stimulator of interferon genes (STING) dependent pathway.

RESULTS

Biallelic mutations in *LSM11* and *RNU7-1* in patients with AGS

Within our cohort of individuals exhibiting typical features of AGS ascertained over a period of 20 years, a subset of patients has remained genetically undefined. Whole exome sequencing of two such affected siblings, born to first cousin parents of Pakistani origin, identified a homozygous c.631G>A transition in *LSM11*, leading to the substitution of an evolutionarily conserved glycine at codon 211 for a serine (p.(Gly211Ser)) in the encoded protein LSM11 (Fig. 1, a and b). This change is not recorded in publicly available databases, and is predicted as damaging by *in silico* algorithms (table S1). LSM11 forms part of the U7 small nuclear ribonucleoprotein (snRNP) complex involved in the processing of RDH pre-mRNAs (Fig. 1c)[21]. We did not observe any other patients with pathogenic variants in *LSM11*. However, using whole genome and targeted Sanger sequencing, 16 patients belonging to 11 independent pedigrees were found to harbor biallelic variants, frequency of ≤ 0.005 alleles in the Genome Aggregation Database (gnomAD), in the gene *RNU7-1* encoding the snRNA U7 (Fig. 1d, fig. S1 and table S2). Parental genotypes and familial segregation were consistent with autosomal recessive inheritance (fig. S2). A further seven molecularly unresolved probands demonstrating a convincing AGS phenotype were negative for rare variants in both *LSM11* and

RNU7-1. Screening of a panel of 663 controls did not identify any individuals with biallelic rare variants in *RNU7-1* (0 of 663 controls versus 11 of 18 probands: $p < 5 \times 10^{-6}$), and no control individual carried a variant not already recorded in gnomAD (four of 22 patient alleles were novel, versus 0 of 1326 control alleles: Chi-squared $p < 5 \times 10^{-6}$)(tables S3 and S4). Importantly, we observed putative mutant alleles to be shared across our cohort; specifically, four rare variants were seen in two or more families, with the n.28C>T transition, a substitution previously demonstrated to abrogate U7 processing of a chimeric mouse histone H4 in *Xenopus laevis* oocytes[22], present in seven pedigrees (seven of 22 proband-related alleles versus 16 of 31,392 control alleles on gnomAD; Chi-squared $p < 5 \times 10^{-6}$). These genetic data indicate that mutations in *LSM11* and *RNU7-1* represent novel causes of AGS (tables S5 and S6), referred to here as AGS8 and AGS9 respectively. Although AGS9 is a rare autosomal recessive trait, we recorded neither parental consanguinity nor mutational homozygosity for this genotype, suggesting that it is the combination of a severe and a mild mutation which leads to disease, with biallelic severe mutations likely resulting in embryonic lethality, and the combination of two milder variants causing no phenotype or a phenotype not ascertained here. This situation is reminiscent of Mendelian disease-causing mutations observed in another small RNA, U8[23].

Dysregulation of histone RNA processing due to *LSM11* and *RNU7-1* mutations

RDH genes encoding the four core histones (H2A, H2B, H3 and H4) and the linker histone family (H1) differ from all other genes since they contain no introns and encode the only known eukaryotic mRNAs without a poly(A) tail at the 3´ end, terminating instead with a stem-loop that has been conserved in all metazoans[24]. This unique 3´ end is formed by endo-nucleolytic cleavage of the pre-mRNAs between the stem-loop and a purine-rich sequence termed the histone downstream element (HDE)[25,26] (Fig. 1c). Two transacting factors, stem-loop binding protein (SLBP)[27,28] and the U7 snRNP[29], binding to the stem-loop and HDE respectively, are essential for the processing of RDH pre-mRNAs. A failure of histone pre-mRNA processing can lead to the production of aberrant mRNA isoforms that possess a poly(A) tail due to the use of cryptic polyadenylation signals downstream of the U7-dependent cleavage site[30].

To investigate the functional consequences of the *LSM11* and *RNU7-1* variants that we identified, we performed reverse transcription quantitative PCR (RT-qPCR) in patient and control fibroblasts. Polyadenylated RNA was reverse transcribed, and cDNA amplified using a primer spanning the endo-nucleolytic cleavage site between the stem-loop and HDE to determine the presence of misprocessed transcripts (Fig. 2a). Compared to controls, patient cells were significantly enriched for misprocessed (i.e. aberrantly polyadenylated) forms of RDH mRNAs derived from different histone clusters[31] (Fig. 2b). Similar changes were observed when we performed siRNA knockdown of *LSM11* in HEK293T (Fig. 2c), HCT116 and U2OS (fig. S3) cells. Using random hexamers for reverse transcription, we then examined the expression of all replication-dependent linker histone genes and a set of core histone genes in patient cells, compared to controls and to other AGS genotypes. Whilst equivalent mRNA levels derived from the replication-independent histone genes *H1FX*, *H1F0* and *H3F3A*, transcripts which are normally polyadenylated, were seen in all four groups, we recorded important differences in the expression of linker and core histone mRNAs in cells from patients with mutations in *LSM11* or *RNU7-1* (Fig. 2d). In contrast to the other single gene-copy linker histones, levels of *HIST1H1E*, encoding H1.4, were significantly reduced in AGS8/9 patient fibroblasts.

To extend these findings we went on to assess genome-wide transcript expression in patient and control fibroblasts using RNA-Seq. In this analysis, transcripts remaining after an initial selection for polyadenylated RNA were ribo-depleted, and both poly(A)+ and poly(A)- fractions then sequenced (Fig. 3a). In doing so, we observed a global increase in polyadenylated RDH mRNA transcripts in cells from patients compared to controls, with a concomitant reduction in mature, non-polyadenylated, RDH gene transcripts (Fig. 3b). The same pattern of increased polyadenylated and reduced non-polyadenylated transcript expression was seen when linker and core histones were analyzed as sub-groups (fig. S4; table S7), and when histone genes were considered individually (and compared to replication independent genes, e.g. *H2AFJ*, whose expression did not change) (Fig. 3c). Taken together, these data indicate that the *LSM11* and *RNU7-1* variants identified in this study represent pathogenic loss-of-function mutations resulting in a disturbance of RDH pre-mRNA processing.

RDH pre-mRNA misprocessing is associated with increased type I interferon signaling

AGS has been classified as a type I interferonopathy, in which enhanced type I interferon production is considered as directly relevant to pathogenesis. Using a previously validated *ex vivo* assay of interferon signaling status[32], we recorded an upregulation of interferon stimulated gene (ISG) expression in the blood of all 16 patients tested (Fig. 4a), with an abnormal result seen on 44 of 45 occasions, including eight patients assessed three or more times over a period of greater than two years (Fig. 4b, fig. S5). Similarly, we observed increased expression of ISGs in patient fibroblasts, determined by targeted RT-qPCR (Fig. 4c), protein levels (Fig. 4d) and gene set enrichment analysis of the poly(A)+ fraction of our RNA-Seq experiment (fig. S6). ISG expression fell following the treatment of patient fibroblasts with the TANK-binding kinase 1 (TBK1) chemical inhibitor BX795 and the JAK1/2 inhibitor ruxolitinib, which block interferon induction and type I interferon receptor signaling respectively (fig. S7).

Increased type I interferon signaling is dependent on cGAS and STING

Having identified a link between mutations in the U7-snRNP complex, histone pre-mRNA misprocessing and up-regulated interferon signaling, we wished to explore the pathway mediating the observed interferon induction. In mammalian cells, immune responses to viral infection involve host-encoded nucleic acid-binding pattern-recognition receptors. These include endosomal Toll-like receptors (TLRs), the cytosolic sensors of RNA (melanoma differentiation associated protein 5, MDA5, and retinoic acid inducible gene I, RIG-I, both of which signal through mitochondrial antiviral signaling protein (MAVS)) and dsDNA sensors, most particularly cGAS signaling through STING[33,34].

Lentiviral transduction of THP-1 cells, a human monocytic cell line, with two different shRNAs targeting *LSM11*, was associated with a reduction in *LSM11* RNA (Fig. 5a), and recapitulated an increase in misprocessed RDH mRNAs (assessed using random hexamers for reverse transcription and primers spanning the endonucleolytic site; Fig. 5b). Further consistent with data in patient cells, these changes were accompanied by the induction of ISGs and interferon beta (Fig. 5c; fig. S8). We then compared the results of knockdown of *LSM11* in wild-type (WT) THP-1 cells, and in functionally characterized THP-1 cells null for *cGAS*, *STING* or *MAVS*. The enhanced interferon signaling seen in WT cells was preserved in THP-1 cells null for *MAVS* (Fig. 5d; fig. S9). In contrast, the absence of either cGAS or STING abrogated the enhanced interferon response seen in WT THP-1 cells following *LSM11* knockdown. Changes in RNA expression were mirrored by ISG15 protein levels and phospho-STING (Fig. 5e; fig. S9; fig. S10).

To extend these findings to patient cells, we used siRNA knockdown of the key anti-viral sensors in fibroblasts from individuals with mutations in *LSM11* or *RNU7-1*. As in THP-1 cells, we observed an abrogation of ISG induction upon knockdown of *cGAS* and *STING*, whilst knockdown of *MAVS*, and of myeloid differentiation primary response 88 (*MYD88*), which transduces endosomal signaling of DNA via TLR9 (and RNA by TLR7/8), had no such effect (Fig. 5f; fig. S11; fig. S12). Concordant with these data, we recorded increased levels of the second messenger cyclic GMP-AMP (cGAMP), synthesized by cGAS upon binding of stimulatory DNA, in patient fibroblasts compared to control cells (Fig. 5g).

Mutations in *LSM11* and *RNU7-1* are associated with a disturbance of chromatin-bound histone stoichiometry and an altered distribution of nuclear cGAS

The above experiments indicated that it was DNA, rather than RNA, that was responsible for increased interferon signaling upon dysregulation of histone pre-mRNA processing. This led us to consider if a perturbation of cellular histone levels and chromatin structure might result in cGAS induction. To explore this possibility, we first isolated protein from the nuclear fraction of control and patient fibroblasts. By Coomassie gel staining, we noted an overall reduction of linker histones in patient compared to control cells, while the level of core histones was not obviously affected (fig. S13). Using histone sub-type specific antibodies, we went on to measure the level of histone proteins in fibroblasts by western blot. Again, we observed no difference in the total level of core histones in patient and control cells (Fig. 6a). In contrast, an examination of chromatin-bound proteins revealed a marked reduction in the ratio of linker H1.4 compared to H1.2 in patient cells (Fig. 6b; fig. S14), with *HIST1H1E* (encoding H1.4) and *HIST1H1C* (encoding H1.2) together representing the majority of linker histone transcripts expressed in fibroblasts (fig. S4; table S7). Notably then, knockdown of *H1.4* in THP-1 cells led to a robust induction of ISGs (Fig. 6c), suggesting that reduced levels of this linker histone might contribute to cGAS-STING dependent inflammatory signaling.

The nuclear localisation of cGAS implies that genomic DNA is somehow rendered immunologically inert. Of importance, we have previously shown that cGAS can bind to, and be activated by, chromatin *in vitro*[5]. However, while nucleosomes have a higher affinity for cGAS than naked DNA, they apparently confer a relative inhibition of DNA-dependent cGAS activation[8]. We therefore postulated that the histone composition of nuclear DNA might affect cGAS binding and/or activation. Consistent with this, wide-field microscopy and immunofluorescence demonstrated an altered nuclear distribution of cGAS in AGS8/9 fibroblasts compared to control cells (Fig. 6d). Patient fibroblasts also exhibited an increased frequency of misshapen nuclei (Fig. 6e), including nuclear membrane herniations which were sometimes associated with the presence of intense foci of cGAS (Fig. 6f). Considering the increased level of cGAMP in patient cells (Fig. 5g), and the induction of ISGs upon *H1.4* knockdown in THP-1 cells (Fig. 6c), we assembled chromatin *in vitro* with or without linker histone. Consistent with our previous findings[5], recombinant cGAS produced cGAMP in response to chromatin in the absence of linker histone. In contrast, after incubation with synthetic chromatin prepared in the presence of linker histone, cGAMP levels did not reach above those seen for the negative control (Fig. 6g).

CONCLUSIONS

Here we demonstrate that autosomal recessive, loss-of-function mutations in *LSM11* and *RNU7-1*, encoding two components of the canonical RDH pre-mRNA processing complex, result in a phenotype consistent with the type I interferonopathy AGS. We provide three lines of evidence to support this assertion: 1. The observation of rare, biallelic sequence variants in affected individuals, and the sharing of mutations across unrelated pedigrees; 2. The finding of a disturbance of RDH mRNA expression and histone protein status in patient compared to control fibroblasts. 3. The demonstration that knockdown of *LSM11* in THP-1 cells results in an increase in misprocessed RDH mRNA and interferon signaling.

Given a defect in RDH pre-mRNA processing, we initially hypothesized that interferon would be induced through an RNA-mediated pathway, resulting from an accumulation of polyadenylated histone pre-mRNAs. However, experiments in THP-1 cells and patient fibroblasts indicated that aberrant interferon signaling was transduced by DNA via cGAS and STING, rather than through the cytosolic RNA signaling accessory protein MAVS. Taken together with the observation of a disturbance in the ratio of linker histone proteins in patient fibroblasts, these data led us to consider the mechanism(s) by which changes in nuclear histone protein composition might trigger chronic type I interferon signaling.

Combined depletion of H1.1 and H1.4 has been shown to result in a type I interferon response in T47D cells, suggested to be driven by an increase in the expression of heterochromatic repeats or endogenous retroelements[35]. However, we found no evidence of increased LINE-1 element expression in patient fibroblasts, nor of increased retroelement-derived transcripts in RNA-Seq data (fig. S15). As an alternative possibility, we took account of recent observations that the dsDNA receptor cGAS is present in the nucleus[14], that cGAS can bind to and be activated by chromatin[5], and that nucleosomes limit the activation of cGAS by naked DNA[8]. In doing so, we recorded an abnormal distribution of nuclear cGAS in patient fibroblasts, and *in vitro* cGAMP production by recombinant cGAS when exposed to chromatin lacking linker histone, but not chromatin containing linker histone. Altogether, these data support the possibility that chromatin-bound histone stoichiometry plays an essential physiological role in limiting cGAS activation in the nucleus (fig. S16).

ACKNOWLEDGEMENTS

We gratefully acknowledge help and advice from Wendy Bickmore, Francesca Taglini, Patricia Heyn, Tamir Chandra, Ray O'Keefe, Zbigniew Dominski, William Marzluff, Laura Murphy, Luis Sanchez-Pulido and Chris Ponting. We acknowledge the help of the Genetics Core, Edinburgh Clinical Research Facility, University of Edinburgh. **Funding**: YJC acknowledges the European Research Council (GA309449 and 786142-E-T1IFNs), a state subsidy managed by the National Research Agency (France) under the 'Investments for the Future' programme bearing the reference ANR-10-IAHU-01, and the NIHR UK Rare Genetic Disease Research Consortium. The project was supported by MSDAVENIR (Devo-Decode Project) and the Foundation Maladies Rare (GenOmics of rare diseases 2016-1). APJ and MAMR are supported by a UK Medical Research Council Human Genetics Unit core grant (MRC, U127580972). JR acknowledges the UK Medical Research Council (MRC core funding of the MRC Human Immunology Unit). JH was supported by the European Commission under the Horizon 2020 programme. We acknowledge the Edinburgh Super-Resolution Imaging Consortium (ESRIC), supported by the Wellcome Trust (grant number 208345/Z/17/Z). **Author contributions**: C.U., A.L., M.D., A.P.B., M.P.R., M-T.e-D., G.I.R., A.P.W., A.D., W.A., M-L.F., L.S., J.D., N.B., M.J.M-N., E.DM., D.S., J.R., J.H., M.A.M.R., N.G. performed experiments; S.D., A.R-R., and J.L.G-P. performed bioinformatic analyses; J-F.D., A.B-A., and R.O. were responsible for genome sequencing; J.B., K.B., J.B., V.C., C.C., L.M.DW., A.D., D.D., F.E., M.K-H., R.K., K.L., J.H.L., A.M., C.ML., S.O., S.P., K.R., C.A.S., C.S., D.T., G.T., M.V., H.vdL., H.VE., M.V., K.W., ascertained and diagnosed clinical cases. C.U.,

A.P.J., M.A.M.R., N.G., and Y.J.C. wrote and edited the manuscript. Y.J.C. was responsible for overall supervision of the project. **Competing interests**: The authors declare no competing interests.

REFERENCES

1. Roers, A., Hiller, B. & Hornung, V. Recognition of Endogenous Nucleic Acids by the Innate Immune System. *Immunity* **44**, 739-54 (2016).
2. Uggenti, C., Lepelley, A. & Crow, Y.J. Self-Awareness: Nucleic Acid-Driven Inflammation and the Type I Interferonopathies. *Annu Rev Immunol* **37**, 247-267 (2019).
3. Bartsch, K. *et al.* Absence of RNase H2 triggers generation of immunogenic micronuclei removed by autophagy. *Hum Mol Genet* **26**, 3960-3972 (2017).
4. Harding, S.M. *et al.* Mitotic progression following DNA damage enables pattern recognition within micronuclei. *Nature* **548**, 466-470 (2017).
5. Mackenzie, K.J. *et al.* cGAS surveillance of micronuclei links genome instability to innate immunity. *Nature* **548**, 461-465 (2017).
6. Gentili, M. *et al.* The N-Terminal Domain of cGAS Determines Preferential Association with Centromeric DNA and Innate Immune Activation in the Nucleus. *Cell Rep* **26**, 3798 (2019).
7. Yang, H., Wang, H., Ren, J., Chen, Q. & Chen, Z.J. cGAS is essential for cellular senescence. *Proc Natl Acad Sci U S A* **114**, E4612-E4620 (2017).
8. Zierhut, C. *et al.* The Cytoplasmic DNA Sensor cGAS Promotes Mitotic Cell Death. *Cell* **178**, 302-315.e23 (2019).
9. Denais, C.M. *et al.* Nuclear envelope rupture and repair during cancer cell migration. *Science* **352**, 353-8 (2016).
10. Raab, M. *et al.* ESCRT III repairs nuclear envelope ruptures during cell migration to limit DNA damage and cell death. *Science* **352**, 359-62 (2016).
11. Jiang, H. *et al.* Chromatin-bound cGAS is an inhibitor of DNA repair and hence accelerates genome destabilization and cell death. *EMBO J* **38**, e102718 (2019).
12. Lahaye, X. *et al.* NONO Detects the Nuclear HIV Capsid to Promote cGAS-Mediated Innate Immune Activation. *Cell* **175**, 488-501.e22 (2018).
13. Liu, H. *et al.* Nuclear cGAS suppresses DNA repair and promotes tumorigenesis. *Nature* **563**, 131-136 (2018).
14. Volkman, H.E., Cambier, S., Gray, E.E. & Stetson, D.B. Tight nuclear tethering of cGAS is essential for preventing autoreactivity. *Elife* **8**, e47491 (2019).
15. Crow, Y.J. *et al.* Mutations in the gene encoding the 3'-5' DNA exonuclease TREX1 cause Aicardi-Goutières syndrome at the AGS1 locus. *Nat Genet* **38**, 917-20 (2006).
16. Crow, Y.J. *et al.* Mutations in genes encoding ribonuclease H2 subunits cause Aicardi-Goutières syndrome and mimic congenital viral brain infection. *Nat Genet* **38**, 910-6 (2006).
17. Rice, G.I. *et al.* Mutations involved in Aicardi-Goutières syndrome implicate SAMHD1 as regulator of the innate immune response. *Nat Genet* **41**, 829-32 (2009).
18. Rice, G.I. *et al.* Mutations in ADAR1 cause Aicardi-Goutières syndrome associated with a type I interferon signature. *Nat Genet* **44**, 1243-8 (2012).
19. Rice, G.I. *et al.* Gain-of-function mutations in IFIH1 cause a spectrum of human disease phenotypes associated with upregulated type I interferon signaling. *Nat Genet* **46**, 503-509 (2014).
20. Crow, Y.J. & Manel, N. Aicardi-Goutières syndrome and the type I interferonopathies. *Nat Rev Immunol* **15**, 429-40 (2015).

21. Pillai, R.S. *et al.* Unique Sm core structure of U7 snRNPs: assembly by a specialized SMN complex and the role of a new component, Lsm11, in histone RNA processing. *Genes Dev* **17**, 2321-33 (2003).
22. Kolev, N.G. & Steitz, J.A. In vivo assembly of functional U7 snRNP requires RNA backbone flexibility within the Sm-binding site. *Nat Struct Mol Biol* **13**, 347-53 (2006).
23. Badrock, A.P. *et al.* Analysis of U8 snoRNA Variants in Zebrafish Reveals How Bi-allelic Variants Cause Leukoencephalopathy with Calcifications and Cysts. *Am J Hum Genet* **106**, 694-706 (2020).
24. Marzluff, W.F., Wagner, E.J. & Duronio, R.J. Metabolism and regulation of canonical histone mRNAs: life without a poly(A) tail. *Nat Rev Genet* **9**, 843-54 (2008).
25. Muller, B. & Schumperli, D. The U7 snRNP and the hairpin binding protein: Key players in histone mRNA metabolism. *Semin. Cell Dev. Biol* **8**, 567-576 (1997).
26. Dominski, Z. & Marzluff, W.F. Formation of the 3' end of histone mRNA: getting closer to the end. *Gene* **396**, 373-90 (2007).
27. Wang, Z.-F., Whitfield, M.L., Ingledue, T.I., Dominski, Z. & Marzluff, W.F. The protein which binds the 3' end of histone mRNA: a novel RNA- binding protein required for histone pre-mRNA processing. *Genes Dev* **10**, 3028-3040 (1996).
28. Martin, F., Schaller, A., Eglite, S., Schümperli, D. & Müller, B. The gene for histone RNA hairpin binding protein is located on human chromosome 4 and encodes a novel type of RNA binding protein. *EMBO J* **16**, 769-78 (1997).
29. Sabath, I. *et al.* 3'-End processing of histone pre-mRNAs in Drosophila: U7 snRNP is associated with FLASH and polyadenylation factors. *RNA* **19**, 1726-44 (2013).
30. Sullivan, E. *et al.* Drosophila stem loop binding protein coordinates accumulation of mature histone mRNA with cell cycle progression. *Genes Dev* **15**, 173-187 (2001).
31. Marzluff, W.F., Gongidi, P., Woods, K.R., Jin, J. & Maltais, L.J. The human and mouse replication-dependent histone genes. *Genomics* **80**, 487-98 (2002).
32. Rice, G.I. *et al.* Assessment of interferon-related biomarkers in Aicardi-Goutières syndrome associated with mutations in TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, and ADAR: a case-control study. *Lancet Neurol* **12**, 1159-69 (2013).
33. Chen, Q., Sun, L. & Chen, Z.J. Regulation and function of the cGAS-STING pathway of cytosolic DNA sensing. *Nat Immunol* **17**, 1142-9 (2016).
34. Streicher, F. & Jouvenet, N. Stimulation of Innate Immunity by Host and Viral RNAs. *Trends Immunol* **40**, 1134-1148 (2019).
35. Izquierdo-Bouldstridge, A. *et al.* Histone H1 depletion triggers an interferon response in cancer cells via activation of heterochromatic repeats. *Nucleic Acids Res* **45**, 11622-11642 (2017).

FIGURE LEGENDS

Fig. **1**. **Biallelic mutations in LSM11 and U7 snRNA in patients with Aicardi-Goutières syndrome (AGS)**. (**a**) Family pedigree (AGS114, F1) where the affected individuals are homozygous (hom) for a glycine to serine substitution at codon 211 in LSM11. Circles and squares indicate female and male family members respectively. Black symbols represent affected individuals. Diagonal line indicates deceased status. Double line indicates consanguinity (first cousins in this case). WT = wildtype. (**b**) CLUSTAL Omega alignment of LSM11 homologs. The position of the Gly211Ser substitution is indicated in yellow, and the human LSM11 Sm motif 1 boxed in magenta. (**c**) Schematic illustrating the U7 snRNP complex involved in the processing of replication-dependent histone (RDH) RNAs. SL = stem loop; SLBP = stem loop binding protein; HDE = histone downstream element; PAS = polyadenylation site. (**d**) Schematic of *RNU7-1* encoding the small nuclear RNA (snRNA) U7, with rare (≤ 0.005 allelic frequency on gnomAD) variants (red indicates point mutations; brackets refer to deletions) identified in AGS probands annotated according to family.

Fig. **2**. **Misprocessing of replication-dependent histone (RDH) pre-mRNAs**. (**a**) Strategy to investigate the presence of misprocessed mRNAs. Upon selection of polyadenylated RNA by oligo(dT) primed reverse transcription, primers surrounding the cleavage site were used to assess the presence of polyadenylated RDH transcripts in patient-derived primary fibroblasts (**b**) and after knockdown of *LSM11* by siRNA in HEK293T cells (**c**). (**d**) Violin plots showing the expression of total RDH mRNAs, measured by RT-qPCR, in primary fibroblasts derived from controls and from patients carrying mutations in *LSM11* (yellow), *RNU7-1* (magenta) or other known AGS-causing genes (*TREX1*, *SAMHD1*, *ADAR*, *IFIH1*; light grey). Mann-Whitney U test was used to compare differences between two groups, while differences in expression of RDH mRNA in D were assessed by Kruskal-Wallis H test, followed by Dunn's multiple comparison test. Experiments were performed with a minimum of 3 biological and/or technical replicates. Error bars indicate standard deviation (SD). P values: * = <0.05; ** = <0.01; *** = < 0.001.

Fig. **3**. **Global analysis of transcript expression using RNA-Seq**. (**a**) Schematic of the RNA-Seq sequential protocol for the analysis of poly(A)+ and poly(A)- transcripts. (**b**) Box plot showing overall expression levels of polyadenylated and non-polyadenylated RDH transcripts in patients (n = 3) and controls (n = 3). (**c**) Genome Browser view of a selected RDH mRNA (*HIST1H1C*) compared to a non-canonical histone (*H2AFJ*). Error bars represent standard deviation. *P* values were calculated using Wilcoxon rank-sum test. * = <0.05; *** = < 0.001.

Fig. **4**. **Interferon stimulated gene expression in whole blood and in fibroblasts from AGS8/9 mutation-positive patients**. (**a**) Dot plot showing the interferon score calculated from the median fold change of six interferon stimulated genes (ISGs) measured in blood. (**b**) Serial ISG expression in three affected individuals (.1, .2, .3) from AGS520 (F3) with biallelic mutations in *RNU7-1*. Numbers in brackets refer to decimalized age at sampling, followed by interferon score. Colours denote individuals, with repeat samples represented by different bars of the same colour. ISG mRNA expression in control versus patient-derived fibroblasts (**c**), and protein expression of two representative ISGs (ISG15 and MX1) assessed by western blot (**d**), with Cofilin and vinculin used as loading controls for low and high molecular weight proteins respectively. Violin plots show distribution of the raw data, with data points indicating the average value for each individual. Mann-Whitney U test was used to compare differences between two groups. Experiments were performed with a minimum of 3 biological and/or technical replicates. Error bars indicate standard deviation (SD). P values are: * = <0.05; **** = <0.0001.

Fig. 5. Interferon induction secondary to RDH pre-mRNA processing is mediated by the cGAS-STING pathway. (**a**) Knockdown of *LSM11* by lentiviral transduction of THP-1 cells using either of two shRNAs. Following knockdown of *LSM11*, RDH mRNA processing was assessed by RT-qPCR using primers surrounding the cleavage site of two representative canonical histone genes *HIST1H2AC* and *HIST2H3A* (**b**), and the expression of the ISGs *IFI27* and *IFI44L*, and of *IFNß*, measured by RT-qPCR (**c**). mRNA expression of *IFI27, IFI44L* (**d**) and expression of ISG15 protein (**e**) following knockdown of *LSM1* in THP-1 cells WT or null for *cGAS*, *STING* or *MAVS*. Cofilin was used as a housekeeping protein. The white asterisk indicates a non-specific band. (**f**) *IFIT1* and *ISG15* gene (left), and MX1 and ISG15 protein (right), expression in patient and control fibroblasts following knockdown of either *STING*, *cGAS*, *MYD88* or *MAVS*. (**g**) Measurement of cGAMP in patient and control fibroblasts assessed by ELISA. Stimulation of a BJ-5ta fibroblast cell line with 1µg/ml DNA was used as a positive control. Data points in the box plots show the value for individual replicates. Violin plots show the distribution of experimental replicates, with data points representing the average value per cell line. Mann-Whitney U test and Kruskal-Wallis H test (followed by Dunn's multiple comparison test) were used to compare differences between two or more groups respectively. Experiments were performed with a minimum of 3 biological and/or technical replicates. Error bars indicate standard deviation (SD). P values are: * = <0.05; ** = <0.01; *** = < 0.001; **** = < 0.0001. ns = not significant. Scr = scrambled, shLSM11 = short hairpin RNA against LSM11.

Fig. 6. Histone stoichiometry, nuclear morphology and cGAS distribution in patient and control fibroblasts, and *in vitro* cGAMP production. (**a**) Quantification of core histone proteins in total cell lysates and chromatin enriched fractions. (**b**) Western blot of linker histones H1.4 and H1.2 in the chromatin enriched fraction of patient and control fibroblasts. (**c**) Expression of *IFI27* and *IFI44L* following lentivirus-mediated knockdown of *H1.4* in THP-1 cells. (**d**) Quantification of cGAS signal at the nuclear periphery of patient fibroblasts. cGAS signal was measured across a set of bins (rings) equally spaced within each nuclei (schematic). The numbering of bins is from 1 (innermost) to 4 (outermost), where 4 is the total number of bins specified for each object. FracAdT_4 indicates integrated total raw intensity at ring 4. Numbers in brackets indicate numbers of cells analysed. Fluorescent microscopy shows representative nuclear localisation of cGAS (yellow) in control and patient cells. Nuclei are stained with DAPI (blue). (**e**) Solidity (area/convex area) and form factor ($4*\pi*\text{area}/\text{perimeter}^2$) measurements of patient and control fibroblast nuclei. Both parameters vary between 0 and 1, with 1 indicating perfectly circularity. (**f**) Representative DAPI stained images and immunofluorescent microscopy of control and patient-derived fibroblasts. Scale bar is 8 µm. (**g**) Quantification of cGAMP production (mean: error bars SD) by recombinant cGAS following exposure to 601 chromatin with (601 +H5) and without (601 -H5) chicken linker histone H5. Experiments were performed using 1mM NaCl, and following glutaraldehyde-induced crosslinking of DNA and chromatin. No DNA, negative control. Mann-Whitney U test and Kruskal-Wallis H test (followed by Dunn's multiple comparison test) were used to compare differences between two or more groups respectively. Experiments were performed with a minimum of 3 biological and/or technical replicates. Error bars indicate standard deviation (SD). P values are: * = <0.05; ** = <0.01; *** = < 0.001; **** = < 0.0001.

Figure 1. Biallelic mutations in LSM11 and U7 in patients with Aicardi-Goutières syndrome (AGS).

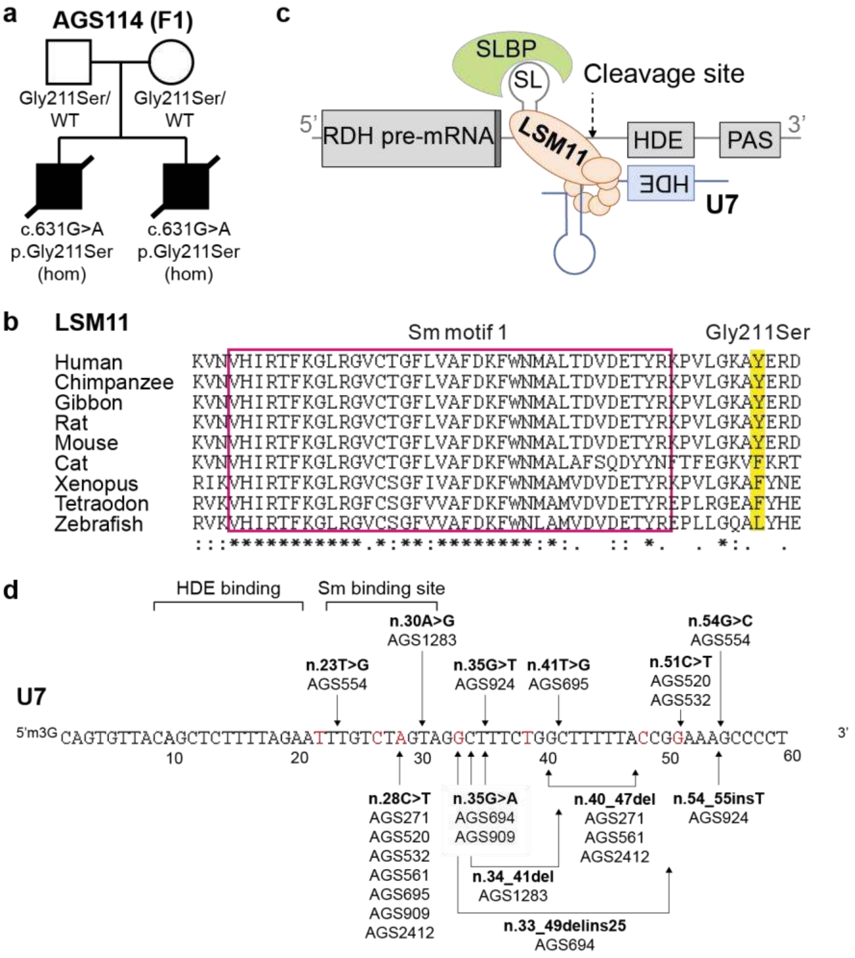

a
AGS114 (F1)
Gly211Ser/
WT
Gly211Ser/
WT
c.631G>A
p.Gly211Ser
(hom)
c.631G>A
p.Gly211Ser
(hom)
c
SLBP
SL
Cleavage site
5'
RDH pre-mRNA
LSM11
HDE
PAS
3'
HDE
U7
b
LSM11
Sm motif 1
Gly211Ser
Human KVNVHIRTFKGLRGVCTGFLVAFDKFWNMALTDVDETYRKPVLGKAYERD
Chimpanzee KVNVHIRTFKGLRGVCTGFLVAFDKFWNMALTDVDETYRKPVLGKAYERD
Gibbon KVNVHIRTFKGLRGVCTGFLVAFDKFWNMALTDVDETYRKPVLGKAYERD
Rat KVNVHIRTFKGLRGVCTGFLVAFDKFWNMALTDVDETYRKPVLGKAYERD
Mouse KVNVHIRTFKGLRGVCTGFLVAFDKFWNMALTDVDETYRKPVLGKAYERD
Cat KVNVHIRTFKGLRGVCTGFLVAFDKFWNMALAFSQDYYNFTFEGKVFKRT
Xenopus RIKVHIRTFKGLRGVCSGFIVAFDKFWNMAMVDVDETYRKPVLGKAFYNE
Tetraodon RVKVHIRTFKGLRGFCSGFVVAFDKFWNMAMVDVDETYREPLRGEAFYHE
Zebrafish RVKVHIRTFKGLRGVCSGFVVAFDKFWNLAMVDVDETYREPLLGQALYHE
:::***********.*:**:********:*:. :: *. . *:. .
d
HDE binding
Sm binding site
U7
n.30A>G
AGS1283
n.54G>C
AGS554
n.23T>G
AGS554
n.35G>T
AGS924
n.41T>G
AGS695
n.51C>T
AGS520
AGS532
5'm3G CAGTGTTACAGCTCTTTTAGAATTTGTCTAGTAGGCTTTCTGGCTTTTTACCGGAAAGCCCCT 3'
10
20
30
40
50
60
n.28C>T
AGS271
AGS520
AGS532
AGS561
AGS695
AGS909
AGS2412
n.35G>A
AGS694
AGS909
n.40_47del
AGS271
AGS561
AGS2412
n.54_55insT
AGS924
n.34_41del
AGS1283
n.33_49delins25
AGS694

Figure 2. Misprocessing of replication-dependent histone (RDH) pre-mRNAs.

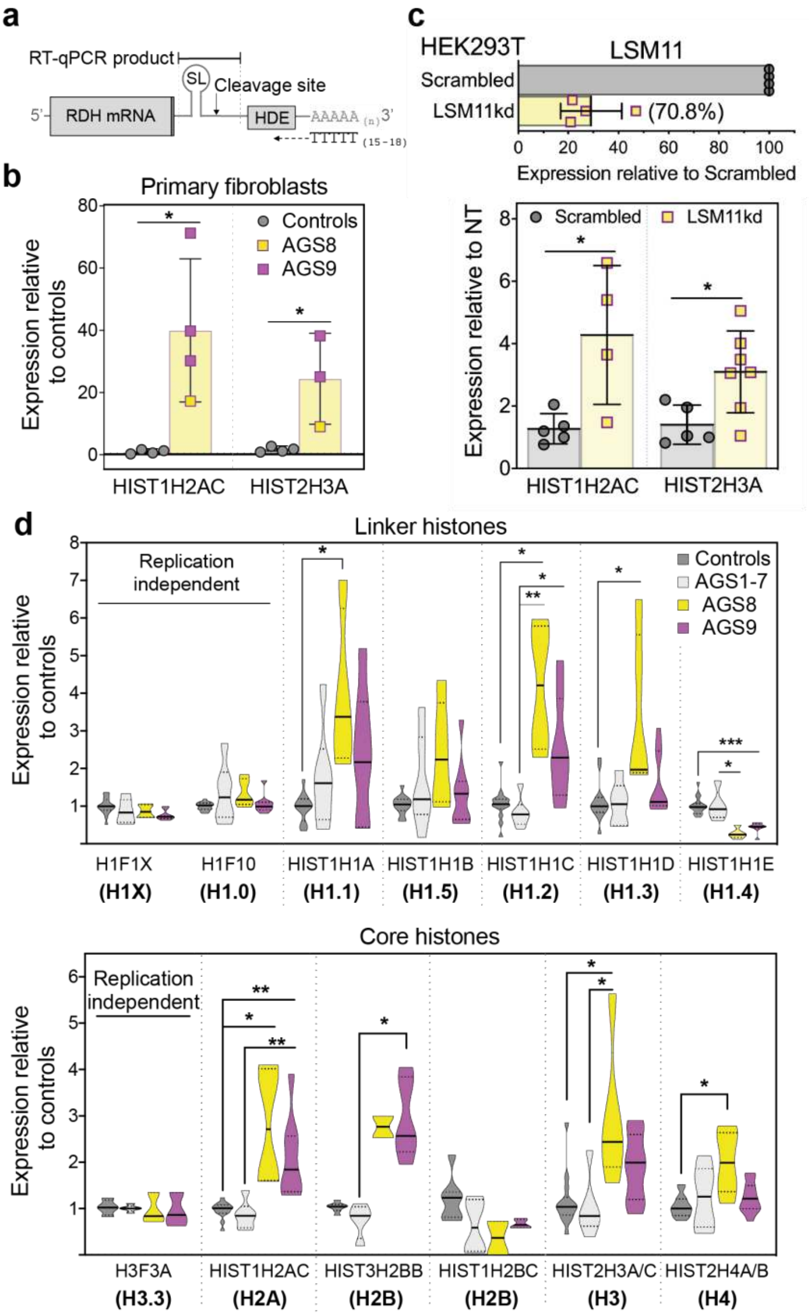


Figure 3. Global analysis of transcript expression using RNA-Seq.

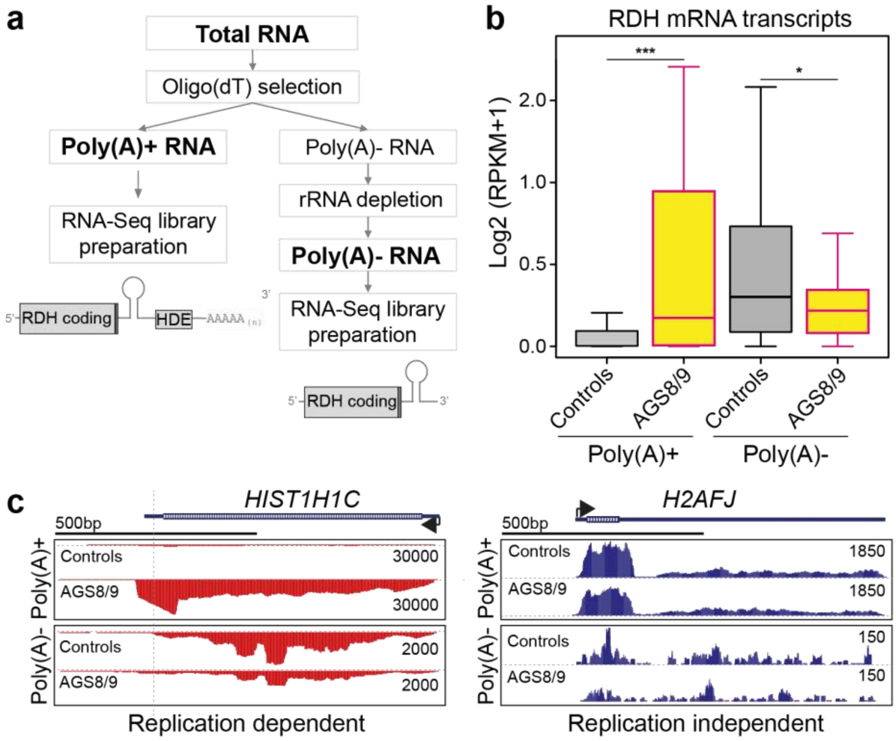


Figure 4. Interferon stimulated gene expression in whole blood and in fibroblasts from AGS8/9 mutation-positive patients.

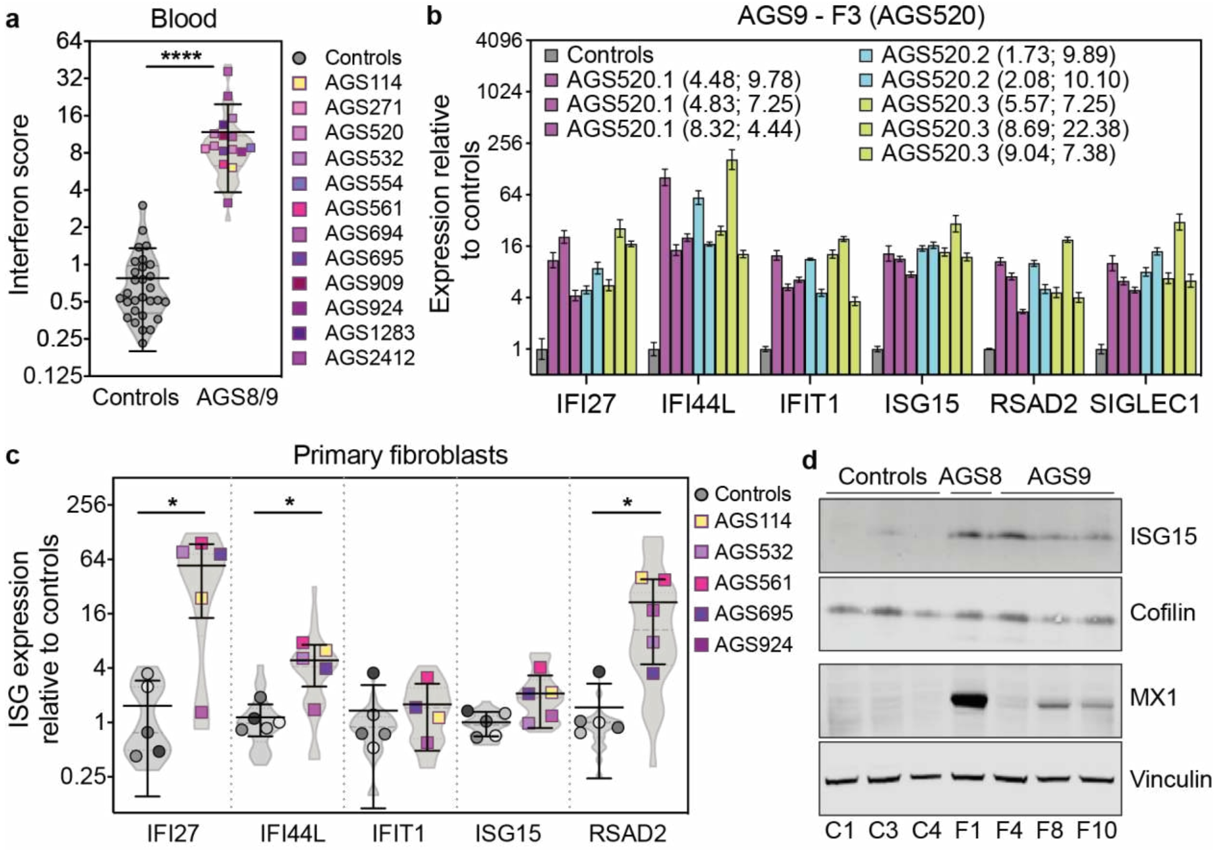
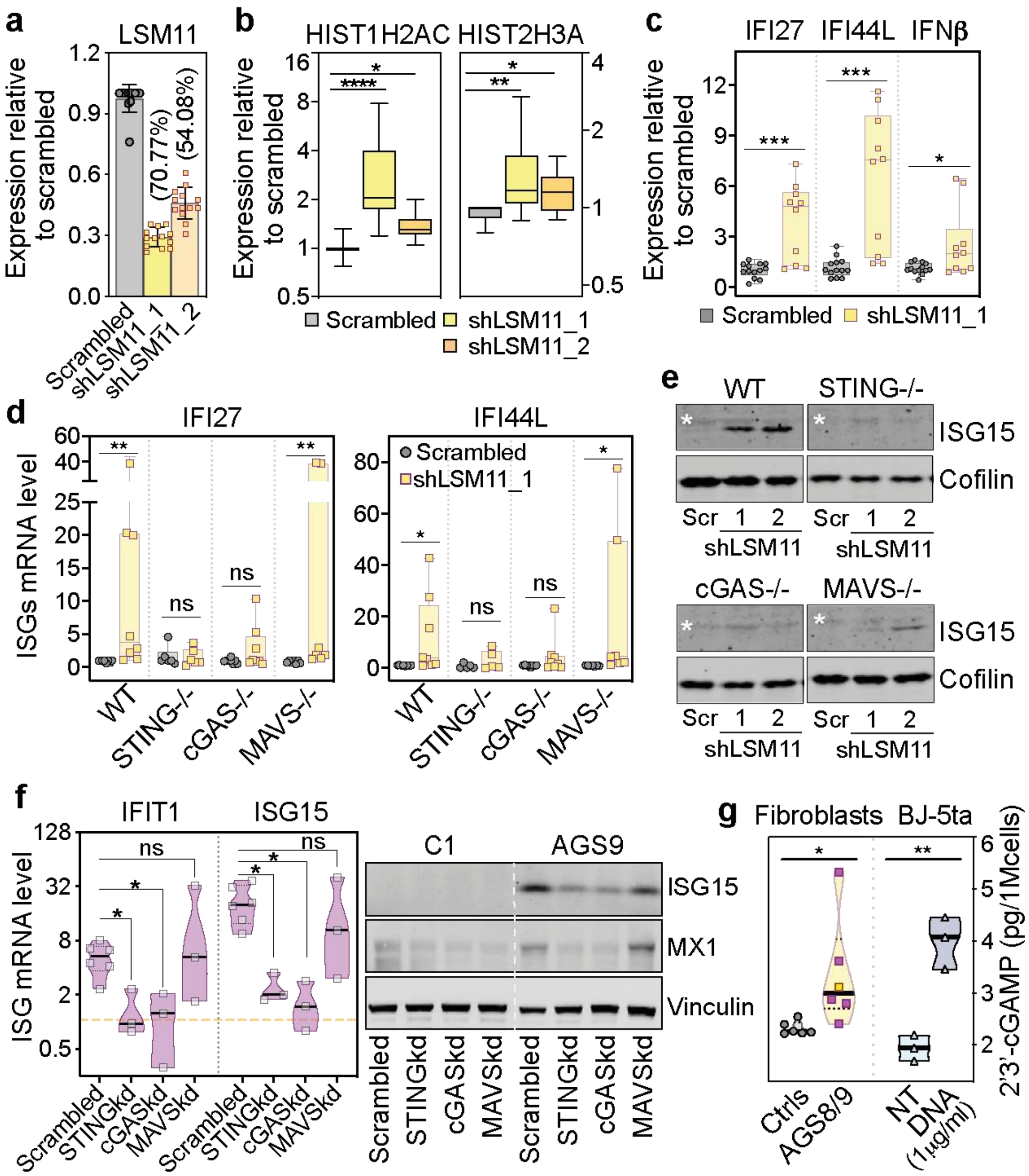
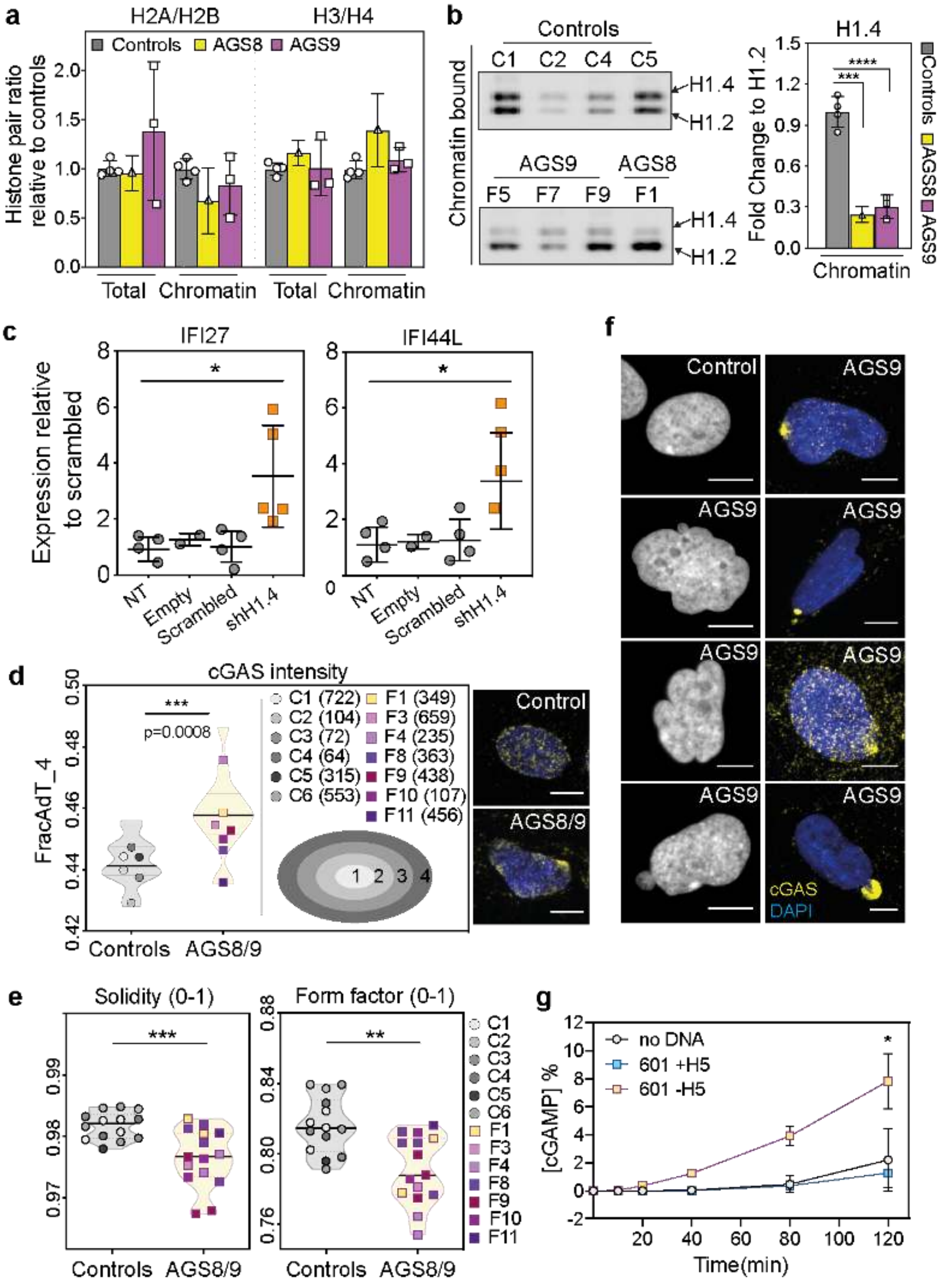


Figure 5. Interferon induction secondary to RDH pre-mRNA misprocessing is mediated by the cGAS-STING pathway.



Figure 6. Histone stoichiometry, nuclear morphology and cGAS distribution in patient and control fibroblasts, and in vitro cGAMP production.