Duplications disrupt chromatin architecture and rewire *GPR101*-enhancer communication in X-linked acrogigantism

Authors

Martin Franke, Adrian F. Daly, Leonor Palmeira, ..., Albert Beckers, Constantine A. Stratakis, Giampaolo Trivellin

Correspondence

giampaolo.trivellin@humanitasresearch.it

Franke et al., 2022, The American Journal of Human Genetics *109*, 1–18
April 7, 2022 © 2022
https://doi.org/10.1016/j.ajhg.2022.02.002

Please cite this article in press as: Franke et al., Duplications disrupt chromatin architecture and rewire *GPR101*-enhancer communication in X-linked acrogigantism, The American Journal of Human Genetics (2022), https://doi.org/10.1016/j.ajhg.2022.02.002

ARTICLE

Duplications disrupt chromatin architecture and rewire *GPR101*-enhancer communication in X-linked acrogigantism

Martin Franke,[1,13] Adrian F. Daly,[2,13] Leonor Palmeira,[3] Amit Tirosh,[4] Antonio Stigliano,[5,6] Eszter Trifan,[7] Fabio R. Faucz,[6] Dayana Abboud,[8] Patrick Petrossians,[2] Juan J. Tena,[1] Eleonora Vitali,[7,9] Andrea G. Lania,[9] José L. Gómez-Skarmeta,[1,15] Albert Beckers,[2,14] Constantine A. Stratakis,[6,10,11,14] and Giampaolo Trivellin[12,14,*]

Abstract

X-linked acrogigantism (X-LAG) is the most severe form of pituitary gigantism and is characterized by aggressive growth hormone (GH)-secreting pituitary tumors that occur in early childhood. X-LAG is associated with chromosome Xq26.3 duplications (the X-LAG locus typically includes *VGLL1*, *CD40LG*, *ARHGEF6*, *RBMX*, and *GPR101*) that lead to massive pituitary tumoral expression of *GPR101*, a novel regulator of GH secretion. The mechanism by which the duplications lead to marked pituitary misexpression of *GPR101* alone was previously unclear. Using Hi-C and 4C-seq, we characterized the normal chromatin structure at the X-LAG locus. We showed that *GPR101* is located within a topologically associating domain (TAD) delineated by a tissue-invariant border that separates it from centromeric genes and regulatory sequences. Next, using 4C-seq with *GPR101*, *RBMX*, and *VGLL1* viewpoints, we showed that the duplications in multiple X-LAG-affected individuals led to ectopic interactions that crossed the invariant TAD border, indicating the existence of a similar and consistent mechanism of neo-TAD formation in X-LAG. We then identified several pituitary active *cis*-regulatory elements (CREs) within the neo-TAD and demonstrated *in vitro* that one of them significantly enhanced reporter gene expression. At the same time, we showed that the *GPR101* promoter permits the incorporation of new regulatory information. Our results indicate that X-LAG is a TADopathy of the endocrine system in which Xq26.3 duplications disrupt the local chromatin architecture forming a neo-TAD. Rewiring *GPR101*-enhancer interaction within the new regulatory unit is likely to cause the high levels of aberrant expression of *GPR101* in pituitary tumors caused by X-LAG.

Introduction

The correct expression of genes in time and space is mediated predominantly by noncoding *cis*-regulatory elements (CREs), known as promoters and enhancers, and by *trans*-acting factors, such as transcription factors (TFs). Promoters are located near the transcription start site (TSS) of a gene, while enhancers can be located >1 Mb away.[1] Enhancers increase gene expression by physically interacting with their target promoters via looping of DNA, an interaction that is mediated by tissue-specific TFs. The specificity of enhancer-promoter interactions is achieved, in part, by compartmentalization of the genome into discrete regulatory units termed topologically associating domains (TADs).[2–4] TADs are typically megabase-sized chromatin domains with high levels of internal interaction that are separated from each other by regions of low interaction (TAD borders). TAD borders are enriched for the DNA-binding factor CTCF that acts together with other architectural proteins to establish chromatin interactions and TADs in the human genome.[5–7] Consistent with this proposed role in genome organization, TAD border positions are generally stable across cell types and species.[2,8,9] TADs function as a dynamic scaffold for enhancer-promoter DNA loops, guiding them spatially and facilitating interactions. Disruption of normal TAD organization by genomic rearrangements such as deletions or duplications can expose promoters to external ectopic enhancers (a mechanism called enhancer adoption or hijacking) and can lead to the formation of a new TAD (neo-TAD).[10,11,12] Indeed,

[1]Centro Andaluz de Biología del Desarrollo (CABD), Consejo Superior de Investigaciones Científicas/Universidad Pablo de Olavide, 41013 Seville, Spain; [2]Department of Endocrinology, Centre Hospitalier Universitaire de Liège, University of Liège, Domaine Universitaire du Sart-Tilman, 4000 Liège, Belgium; [3]Department of Human Genetics, Centre Hospitalier Universitaire de Liège, University of Liège, Domaine Universitaire du Sart-Tilman, 4000 Liège, Belgium; [4]NET Service and Endocrine Oncology Bioinformatics Lab, Sheba Medical Center and Sackler Faculty of Medicine, Tel Aviv University, 5265601 Ramat Gan, Israel; [5]Endocrinology - Department of Clinical and Molecular Medicine, Sant'Andrea Hospital - Sapienza University of Rome, 00189 Rome, Italy; [6]Section on Endocrinology and Genetics, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA; [7]Laboratory of Cellular and Molecular Endocrinology, Humanitas Research Hospital – IRCCS, 20089 Rozzano (Mi), Italy; [8]Laboratory of Molecular Pharmacology, GIGA-Molecular Biology of Diseases, University of Liège, 4000 Liège, Belgium; [9]Department of Biomedical Sciences, Humanitas University, Via Rita Levi Montalcini 4, 20072 Pieve Emanuele, Milan, Italy; [10]Human Genetics & Precision Medicine, IMBB, Foundation for Research & Technology Hellas, 70013 Heraklion, Crete, Greece; [11]Research Institute, ELPEN, Pikermi, 19009 Athens, Greece; [12]Laboratory of Cellular and Molecular Endocrinology and Laboratory of Pharmacology and Brain Pathology, Humanitas Research Hospital – IRCCS, 20089 Rozzano (Mi), Italy
[13]These authors contributed equally
[14]Senior author
[15]Deceased
*Correspondence: giampaolo.trivellin@humanitasresearch.it
https://doi.org/10.1016/j.ajhg.2022.02.002.

experimental introduction of TAD border elements outside of their normal genomic context can induce chromatin reorganization and TAD disruption.[13] This rewiring of enhancer-promoter interactions has been implicated as a novel pathogenic mechanism, although few disease phenotypes ("TADopathies") have been described to date.[10,14,15]

X-linked acrogigantism (X-LAG [MIM: 300942]) is a rare disease that leads to large, treatment-resistant growth hormone (GH)-secreting pituitary tumors in neonates and toddlers.[16–18] Persons with X-LAG have tandem, non-recurrent, constitutive (in females) or somatic mosaic (in males) microduplications (average size approximately 600 kb) on chromosome Xq26.3 that invariably include *GPR101* (MIM: 300393) encoding an orphan G protein-coupled receptor (GPCR).[14,17–19] Marked upregulation of *GPR101* at the RNA and protein level is a hallmark of pituitary tumors in X-LAG-affected individuals.[16] Experimental studies have highlighted an important role for GPR101 in promoting pituitary GH and prolactin secretion due to potent constitutive activity via G_s and $G_{q/11}$.[19] Other genes commonly duplicated alongside *GPR101* at the X-LAG locus are *VGLL1* (MIM: 300583), *CD40LG* (MIM: 300386), *ARHGEF6* (MIM: 300267), and *RBMX* (MIM: 300199), but their expression is unaltered in pituitary tumors in X-LAG.[16,17]

In this study, we aimed to unravel the molecular mechanism underlying *GPR101* misexpression in X-LAG tumors. We provide strong evidence that the duplication-induced rearrangement of local chromatin creates a neo-TAD, thereby permitting the *GPR101* promoter to interact ectopically with nearby CREs.

Material and methods

Subjects

Studies outlined below were performed using genetic samples from six individuals with X-LAG and two unaffected mothers of X-LAG-affected subjects (Table S1). Subjects were recruited under the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health protocol 97-CH-0076 (ClinicalTrials.gov: NCT00001595) and under the University of Liège protocol B707201420418. The Institutional Review Boards of both Centers approved this study, and informed assent/consent was obtained from all the subjects and their parents.

Circular chromosome conformation capture (4C-seq)

3C and 4C library preparation was performed as previously described with modifications described below and elsewhere.[20] 4C-seq primer sequences, viewpoint fragment coordinates, and corresponding digestion strategies for primary and secondary restriction enzymes are listed in Table S2. Experiments from X-LAG and control samples were performed as singletons.

Cell fixation and nuclei extraction

Approximately 1×10^6 to 2.5×10^6 cells from cultured peripheral blood mononuclear cells (PBMCs from all control subjects and X-LAG-affected individuals S9 and S13) or nucleated cell isolates from peripheral blood samples (X-LAG-affected individuals S2, S6, S7, and S17) were used as input material for 3C library preparation. Trypsinized or disaggregated cells were filtered through a 40 μm cell strainer (Corning, #352340) and pelleted by centrifugation at 500 × *g*. Cells were fixed in 5 mL of 2% formaldehyde in 10% FCS/PBS and incubated for 10 min at room temperature to cross-link chromatin. The reaction was quenched with glycine at a final concentration of 0.125 M and incubation at room temperature for 5 min. Cells were pelleted by centrifugation for 5 min at 500 × *g* at 4°C and washed on ice twice with 1× PBS. Cell pellets were either snap frozen in liquid nitrogen or processed for nuclei extraction. For nuclei extraction, fixed cells were resuspended in 2 mL freshly prepared lysis buffer (50 mM Tris [pH 7.5]; 150 mM NaCl; 5 mM EDTA; 0.5% NP-40; 1.15% Triton X-100; 1x Roche Complete protease inhibitors) and incubated for 10 min on ice. Nuclei were pelleted by centrifugation for 5 min, 800 × *g* at 4°C and washed with 1× PBS. Pelleted nuclei were either snap-frozen in liquid nitrogen or further processed.

Chromatin digestion and proximity ligation for 3C library preparation

Nuclei pellets were resuspended in 100 μL of 0.5% SDS and incubated for 10 min at 62°C, without shaking. 292 μL water and 50 μL 10% Triton X-100 were added to each sample, mixed, and incubated for 15 min at 37°C to quench remaining SDS. 50 μL of 10× restriction enzyme buffer and a total of 400 units of primary restriction enzyme were added to each sample, mixed, and incubated overnight at 37°C with 900 rpm shaking. The restriction enzyme was heat inactivated according to the manufacturer's instructions. Nuclei were pelleted at 800 × *g* for 10 min at 4°C and resuspended in 500 μL water. To ligate restriction fragment ends, 500 μL of 2× ligation mix (100 μL of 10× ligation buffer [NEB], 100 μL of 10% Triton X-100, 10 μL of 10 mg/mL BSA, 6.5 μL of T4 DNA ligase [NEB, M0202L], 283.5 μL water) were added to each sample and incubated overnight at 16°C and 800 rpm shaking.

Cross-link reversal and DNA purification

Nuclei were pelleted by centrifugation for 10 min, 800 × *g* at 4°C and sample volume was reduced to a total of 200 μL. 230 μL of 10 mM Tris HCL (pH 7.5), 20 μL of Proteinase K (10 mg/mL), and 50 μL of 10% SDS were added, mixed by pipetting, and incubated for 30 min at 55°C. Subsequently, 40 μL of 4 M NaCl were added and samples were incubated overnight at 65°C with 700 rpm shaking. Next, 5 μL of RNase A (10 mg/mL) were added, followed by incubation at 37°C for 30 min at 700 rpm. 20 μL Proteinase K (10 mg/mL) were added to the sample and incubated at 55°C for 1–2 h at 700 rpm. DNA was purified by phenol-chloroform extraction. Following DNA precipitation, the dried DNA pellet was reconstituted in 100 μL 10 mM Tris-HCl (pH 7.5). DNA concentration from resulting 3C libraries was measured with Qubit.

Second restriction digestion and intra-molecular ligation for 4C-seq library preparation

3C libraries were digested with secondary restriction enzyme and incubated overnight at 37°C in a 500 μL reaction volume, containing 1× restriction enzyme buffer and 5 units of restriction enzyme per 1 μg DNA. The restriction enzyme was heat-inactivated following the manufacturer's instructions. Samples were transferred to 50 mL falcon tubes to increase ligation reaction volume with water and to promote intra-molecular ligation events. Samples were ligated overnight at 16°C in a 7 mL ligation reaction, containing 1× ligation buffer, 100 units of T4-DNA ligase (Thermo Fisher Scientific, #EL0011). Samples were further diluted with water to 14 mL. For DNA precipitation, samples were mixed with 1.4 mL 3 M sodium acetate (pH 5.2), 700 μg glycogen, and 35 mL 100% ethanol. Samples were placed at −80°C until solution

became viscous. DNA was pelleted by centrifugation at full speed for 1 h at 4°C. The pellet was washed with 70% ethanol and DNA resuspended in 200 μL of 10 mM Tris-HCl (pH 7.5). Samples were subsequently purified using AMPure XP beads (Agencourt, A63881) as follows: 1.8× volume of AMPure beads were added to the sample, mixed by pipetting, and incubated for 10 min at room temperature. Beads were separated on a magnet, and clear supernatant was discarded. Beads were washed twice with 70% EtOH, air-dried for 5 min and DNA was eluted in 100 μL 10 mM Tris-HCl (pH 7.5). DNA concentration of 4C libraries was measured with Qubit.

Inverse 4C PCR and sequencing of 4C-seq libraries

4C libraries were used as templates for inverse PCR using primers designed for the viewpoint digestion fragment (Table S2). Primers were designed as previously described.[21] The 5′ ends of primers contained the following partial Illumina TrueSeq adapters, read primer 1: 5′-CTACACGACGCTCTTCCGATCT-3′; primer 2: 5′-CAGACGTGTGCTCTTCCGATCT-3′. Primer pairs were tested for optimal PCR amplification using a template dilution series with 0.5, 1, 2, and 4 ng/μL final concentration in PCR. Optimal template concentration determined by gel electrophoresis corresponded to the highest template concentration producing reproducible banding pattern and amplification in a linear range. PCR conditions in 50 μL (Expand Long Template PCR System, Roche, #11759060001): 1× PCR buffer, 0.2 mM dNTPs, 1 μM primers, 0.075 U/μL DNA polymerase, template as determined. PCR cycler setup: (1) 94°C for 2 min, (2) 94°C for 15 s, (3) 55°C for 1 min, (4) 68°C for 3 min, (5) repeat step (2) 29 times, (6) 68°C for 7 min, (7) 4°C. Parallel PCR reactions were performed to amplify from a total of 1 μg template per 4C library and viewpoint. PCR reactions were pooled and purified using 1.8× volume of AMPure XP beads (Agencourt, A63881), see procedure above.

Finally, 4C-seq libraries were indexed using TruSeq index primers (e.g., NEBNext Index Primer for Illumina [NEB #E7335S] and NEBNext High-Fidelity 2X PCR Master Mix [NEB]). 40 μL PCR reaction, containing 1× NEBNext High-Fidelity PCR Master Mix, 0.75 μM NEBNext Universal Primer, and NEBNext Index Primer, 50 ng of inverse PCR product. PCR cycler setup: (1) 98°C for 60 s; (2) 98°C for 10 s; (3) 65°C for 30 s; (4) 72°C for 30 s; (5) repeat step (2) 7 times; (6) 72°C for 5 min. For each sample, 4 independent PCR reactions were performed then pooled for AMPure beads purification, using 1.0× volume of bead solution. Final 4C-seq libraries were resuspended in 50 μL water, multiplexed, and sequenced using HiSeq technology to produce 50 bp single-end reads and approximately 10 million raw sequencing reads for each viewpoint and sample.

4C-seq data analysis and data visualization

4C-seq data were analyzed as previously described.[22] Briefly, sequencing data were demultiplexed and mapped to the human reference genome (GRCh37/hg19). Reads located in restriction fragments flanked by restriction recognition sites of the same enzyme, in fragments smaller than 40 bp or within a window of 10 kb around the viewpoint were filtered out. Mapped reads were then converted to read counts per restriction fragment of the first restriction enzyme. For visualization, all samples were normalized for million mapped reads (reads per million, RPM) for a given genomic region (chrX:135,000,000–137,000,000, hg19) and smoothened using a 10-fragment running window algorithm. To compare interaction profiles of X-LAG subjects to control, subtraction of normalized reads was applied. To reduce variability among control samples, an average profile from three independent control samples was used for subtraction.

Chromatin contact enrichment of candidate enhancers with GPR101 promoter

Normalized read counts for restriction fragments overlapping with candidate enhancer regions (eHTATSF1, chrX:135,577,559–135,581,959; eVGLL1-intronic, chrX:135,625,877–135,641,072; eVGLL1-distal, chrX:135,656,769–135,660,247; ARHGEF6-intronic, chrX:135,846,959–135,851,769; eRBMX, chrX:135,959,959–135,963,959; eAK055694, chrX:135,990,759–135,994,160) were extracted from the *GPR101* viewpoint for X-LAG and control samples. Total read count in each candidate region was averaged and normalized by read-covered fragments to correct PCR bias for spurious and high read counts in single fragments. Fold enrichment of chromatin contacts relative to controls (average from three independent control samples) was calculated and plotted on a logarithmic scale.

Hi-C data visualization and TAD coordinates

GM12878 Hi-C data were visualized in the UCSC genome browser (hg19) at 5 kb resolution in Figure 1. The predicted TAD positions for different cell types (Figures 1 and S1) were obtained from the 3D genome browser[23] (Table S3). TAD positions were predicted from Hi-C data with the hidden Markov model (HMM) pipeline from Dixon et al.[2,23,24] TAD coordinates for all cell types were available in the hg19 genome version, except data for GM12878, which we retrieved in the hg38 genome version and subsequently converted to hg19, using the UCSC liftOver tool.[7,24–28]

CRE annotation

GPR101 promoter annotation was carried out using the Gene2Promoter tool of the Genomatix suite (Genomatix) and the MPromDb software.[29] Two promoter sequences were identified with these methods: a proximal 1,101-bp-long region and a distal 1,097-bp-long region (Figure S5). These sequences were then intersected with specific tracks (CpG Islands, H3K27ac, and H3K4me3 marks from ENCODE) retrieved from the UCSC genome browser and the TSS annotations from the FANTOM5 consortium.

The genomic locations of the regions containing candidate enhancers identified by interrogating the H3K27ac ChIP-seq tracks retrieved from NCBI Gene Expression Omnibus (GEO) GSM1119175 and GSM1119152[30] and the ATAC-seq tracks GEO: GSM3579919[31] were intersected with the list of regulatory elements retrieved from the genome-wide GeneHancer database.[32] This approach allowed the identification of four known putative enhancers sequences: GH0XJ136495, a 4,401-bp-long sequence overlapping the *HTASF1* promoter; GH0XJ136764, a 4,811-bp-long sequence located within an intron of *ARHGEF6*; GH0XJ136878, a 4,001-bp-long sequence overlapping the *RBMX* promoter; GH0XJ136908, a 3,402-bp-long sequence located within the *AK055694* (*ENSG00000234062*) pseudogene. Two additional candidate enhancers were identified within and downstream of *VGLL1* and labeled as eVGLL1-intronic (1,357-bp-long) and eVGLL1-distal (4,101-bp-long), respectively.

Cloning of *GPR101* promoter sequences and candidate enhancers into reporter vectors

The two *in silico*-identified promoter sequences for *GPR101* were PCR amplified from the genomic DNA of a healthy control donor and cloned into the pGEM-T-Easy vector (A1360, Promega) using the SacI restriction site. Cloning was performed using the In-Fusion Cloning System (Takara Bio Europe) and the primers listed in Table S2. The primers were designed using the online In-Fusion Cloning Primer Design Tool (Takara Bio Europe). Both SacI

Please cite this article in press as: Franke et al., Duplications disrupt chromatin architecture and rewire *GPR101*-enhancer communication in X-linked acrogigantism, The American Journal of Human Genetics (2022), https://doi.org/10.1016/j.ajhg.2022.02.002

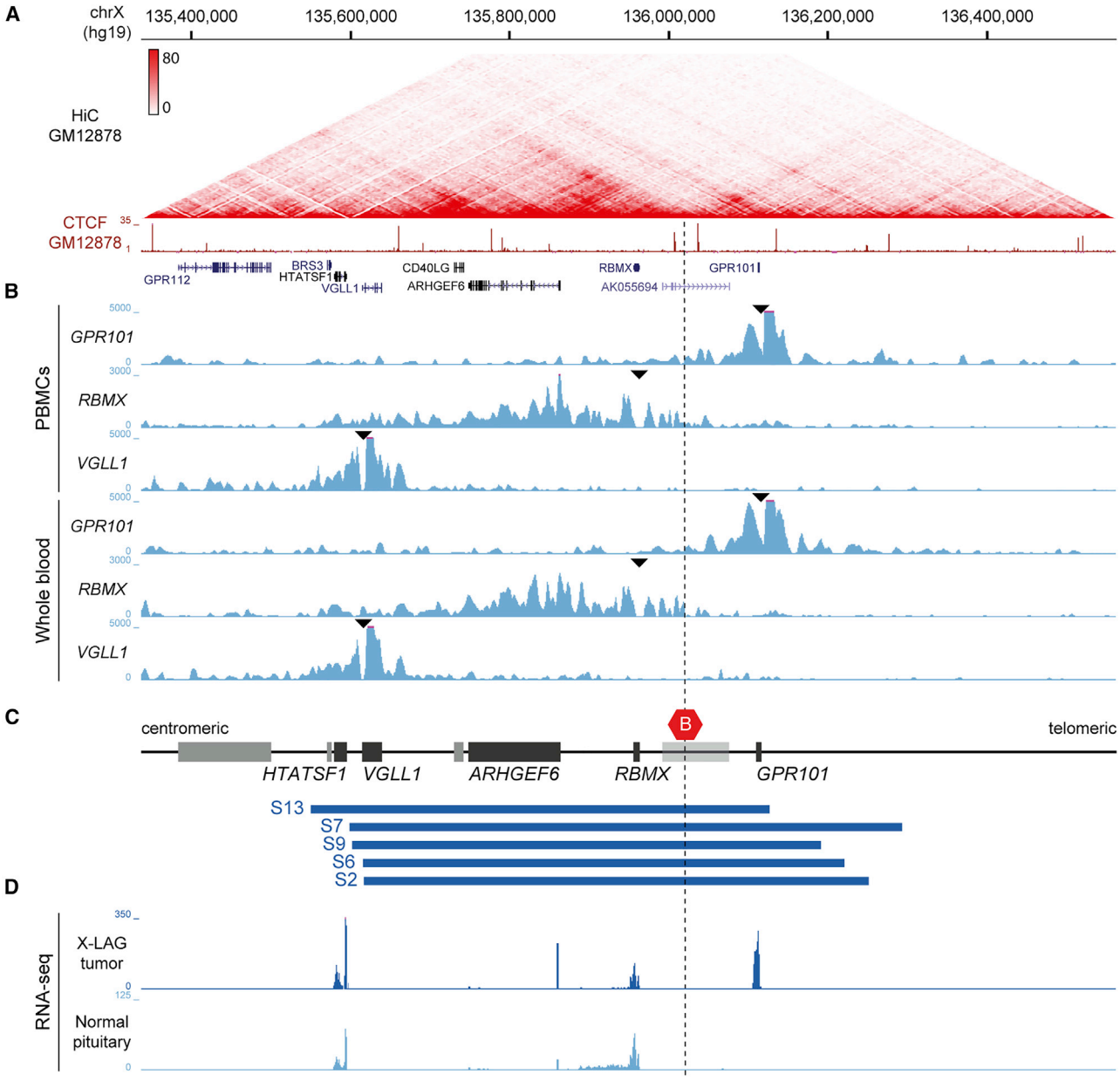

Figure 1. Chromatin organization at the X-LAG locus

(A) Hi-C data from GM12878 lymphoblastoid cells, visualized in the UCSC genome browser at 5 kb resolution (hg19, chrX:135,336,766–136,561,684), showing TAD organization at the X-LAG locus. The ChIP-seq signal for the transcriptional regulator CTCF from GM12878 lymphoblastoid cells and gene annotation are shown underneath. Frequent chromatin interactions within TADs and prominent loop interactions are delimited by CTCF binding sites.

(B) 4C-seq interaction profiles of control samples with viewpoints (black triangles) in the promoters of *GPR101*, *RBMX*, and *VGLL1* in samples of peripheral blood mononuclear cells (PBMCs, upper three tracks) and nucleated blood cells from peripheral blood samples (lower three tracks). Prominent chromatin interactions from each promoter are confined to the respective TADs, recapitulating the TAD organization observed by Hi-C.

(C) Schematic representation of the X-LAG locus and genomic positions of five different tandem duplications from subjects with X-LAG (blue boxes). All duplications involve the invariant TAD border (red hexagon) that under normal conditions separates *GPR101* and genomic sequences located centromerically, including *RBMX* and *ARHGEF6*. For the sake of visualization, genes that were studied in the current experiments are indicated as black boxes while genes not of relevance to the current work are shown in gray boxes. Note that the gene body of the pseudogene *AK055694* (light gray box) overlaps with the TAD border but its putative promoter is located centromeric to the TAD border.

(D) Mean RNA expression levels in four X-LAG tumors and three normal pituitaries show consistent upregulation of *GPR101* in individuals with X-LAG duplications.

All panels are aligned to have the same start and end genomic coordinates.

fragments were subcloned into a promoter-less luciferase reporter vector (pLightSwitch_Prom, S790005, Switchgear Genomics) using the SacI site located upstream of the start codon of the *RenSP* (synthetic *Renilla* luciferase that includes mODC PEST) reporter gene.

The eARHGEF6-intronic (GeneHancer ID: GH0XJ136764) and eVGLL1-intronic fragments were PCR amplified from the genomic DNA of a healthy control donor and cloned into an enhancer reporter vector (pLightSwitch_LR, S990005, Switchgear Genomics) using, respectively, the HindIII or the NheI (5′) and XhoI (3′) sites located upstream of the weak, constitutive herpes simplex virus thymidine kinase (*TK*) gene promoter. Cloning was performed using the In-Fusion Cloning System (Takara Bio Europe) and the primers are listed in Table S2. The eHTATSF1 (GeneHancer ID: GH0XJ136495), eRBMX (GeneHancer ID: GH0XJ136878), eAK055694 (GeneHancer ID: GHOXJ136908), and eVGLL1-distal fragments were cloned within the same enhancer reporter vector using either the KpnI site (eHTATSF1, eRBMX, eAK055694) or the NheI (5′) and XhoI (3′) sites (eVGLL1-distal). These sequences were synthesized and cloned by GeneWiz Inc. All cloned fragments were verified by Sanger sequencing. The preparations of all plasmid DNAs used in the experiments were checked for integrity by restriction digestion with single or double cutters followed by a run on a 1% agarose gel.

Cell culture

The rat pituitary tumor GH3 cell line and the Human Embryonic Kidney (HEK)-293AD cell line were purchased from ATCC (CCL-82.1 and CRL-1573, respectively). Cells were tested for mycoplasma contamination and tested negative. Cells were maintained in Dulbecco's modified Eagle's medium (DMEM, low glucose, pyruvate, no glutamine; GIBCO) supplemented with 10% fetal bovine serum (Gemini Bio Products), 10 mM HEPES, and 1% antibiotic-antimycotic (GIBCO) in a humidified atmosphere at 37°C with 5% CO_2.

CRE luciferase reporter assays

GH3 and HEK293AD cells were seeded in 96-well plates (white clear bottom, poly-d-lysine coated, 354651, Corning Biocoat) at a density of 3.2 and 2.5 × 10^4 cells per well, respectively. After 24 h, cells were transfected with TurboFect (R0531, Thermo Fisher Scientific) according to the manufacturer's protocol, using 100 ng of each reporter vector. A reporter construct containing the strong, constitutive promoter of the human housekeeping gene beta actin (S717678, Switchgear Genomics) was used in each experiment and served as positive transfection control. At 24 h after transfection, cells were lysed and *Renilla* luciferase activity measured using the Renilla Luciferase Assay System (E2810, Promega) following the manufacturer's protocol. All experiments were repeated at least three times with three technical replicates for each construct.

RNA-seq analysis

Each sample was prepared as follows: total RNA was extracted from surgical samples of pituitary tumors in four individuals with X-LAG (three females, one male) and also from three samples of normal adult pituitaries (two males, one female) obtained at autopsy (Cureline Inc.). Total RNA-seq libraries were prepared, libraries were sequenced on an Illumina HiSeq with a 50 bp single- or paired-end protocol and demultiplexed (Zymo Research). Quality of raw data files (fastq) were evaluated using the fastqc (see web resources) and trimmed with Trimmomatic.[33] Then, only high-quality reads were aligned to the human reference genome hg19 using Star[34] in a two-pass approach. A first STAR pass allowed us to obtain a database of splice junctions, from which we filtered out any non-canonical junctions. A new genome was generated with the splice junctions and a second STAR pass was run. Coverage per sample was computed in 1 kb bins and a per sample upper quartile normalization was performed. The normalized average coverage was then computed for all X-LAG samples in the one hand and for the normal samples on the other. Raw count of reads per gene per sample was obtained using the feature Counts.

As recommended, analyses of differential expression to compare X-LAG and normal pituitary samples gene expression were performed with different analytical packages.[35,36] First, we performed an analysis with edgeR and limma as follows: rows with all zero counts and rows with very low counts in at least two samples were removed. Library sizes were then reset in edgeR and a TMM normalization was applied before transforming the counts into log CPM (counts per million) values. A simple linear model matrix with the two conditions as contrast was fit to the data and empirical Bayes statistics were used to identify differentially expressed genes with a robustification against outlier dispersion. This analysis identified 18 differentially expressed genes based on an adjusted p value < 0.05 threshold. The complete list of differentially expressed genes (genes for which a fold change is observed between X-LAG and control samples), is presented in Table S4.

A second differential expression analysis was performed by using HTSeq-count for expression quantification;[37] the derived counts files were then analyzed using DESeq2.[38] This analysis identified 88 differentially expressed genes based on an adjusted p value < 0.05 threshold. The complete list of differentially expressed genes from this second analysis is also presented in Table S4. A list of common differentially expressed genes was then obtained by intersecting the different analyses' outputs, using as a cutoff the first 500 genes of each list ranked by p value. This list, comprising 55 genes (Table 1), was subjected to pathway enrichment analysis using Metascape.[39]

GTEx

The data used for the analyses described in this manuscript were obtained from the GTEx Portal on 09/30/2021.

Statistical analysis

The graphs presented in Figure 6 were plotted as individual biological replicates with mean ± standard deviation (SD). Data distributions were assessed for approximate normality and differences between experimental groups were analyzed by 1-way ANOVA with Dunnett's post hoc test or a corresponding non-parametric test, as appropriate. Data were analyzed using GraphPad Prism (GraphPad). p values < 0.05 were considered statistically significant.

Results

Chromatin organization and normal TAD structure at the X-LAG locus on chromosome Xq26.3

Structural variations and tandem duplications can interrupt and rearrange the local TAD organization and thereby interfere with gene regulation. To investigate the potential contribution of TAD disruption to the etiology of X-LAG, we first determined the normal structural organization at

Please cite this article in press as: Franke et al., Duplications disrupt chromatin architecture and rewire *GPR101*-enhancer communication in X-linked acrogigantism, The American Journal of Human Genetics (2022), https://doi.org/10.1016/j.ajhg.2022.02.002

Table 1. Transcripts of interest differentially expressed in X-LAG versus normal pituitary

Gene symbol	Description	log2 Fold change	p value	p adjusted
GPR101	G protein-coupled receptor 101	12.63436884	1.29E−07	0.000190613
CBLN1	cerebellin 1 precursor	5.512119335	0.000563223	0.069604995
THBS2	thrombospondin 2	5.308143562	2.21E−05	0.006548969
SHC3	SHC adaptor protein 3	3.511547641	1.30E−05	0.006548969
HEPACAM2	HEPACAM family member 2	3.438855732	0.001902735	0.100777001
ECEL1	endothelin converting enzyme like 1	3.435679699	0.000217119	0.04024841
OTOS	otospiralin	3.02397859	0.001895077	0.100777001
ROBO2	roundabout guidance receptor 2	2.710070311	0.001458344	0.090601717
PDE3A	phosphodiesterase 3A	2.661730163	1.97E−05	0.006548969
FKBP10	FKBP prolyl isomerase 10	2.293691778	0.000235806	0.041141147
RAB27B	RAB27B, member RAS oncogene family	2.273319657	0.000154602	0.035272955
ZNF185	zinc finger protein 185 with LIM domain	2.272444372	0.00077183	0.078939629
ENPP2	ectonucleotide pyrophosphatase/ phosphodiesterase 2	2.215979663	0.000302407	0.049829942
RCN1	reticulocalbin 1	1.79981704	0.001001391	0.08735662
THY1	Thy-1 cell surface antigen	1.693370111	0.003678918	0.132060483
FAM163A	family with sequence similarity 163 member A	1.58328952	0.015330845	0.231996364
MEF2C	myocyte enhancer factor 2C	1.295394391	0.005879336	0.169266482
CNTN2	contactin 2	1.257122224	0.008331689	0.191799297
ACSL6	acyl-CoA synthetase long chain family member 6	1.193527525	0.058994577	0.356575463
NRXN3	neurexin 3	1.118119659	0.003173021	0.132060483
ZNF667	zinc finger protein 667	1.085357541	0.000197451	0.039042708
PTPRK	protein tyrosine phosphatase receptor type K	1.060081218	0.003335589	0.132060483
CDH18	cadherin 18	1.035159724	0.002714759	0.125756193
NCALD	neurocalcin delta	1.021986775	0.025977855	0.281559841
AFAP1	actin filament associated protein 1	0.879150352	0.004961622	0.151713106
JMY	junction mediating and regulatory protein, p53 cofactor	−0.697771242	0.00521534	0.156249487
IPO7	importin 7	−0.703833909	0.006282581	0.169612007
DENND4A	DENN domain containing 4A	−0.820624498	0.00709545	0.179872693
ARHGAP26	Rho GTPase activating protein 26	−0.858268895	0.00593517	0.169266482
HTRA1	HtrA serine peptidase 1	−0.933790787	0.042763447	0.326057545
CHD3	chromodomain helicase DNA binding protein 3	−0.972833851	0.011306594	0.224110606
ENAH	ENAH actin regulator	−1.006496792	0.000133954	0.033108953
PTPRS	protein tyrosine phosphatase receptor type S	−1.078053314	0.001090912	0.089879005
SYNE2	spectrin repeat containing nuclear envelope protein 2	−1.252332472	0.001496792	0.090601717
ST18	ST18 C2H2C-type zinc finger transcription factor	−1.257373855	0.006370766	0.169612007
SLC5A3	solute carrier family 5 member 3	−1.27293174	0.043192658	0.326809756

(*Continued on next page*)

Please cite this article in press as: Franke et al., Duplications disrupt chromatin architecture and rewire *GPR101*-enhancer communication in X-linked acrogigantism, The American Journal of Human Genetics (2022), https://doi.org/10.1016/j.ajhg.2022.02.002

Table 1. ***Continued***

Gene symbol	Description	log2 Fold change	p value	p adjusted
IGFBP5	insulin like growth factor binding protein 5	−1.470642696	0.00204994	0.104829669
MCOLN3	mucolipin TRP cation channel 3	−1.538158909	0.008341911	0.191799297
A2M	alpha-2-macroglobulin	−1.546266649	0.00460635	0.148504708
ANXA1	annexin A1	−1.554518488	0.047250273	0.336077478
RAB3B	RAB3B, member RAS oncogene family	−1.559168136	0.000489003	0.067755285
CACNA1H	calcium voltage-gated channel subunit alpha1 H	−1.615245675	0.000484715	0.067755285
STEAP4	STEAP4 metalloreductase	−1.615302685	7.62E−05	0.020556875
FBXO32	F-box protein 32	−1.626663872	0.001474508	0.090601717
DSP	desmoplakin	−1.805180049	8.06E−06	0.005978382
SOD2	superoxide dismutase 2	−1.853240635	0.042101935	0.323536447
TGFBR3	transforming growth factor beta receptor 3	−1.874227585	1.58E−05	0.006548969
OBSCN	obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF	−2.04135558	0.000920676	0.085335122
SORBS1	sorbin and SH3 domain containing 1	−2.100457177	0.010402284	0.220975104
AEBP1	AE binding protein 1	−2.318738851	0.010430383	0.220975104
COL6A6	collagen type VI alpha 6 chain	−2.618856526	0.012027287	0.224358063
GPC4	glypican 4	−3.013712734	0.000545355	0.069604995
NPTX2	neuronal pentraxin 2	−4.197420142	2.05E−05	0.006548969
POMC	proopiomelanocortin	−4.876589807	3.01E−07	0.000297977
LHB	luteinizing hormone subunit beta	−8.869100817	4.10E−09	1.22E−05

55 differentially regulated genes (p value < 0.05) were identified by both analytical methods (edgeR/limma and DESeq2) that were employed to analyze the RNA-seq data. The transcripts are sorted by log2 fold change. EdgeR/limma-derived values are presented in the table.

the X-LAG locus using Hi-C, a chromosome conformation capture technology that allows for the quantification of genome-wide chromatin contacts. Hi-C data at high resolution in the human lymphoblastoid cell line GM12878 showed that the X-LAG locus is highly structured, forming several TADs that spatially separate the locus (Figure 1A). *GPR101* is organized within a TAD extending telomerically and harboring no other nearby genes within the region that is typically duplicated at the X-LAG locus.[7] The neighboring genes that lie centromeric to *GPR101* (*CD40LG*, *ARHGEF6*, *RBMX*), and that are commonly duplicated alongside *GPR101* in X-LAG, are partitioned within distinct TADs. The TAD border separating *GPR101* from the centromeric genes is characterized by two CTCF-binding sites located within a 30 kb region. Importantly, comparison of Hi-C data from 21 human tissues, including pluripotent and terminally differentiated cell types, revealed that the spatial separation of the locus by this TAD border is largely tissue invariant (Figure S1 and Table S3). Furthermore, comparison between human and mouse data revealed that the general chromatin structure and invariant TAD border observed at the X-LAG locus is evolutionarily conserved (Figure S2).

Using 4C-seq, we probed chromatin interactions along the X-LAG locus from *GPR101*, *RBMX*, and *VGLL1* promoters in healthy individuals (Figure 1B). 4C-seq experiments performed in PBMCs and direct cell isolates from peripheral blood samples revealed similar interaction profiles for both sample types. Those profiles resembled the chromatin topology observed from Hi-C data. Chromatin interactions involving the *GPR101* promoter were restricted to the telomeric end of the X-LAG locus and separated from chromatin interactions established by the *RBMX* and the further centromerically located *VGLL1* promoters. This topological separation observed with both Hi-C and 4C-seq is consistent with the known cell type- and tissue-specific expression of genes at the locus. *RBMX* and *ARHGEF6*, that share the same TAD, show overlapping expression profiles in multiple tissues, including strong expression in the pituitary gland (Figure S3). Pituitary expression, although at a lower level, was also observed for *VGLL1* and *CD40LG*. In contrast, *GPR101* exhibits a restricted expression in brain tissues only and is normally not expressed in the adult pituitary gland.[40,41] This suggests that, under normal conditions, genes and their CREs at the X-LAG locus are separated into different

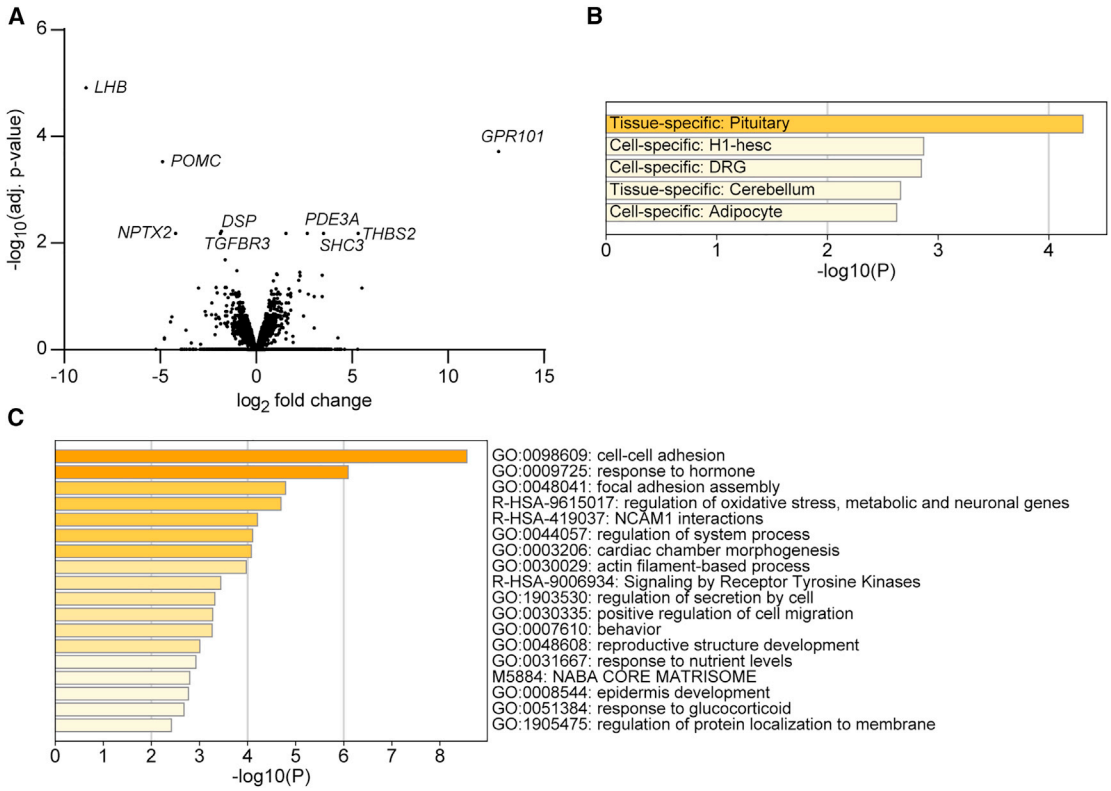

Figure 2. Gene expression analysis in X-LAG tumors versus normal pituitary
Count files from total RNA-seq analyses of four X-LAG tumors and three normal pituitaries were analyzed in parallel using edgeR/limma and DESeq2 methods, which provided highly consistent outputs of differentially expressed genes.
(A) A volcano plot shows the range of significantly up- and downregulated genes, among which *GPR101* was the most markedly dysregulated gene overall. The most significantly downregulated genes were those characteristic for pituitary corticotrope (*POMC*) and gonadotrope (*LHB*) secretion, reflecting a tumoral process in X-LAG that favored somatotrope development and secretion over other pituitary axes. The most dysregulated genes with the lowest adjusted p values are highlighted. EdgeR/limma-derived values are presented in the plot.
(B) Pituitary was identified as the tissue most significantly affected in a pathway analysis using GO categories from PaGenBase.[39,42]
(C) The GO terms that were most significantly altered in the RNA-seq dataset were those for cell-cell adhesion and response to hormone; other GO terms related to pituitary or endocrine function included regulation of secretion by cell, reproductive structure development, and response to glucocorticoid.

TADs and that a strong TAD border separates *GPR101* from centromeric genes and regulatory sequences.

Overlapping the position of X-LAG-associated tandem duplications with the chromatin structure at the locus revealed that all duplications include the observed strong and tissue-invariant TAD border, *GPR101,* and genomic sequences centromeric to this TAD border (Figure S4). In this study we will focus on individuals S2, S6, S7, S9, and S13, as shown in Figure 1C.

RNA-seq analysis in X-LAG tumors and normal pituitary tissue further supported an effect on gene expression that was focused on *GPR101*. As seen in Figure 1D, *GPR101* was markedly upregulated while other duplicated genes at the X-LAG locus were unaltered or remained expressed at low levels (mean raw gene counts at the locus are listed in Table S5).

The differentially expressed genes listed in Table 1 show that *GPR101* was by far the most significantly dysregulated gene overall in X-LAG tumors versus normal pituitary (>12 $\log_2$-fold increase). Interestingly, specific genes for hormones secreted by other pituitary cell types like corticotropes (*POMC*) and gonadotropes (*LHB*) were highly downregulated in X-LAG pituitary (Table 1, Figure 2A). This contributed to a strong tissue-specific pituitary signature in terms of significantly disordered pathways (Figure 2B), including individual GO terms such as response to hormone, regulation of secretion by cell, and reproductive structure development (Figure 2C). These RNA-seq data corroborate and expand on the role of GPR101 in the regulation of GH and prolactin secretion that we previously reported in transgenic mice.[19] The involvement of other GO pathway terms like cell-cell adhesion and focal adhesion assembly may reflect the molecules and signals involved in the tumorigenic processes in pituitary adenomas from the X-LAG subjects.

X-LAG-related chromosome Xq26.3 duplications induce neo-TAD formation and ectopic chromatin interactions

Using 4C-seq, we investigated the effects of chromosome Xq26.3 duplications on TAD structure in PBMCs and

Please cite this article in press as: Franke et al., Duplications disrupt chromatin architecture and rewire *GPR101*-enhancer communication in X-linked acrogigantism, The American Journal of Human Genetics (2022), https://doi.org/10.1016/j.ajhg.2022.02.002

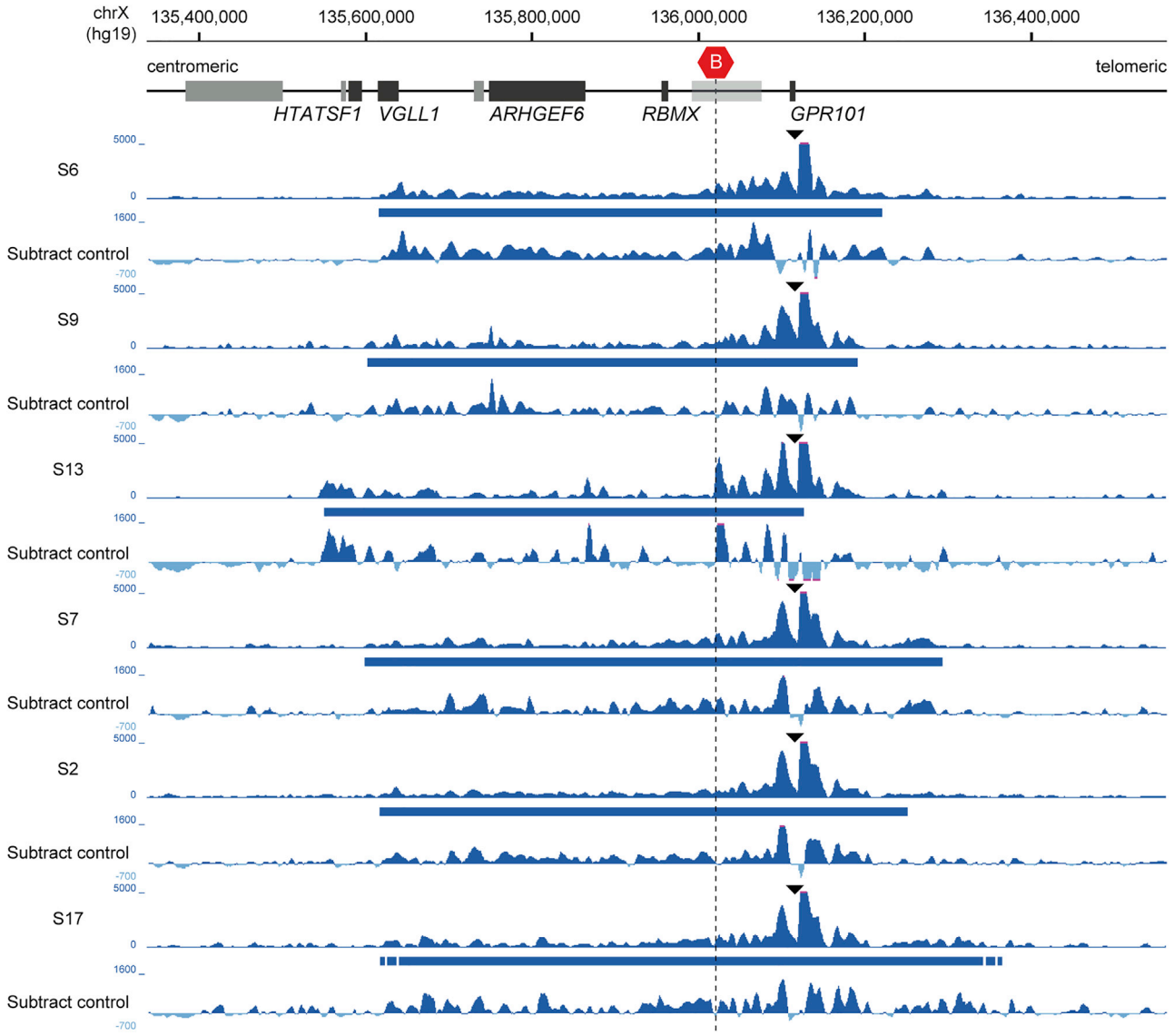

Figure 3. Ectopic chromatin interactions of *GPR101* extend centromerically and cross a TAD border
4C-seq profiles from the *GPR101* viewpoint in six different X-LAG-affected individuals. Size and position of each duplication are indicated below each 4C-seq profile (blue bar) and corresponding subtraction profiles to control samples are presented below. *GPR101* shows consistent ectopic interactions with regions centromeric of the TAD border. Note that the exact breakpoints for the S17 duplication is not determined.

nucleated cell isolates from peripheral blood samples collected from six unrelated X-LAG-affected individuals, as compared with normal control subjects. For these studies, the viewpoint for the 4C-seq analysis was placed on the *GPR101* promoter region. Across the X-LAG-affected individuals, we demonstrated that the chromosome Xq26.3 duplication led to new interactions that crossed the normal TAD boundary surrounding *GPR101*. This, in turn, revealed a landscape of ectopic chromatin interactions of *GPR101* outside of the normal TAD border (Figure 3).

This is illustrated, for example, in subject S6, in whom the duplication creates a unique configuration of genomic sequences around its breakpoint, delineated by the duplicated TAD border (Figure 4A). Using the breakpoint as a unique viewpoint in the subject's genome, we further demonstrated that the resulting 4C-seq profile showed typical enrichment close to the viewpoint (i.e., the breakpoint) but also frequent chromatin interactions across duplicated sequences that cross the usually invariant normal TAD border. Also, genomic regions outside the duplication showed very low levels of interaction, indicating that the duplicated sequences within the neo-TAD are physically separated. Therefore, in subject S6, *GPR101* and *RBMX* viewpoints within the duplication showed ectopic contacts when compared to control samples (Figure 4A). The *GPR101* viewpoint showed additional interactions that extended across the normally invariant TAD border toward the centromeric duplicated sequences, a pattern not observed under normal conditions. The

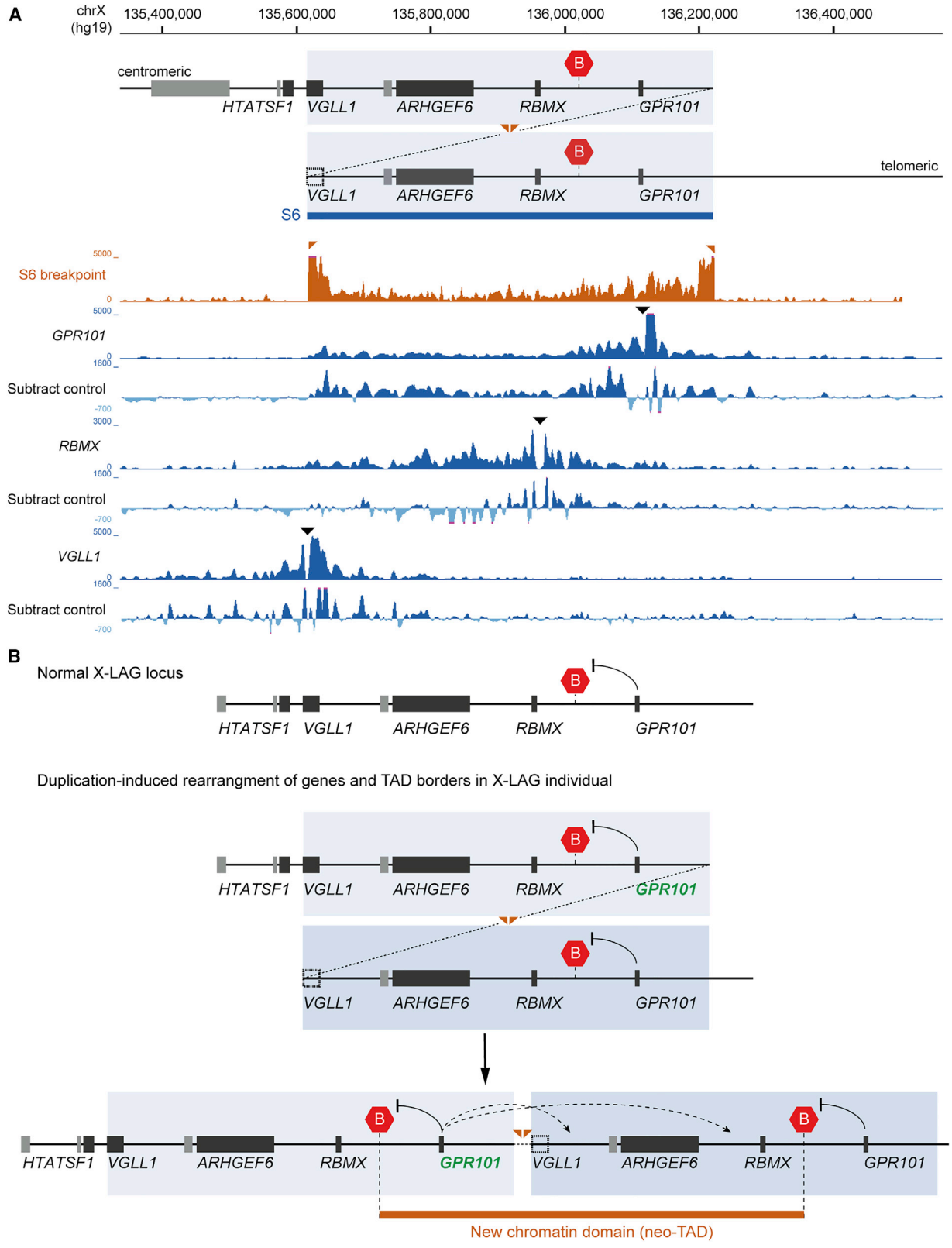

Figure 4. X-LAG duplications create a neo-TAD and ectopic chromatin contacts

(A) 4C-seq experiments in a person harboring a 600 kb tandem duplication (individual S6 reported previously).[16] The schematic depicts the S6 duplicated allele with a duplicated TAD border (red hexagons). Size and position of the duplication is indicated by overlap and blue shaded area. The dotted line, connecting the duplicated sequences, illustrates the duplication breakpoint. Note that the duplication breakpoint disrupts *VGLL1*, excluding the *VGLL1* promoter and viewpoint from the duplication. Below, the 4C-seq interaction profile from the unique viewpoint created by the duplication breakpoint is shown (brown track). The breakpoint is flanked by the duplicated

(legend continued on next page)

RBMX viewpoint also showed ectopic contacts with genomic regions around *GPR101*. In contrast, the *VGLL1* promoter, which was outside of the duplication in this subject, did not show any abnormal chromatin interactions, confirming that sequences within neo-TADs are spatially insulated. The observed aberrant interaction patterns suggest the formation of a neo-TAD, which rearranges genomic sequences that, under normal conditions, do not interact with each other. Overall, the Hi-C, 4C-seq, and RNA-seq data indicate that the inclusion of one copy of *GPR101* within the neo-TAD and the resulting ectopic contacts with potential CREs (Figure 4B) can account for *GPR101* upregulation observed in X-LAG pituitary tumors.

The 4C-seq data showed consistent effects across all X-LAG samples irrespective of the different spans of the Xq26.3 duplication (Figures 3 and 4). Despite the variable duplication breakpoints, *GPR101* showed ectopic interactions that cross the invariant TAD border at the X-LAG locus. These findings strongly suggest that a similar and consistent mechanism of neo-TAD formation exists for all tested duplications in X-LAG.

GPR101 ectopically interacts with candidate pituitary CREs

In X-LAG, the newly established interactions harbor potential CRE sequences such as enhancers that could drive the marked overexpression of *GPR101* that is seen in the X-LAG pituitary tumors. We searched for CREs predicted in the GeneHancer database and for enhancer-associated histone modifications (H3K27ac) from the human pituitary and hypothalamus to identify those CREs with active chromatin signatures (Figure 5A). The genomic region telomeric to the invariant TAD border and containing *GPR101* was largely depleted of predicted CREs and pituitary-specific signals. We identified only two regions showing activity in the pituitary, one in the known promoter of *GPR101*[17,41] and one 21 kb upstream of its start codon. The latter region overlapped with H3K4me3 signals and a CpG island and likely corresponds to a putative far-distal promoter region of *GPR101* (Figure S5). In contrast, the region centromeric to the normal invariant TAD border showed multiple predicted CREs and widespread activity in the hypothalamus and pituitary gland, which is consistent with the transcriptional activity of genes in this region (Figure S3). We found four predicted CREs with strong H3K27ac enrichment in the proximal promoter regions of *HTATSF1* (MIM: 300346), *RBMX*, *AK055694* (a pseudogene telomeric of *RBMX*), and in the intron of *ARHGEF6* (eARHGEF6-intronic). In addition, and although not predicted by GeneHancer, we identified two CRE candidates that showed a pituitary-specific H3K27ac enrichment in *VGLL1* (eVGLL1-intronic) and downstream of *VGLL1* (eVGLL1-distal). Interestingly, these enhancer candidates with pituitary-specific signals contain sequences conserved among mammals, and chromatin accessibility assays in mouse pituitary revealed a functional conservation in somatotrope and lactotrope subpopulations of the anterior lobe of the pituitary (Figures 5A and S6). Collectively, this analysis revealed that the genomic region centromeric to the TAD border contains multiple CREs and shows widespread regulatory activity, including in the pituitary gland. These CRE sequences showed increased interaction frequency with the *GPR101* promoter in cells from all tested X-LAG-affected individuals (Figure 5B).

The rearrangement of large portions of this pituitary active region and *GPR101* within the neo-TAD would favor their spatial proximity and could lead to dysregulation of *GPR101*. To study this further, we overlapped the position of CREs with the smallest regions of overlap (SROs) of known X-LAG-associated duplications. Interestingly, two out of 29 duplications at the locus (individuals S4 and I) are interrupted, with sequence duplications on either side of the invariant TAD border, defining two SROs, one telomeric and one centromeric to the invariant TAD border (Figure S4).[43] The telomeric SRO always contains *GPR101* and its known promoter sequence whereas the centromeric SRO includes a pituitary-active region, containing at least the eVGLL1-intronic enhancer. The induced genomic rearrangements of all known X-LAG duplications reposition one copy of *GPR101* next to active pituitary regions. This scenario supports the model in which spatial proximity of *GPR101* with potential pituitary-active CREs could drive *GPR101* dysregulation in X-LAG-affected subjects.

GPR101 is compatible with pituitary and embryonic developmental programs

GPR101 dysregulation in the pituitary by new CREs has as a prerequisite the compatibility of the *GPR101* promoter to respond functionally to these new ectopic signals. We, therefore, tested the ability of the *GPR101* promoter to activate gene expression in response to pituitary- and embryonic-specific cellular programs. The telomeric SRO overlaps with the known *GPR101* promoter and in this region two promoter sequences (defined here as proximal and distal), each overlapping a CpG island, were identified. Their location is corroborated by our previous RNA-seq

TAD border and the 4C-seq profile shows high interaction frequencies with regions restricted to the duplication, indicating that the neo-TAD is spatially insulated. All reads mapped to a wild-type genome (resulting in split viewpoint for duplication breakpoints). Below this, 4C-seq profiles from the *GPR101*, *RBMX*, *VGLL1* viewpoints and corresponding subtraction profiles to control samples are presented (blue tracks). The viewpoints at *GPR101* and *RBMX* promoters, included in the duplication and neo-TAD, show ectopic chromatin interactions. The *VGLL1* promoter viewpoint, excluded from individual S6's duplication, shows a normal interaction profile.
(B) Model depicting chromatin interactions (arrows) at the X-LAG locus under normal conditions and after duplication-induced rearrangement of genes and the invariant TAD border. Rearranged sequences between the duplicated TAD border creates a new chromatin domain (neo-TAD) that is characterized by ectopic interactions (dotted arrows) and little or no interaction beyond the TAD border.

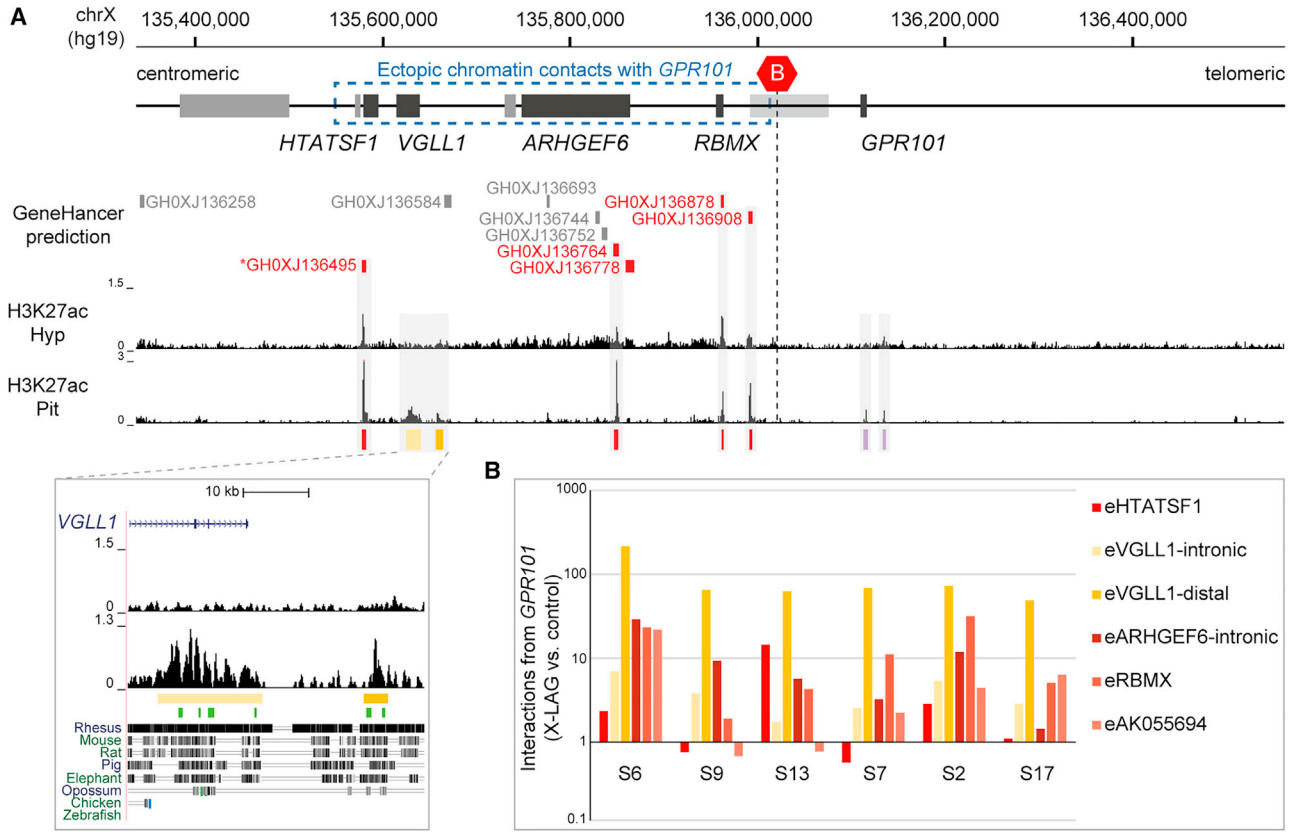

Figure 5. *GPR101* ectopically interacts with candidate pituitary CREs in the neo-TAD

(A) X-LAG locus, depicting the region of ectopic chromatin contacts established by *GPR101* in X-LAG-affected individuals (blue box). The pituitary (Pit) and hypothalamus (Hyp) H3K27ac ChIP-seq tracks were retrieved from NCBI Gene Expression Omnibus (GEO) GSM1119175 and GSM1119152, respectively.[30] Corresponding enriched open chromatin regions identified in mouse somatotropes and lactotropes are indicated by green lines (ATAC-seq tracks retrieved from GEO: GSM3579919).[31] Regions with candidate enhancers that ectopically interact with *GPR101* are indicated with different saturations of yellow (pituitary-specific CREs) and in red (GeneHancer-predicted CREs). The asterisk preceding the candidate GH0XJ136495 CRE (eHTATSF1) indicates an elite gene association. *GPR101* putative promoter regions are indicated in purple.

(B) Enrichment of chromatin contacts from the *GPR101* viewpoint with candidate enhancers (X-LAG-affected individuals versus control). Note that the eHTATSF1 CRE is only included in the duplication of subject S13.

and 5′-RACE data[17,41] and by public CAGE-seq data[44] (Figure S5). We functionally tested both promoter elements by cloning these sequences into a promoter-less luciferase reporter vector. We transfected the constructs into rat pituitary tumoral GH- and prolactin-secreting GH3 cells and human embryonic kidney HEK293AD cells, and measured luciferase activity. Both sequences were found to be functional, with the distal promoter showing higher transcriptional activity and potential to drive *GPR101* expression in both cell types (Figure 6A).

Next, we tested the ability of CREs located in the centromeric region to drive reporter gene expression in GH3 and HEK293AD cells. We selected the regulatory regions that overlapped GeneHancer-predicted CREs and the therein-identified CREs with pituitary-specific H3K27ac signatures (Figure 5B). We cloned each CRE into a reporter vector containing a minimal promoter and then transfected them into both cell lines. Luciferase activity revealed that GeneHancer-predicted candidates (eHTATSF1, eARHGEF6-intronic, eRBMX, eAK055694) reduced reporter gene expression in GH3 cells while pituitary-specific CREs (eVGLL1-intronic and eVGLL1-distal) maintained basal promoter activity (Figure 6B). However, in embryonic HEK293AD cells the eRBMX element significantly enhanced reporter gene expression ($p < 0.0001$ versus basal promoter). All the other CREs either maintained basal promoter activity or repressed it (eVGLL1-intronic, Figure 6B).

Taken together, the results revealed that the *GPR101* distal promoter is capable of driving gene expression in adult rodent pituitary tumor and human embryonic cellular contexts. Therefore, the *GPR101* promoter permits the incorporation of new regulatory information within the neo-TAD. At the same time, CREs located within the centromeric region and that ectopically interact with *GPR101* in X-LAG pituitary cells show context-dependent activity patterns. No transcriptional enhancing function for any tested CRE could be scored in GH3 cells. However, the eRBMX element showed transcriptional enhancing activity in the embryonic cellular context of HEK293AD cells. Interestingly, two CREs repressed basal promoter activity in specific cellular contexts.

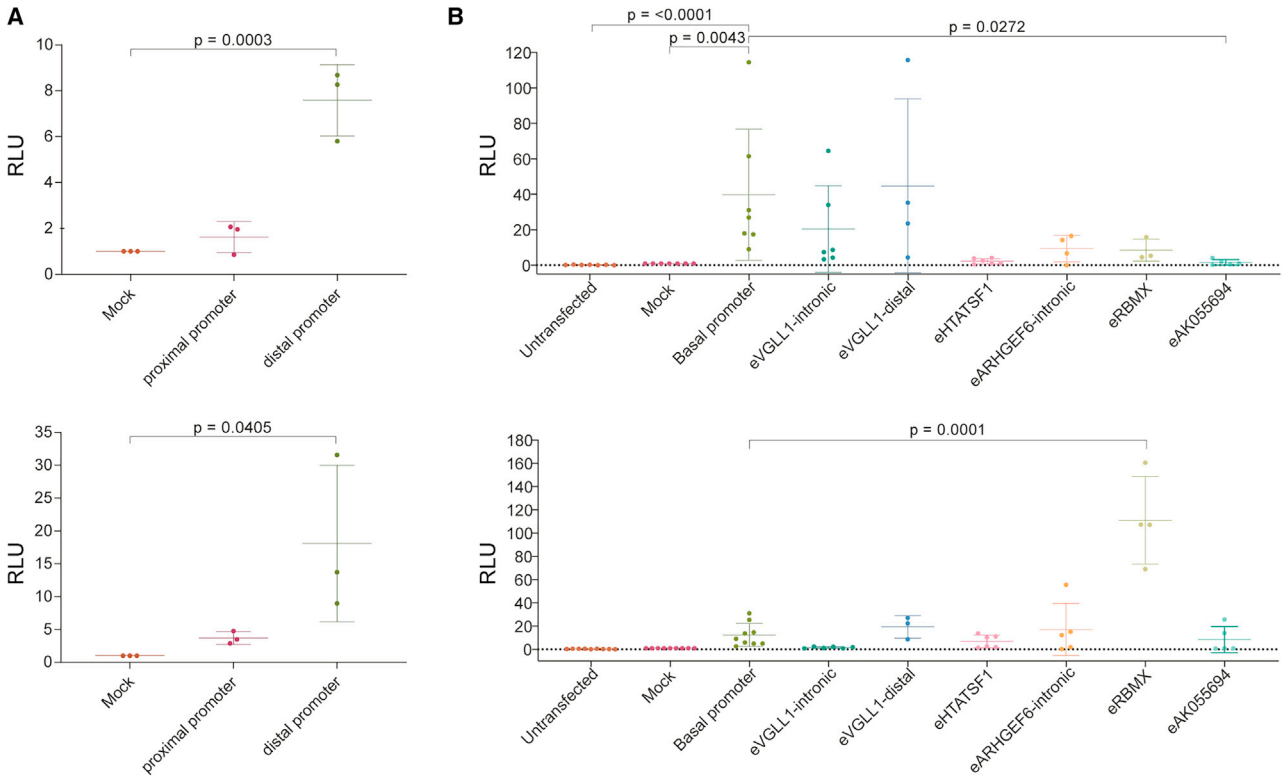

Figure 6. ***In vitro*** **characterization of** ***GPR101*** **promoter and candidate enhancers**
(A) Luciferase reporter assay of proximal and distal promoter elements of *GPR101* in GH3 (top) and HEK293AD (bottom) cells.
(B) Luciferase reporter assay of candidate pituitary active CREs in GH3 (top) and HEK293AD (bottom) cells. The basal promoter construct does not contain any enhancer sequence.
Data in both panels are plotted as individual biological replicates with mean ± standard deviation (SD). All values are relative to Mock (promoter-less luciferase reporter vector) in both panels. Data distributions were assessed for approximate normality and differences between experimental groups were analyzed by 1-way ANOVA with Dunnett's post hoc test or corresponding non-parametric test, as appropriate. RLU, relative luciferase units.

Discussion

Pituitary gigantism is a rare disorder due to GH-secreting pituitary tumors that leads to dramatic and irreversible skeletal overgrowth and multi-organ pathology, including cardiovascular disease.[45] Nearly half of cases of pituitary gigantism are associated with known genetic abnormalities.[46] Among these, X-LAG is the most severe form, with large, highly secretory pituitary tumors developing during neonatal life and early childhood. At a genetic level, X-LAG is associated with chromosome Xq26.3 microduplications that invariably include *GPR101*, leading to extremely elevated *GPR101* expression in the pituitary.[16] GPR101 increases GH and prolactin production and induces overgrowth via constitutive activation of the G protein α subunits G_s, $G_{q/11}$, and $G_{12/13}$,[19] but the mechanism of how a tandem duplication leads to massive overexpression of GPR101 in X-LAG tumors was not explained until now.

Consequences of gene copy-number variations (CNVs) on mRNA expression levels have been studied comprehensively in different species, cell lines, and cells derived from individuals with different types of CNVs. Taken together, these studies concluded that there is an appreciable correlation between mRNA levels and gene copy number.[47] However, for individual genes, mRNA levels often deviate from the expected levels. For example, *GPR101* is upregulated around 1,000-fold in the pituitary tumors of subjects with X-LAG.[16] This observation implies that the phenotypic effect conferred by Xq26.3 duplications in X-LAG is not due simply to increased *GPR101* gene dosage. A potential mechanism that could explain *GPR101* misexpression in X-LAG tumors is the creation of an abnormal fusion gene, whereby the genomic duplication places the *GPR101* protein-coding region under the influence of a strong promoter located close to the breakpoints. However, no fusion events involving *GPR101* were detected in X-LAG tumor samples (n = 4, Table S6). Moreover, no other fusion genes shared among the tumors were detected.

Our results, instead, show that Xq26.3 microduplications alter the chromatin configuration and normal TAD organization at the X-LAG locus. These tandem duplications disrupt a tissue-invariant TAD border that normally separates *GPR101* from genes and regulatory sequences located centromerically. Tandem duplications that cross TAD borders (inter-TAD duplications) have been shown to rearrange the additional copy number of genes and CREs into new TADs, so-called neo-TADs.[10,48] In the

current study, we have shown similar re-organization of the chromatin configuration and neo-TAD formation at the X-LAG locus based on several 4C-seq promoter viewpoints and duplication breakpoint analyses. Furthermore, the neo-TAD formation is consistently seen across samples from multiple X-LAG-affected individuals with different microduplication ranges and breakpoints. The rearranged, duplicated genomic sequences ectopically interact within the neo-TAD and these interactions are spatially restricted by the duplicated TAD border. These results strongly suggest that neo-TAD formation (TADopathy) is the genetic pathophysiology behind *GPR101* misexpression in X-LAG. As such, the endocrine disease X-LAG joins a limited but growing list of TADopathies involving limb malformation, retinal disease, platelet dysfunction, and cancer.[10,12,14,15,48–50]

Within neo-TADs, genes and CREs can interact ectopically, causing aberrant gene expression leading to disease phenotypes. By probing interactions from the *GPR101* promoter in cells from multiple X-LAG-affected individuals with different microduplications, we identified consistent patterns of ectopic interactions. Based on CRE predictions and enhancer-associated chromatin marks, we then identified several potential ectopic CREs that are active in the adult pituitary gland.[41] These findings lead us to suggest that the new regulatory landscape created by neo-TADs in X-LAG may override the physiological regulation of *GPR101* expression. We have previously reported that *GPR101* is linked to the maturation of the somatotrope cell population of the pituitary during embryonic development. *GPR101* expression then switches off during post-natal development, except possibly for the growth spurt phase occurring during adolescence.[41] In contrast, *GPR101* is strongly and consistently upregulated in X-LAG pituitary tumors that have been operated at different ages. Subsequent studies in animal models confirmed the crucial contribution of GPR101 misexpression in the hormonal regulation of body growth. Overexpression of GPR101 in the pituitary of transgenic mice leads to GH and prolactin excess and overgrowth,[19] while its whole-body loss in zebrafish causes reduced body size by perturbing early embryogenesis.[51] These findings indicate that precisely fine-tuning GPR101 expression both temporally (developmental stage) and spatially (specific tissues) is important for proper body growth. In X-LAG, it is conceivable that the exposure of *GPR101* to the new regulatory landscape within the neo-TAD maintains active expression at high levels even after embryonic pituitary development is complete. The manifestation of the phenotype as early as the first months of life[18,52] underscores the high potency of *GPR101* for regulating growth via its permissive role in GH secretion. This may also provide a rationale for the normal localization of *GPR101* alone in a TAD, thereby insulating it from nearby CREs. The stability across cell types and evolutionary conservation of the TAD boundary that separates *GPR101* from centromeric CREs underscores the tight regulation of the normal *GPR101* regulatory unit, as has been reported for other topologically isolated genes with roles in development.[53] Stable boundaries are more intolerant of disruption by structural variants than unique ones,[54] implying that the X-LAG locus is under strong selection. The severity of the clinical effects of gigantism in X-LAG and its extreme rarity (40 individuals described) combine to support a hypothesis that *GPR101* and its surrounding regulatory environment are tightly regulated.

We functionally studied the *GPR101* promoter and pituitary CREs using two *in vitro* models, one a human embryonic cell line (HEK293) to provide species and development stage specificity, and the other an adult rat pituitary tumor cell (GH3) to study tissue specificity. We found that the distal promoter element of *GPR101* is active in an embryonic cellular context and is also compatible with adult rat pituitary tumor cells. The CRE located on the promoter region of *RBMX* (eRBMX) increased basal promoter activity in embryonic cells by about 9-fold. This finding is consistent with the known role of *RBMX* as a regulator of embryogenesis.[55,56]

There was no measurable transcriptional enhancing activity for any of the centromeric CREs in adult rat pituitary tumor cells. Interestingly, we also found that some CREs had an unexpected activity as silencers in the cell models, in particular the eAK055694 CRE in pituitary cells. Similar dual enhancer/silencer activities have been reported at other loci and are active in pituitary cells. Those CREs were validated in an analogous experimental setting to ours.[57] Interestingly, about a quarter of human candidate silencer elements from one cell type can have dual activity and act as active enhancers in different cell types and chromatin contexts.[58] This raises the intriguing possibility that some of our candidate CREs displaying transcriptional silencing in GH3 and HEK293 could actually function as enhancers in the X-LAG-affected subjects' pituitary cells. Moreover, there has been at least one instance reported in which the relocation of a silencer into a different TAD converts it into an activator of its target gene.[59]

Our study has several limitations. First, we employed *in vitro* episomal reporter assays for the functional evaluation of predicted CREs. While episomal vectors are widely used to characterize promoters/enhancers, they inherently suffer from two main disadvantages: (1) their chromatin may have different properties and does not necessarily reflect the endogenous epigenomic chromatin state and, therefore, (2) the intrinsic enhancer activity may not fully recapitulate endogenous target gene expression.[60] Second, multiple neighboring enhancers may exert cumulative activity on a single gene regulatory pathway and be part of a so-called multipartite enhancer[61] or enhancer chains.[62] Our reporter assays evaluated each CRE independently from the others and thus could not measure whether they acted cooperatively or influenced one another. Third, limited biochemical

annotations are publicly available to aid in the identification of candidate CREs that are active in the human pituitary gland. The dataset interrogated in this study included few specimens collected postmortem in adult individuals.[30] This paucity of data could have led to our missing the identification of other CREs that are active in the pituitary only in specific cell sub-types and/or at specific developmental stages. Fourth, the cell types employed for the assays may have not provided the correct cellular context. The unavailability of either embryonic or adult immortalized human somatotrope cell lines is a known limitation in the field of pituitary research.[63] Specific transcription factors or coactivators may be missing or not expressed at the desired levels in the recipient cells we used, thus impairing the activity of the tested CREs.

In conclusion, we show that X-LAG is a TADopathy of the endocrine system in which chromosome Xq26.3 microduplications disrupt the local chromatin architecture forming a neo-TAD that permits ectopic contacts between the promoter of *GPR101* and centromeric CREs. Enhancer adoption/hijacking within the new regulatory unit is the likely cause of the extremely high levels of misexpressed pituitary *GPR101* and the subsequent tumoral GH hypersecretion that epitomizes X-LAG. Our findings raise the possibility that other unexplained forms of hormonal dysfunction in the endocrine and neuroendocrine systems could be explained by similar TADopathy-based enhancer-gene dysregulation mechanisms.

Data and code availability

The 4C-seq and RNA-seq datasets we generated in this manuscript have been deposited in the GEO database under the accession code GSE193114.

Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2022.02.002.

Acknowledgments

The authors thank Jacques Drouin (Laboratoire de Génétique Moléculaire, Institut de Recherches Cliniques de Montréal, Montreal, QC, Canada) for providing information concerning the ATAC-seq tracks interrogated in this study, Marit W. Vermunt and Menmo P. Creyghton (Hubrecht Institute-KNAW & University Medical Center Utrecht, Utrecht, the Netherlands) for guiding us on the use of the H3K27ac pituitary tracks, Julien Hanson (Laboratory of Molecular Pharmacology, GIGA-Molecular Biology of Diseases, University of Liège, Belgium) for discussions on orphan GPCR function and regulation, and Michela Matteoli (Laboratory of Pharmacology and Brain Pathology, Humanitas Research Hospital – IRCCS, Rozzano, Italy) for scientific guidance. The authors also thank Maria Chiara Zatelli (Section of Endocrinology, Department of Medical Sciences, University of Ferrara, Ferrara, Italy), Giovanna Mantovani (Endocrinology and Diabetology Unit, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy), and Catherine Choong (Department of Pediatric Endocrinology and Diabetes, Princess Margaret Hospital for Children, Subiaco, Western Australia, Australia) for providing materials for this study, and thank Lyssikatos Charalampos and Maria de la Luz Sierra (NICHD/NIH) for their help with the collection of biological samples used in this study.

The authors dedicate this paper to the memory of José Luis Gómez-Skarmeta.[64,65]

The work was supported by the following funding sources: Fondazione Telethon, Italy grant no. GGP20130 (to G.T.); Society for Endocrinology equipment grant (to G.T.); Intramural Research Program, *Eunice Kennedy Shriver* National Institute of Child Health & Human Development (NICHD), National Institutes of Health (NIH) Research project Z01-HD008920 (to C.A.S., supporting G.T., F.R.F.); Fonds d'Investissement pour la Recherche Scientifique (FIRS) of the Centre Hospitalier Universitaire de Liège (to A.F.D. and A.B.); the JABBS Foundation, UK (to A.B.); and Novo Nordisk Belgium Educational Grant, Belgium (to A.F.D. and A.B.).

M.F. was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement (#800396) and a Juan de la Cierva-Formación fellowship from the Spanish Ministry of Science and Innovation (FJC2018-038233-I). G.T. was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement (#843843). A.F.D. and D.A. were supported by Action de Recherche Concertée (ARC) Grant 17/21-01 from Liège University. D.A. was supported by grants from Télévie (7461117 F, 7454719 F) and the Léon Fredericq Foundation, Belgium.

Declaration of interests

A.B., A.F.D., F.R.F., C.A.S., and G.T. hold a patent on *GPR101* and its function (US Patent No. 10,350,273, Treatment of Hormonal Disorders of Growth). C.A.S. holds patents on technologies involving *PRKAR1A* and related genes causing adrenal, pituitary, and other tumors. In addition, his laboratory has received research funding support by Pfizer Inc. for investigations on growth-hormone producing pituitary adenomas. C.A.S. also has consulted within the last 12 months with Lundbeck Pharmaceuticals and Sync, LLC, and is currently employed by ELPEN Pharmaceuticals. A.B. and A.F.D. have received research funding from Pfizer Inc. and Novo-Nordisk. The authors declare that they have no conflicts of interest with the contents of this article.

Received: October 14, 2021
Accepted: February 1, 2022
Published: February 23, 2022

Web resources

3D Genome Browser, http://3dgenome.org
ClinicalTrials.gov, https://clinicaltrials.gov
FANTOM5 consortium, https://slidebase.binf.ku.dk/human_promoters/
FastQC, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
GTEx Project, https://www.gtexportal.org/home/
OMIM, https://omim.org/

Primer3, http://bioinfo.ut.ee/primer3-0.4.0/
UCSC Genome Browser, https://genome.ucsc.edu

References

1. Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Hum. Mol. Genet. *12*, 1725–1735.
2. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.
3. Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.
4. Spielmann, M., Lupiáñez, D.G., and Mundlos, S. (2018). Structural variation in the 3D genome. Nat. Rev. Genet. *19*, 453–467.
5. Merkenschlager, M., and Nora, E.P. (2016). CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. Annu. Rev. Genomics Hum. Genet. *17*, 17–43.
6. Nora, E.P., Goloborodko, A., Valton, A.L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930–944.e22.
7. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.
8. Gómez-Marín, C., Tena, J.J., Acemel, R.D., López-Mayorga, M., Naranjo, S., de la Calle-Mustienes, E., Maeso, I., Beccari, L., Aneas, I., Vielmas, E., et al. (2015). Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. Proc. Natl. Acad. Sci. USA *112*, 7542–7547.
9. Vietri Rudan, M., Barrington, C., Henderson, S., Ernst, C., Odom, D.T., Tanay, A., and Hadjur, S. (2015). Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep. *10*, 1297–1309.
10. Franke, M., Ibrahim, D.M., Andrey, G., Schwarzer, W., Heinrich, V., Schöpflin, R., Kraft, K., Kempfer, R., Jerković, I., Chan, W.L., et al. (2016). Formation of new chromatin domains determines pathogenicity of genomic duplications. Nature *538*, 265–269.
11. Lettice, L.A., Daniels, S., Sweeney, E., Venkataraman, S., Devenney, P.S., Gautier, P., Morrison, H., Fantes, J., Hill, R.E., and FitzPatrick, D.R. (2011). Enhancer-adoption as a mechanism of human developmental disease. Hum. Mutat. *32*, 1492–1499.
12. Weischenfeldt, J., Dubash, T., Drainas, A.P., Mardin, B.R., Chen, Y., Stütz, A.M., Waszak, S.M., Bosco, G., Halvorsen, A.R., Raeder, B., et al. (2017). Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. Nat. Genet. *49*, 65–74.
13. Willemin, A., Lopez-Delisle, L., Bolt, C.C., Gadolini, M.L., Duboule, D., and Rodriguez-Carballo, E. (2021). Induction of a chromatin boundary in vivo upon insertion of a TAD border. PLoS Genet. *17*, e1009691.
14. Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025.
15. Liang, M., Soomro, A., Tasneem, S., Abatti, L.E., Alizada, A., Yuan, X., Uusküla-Reimand, L., Antounians, L., Alvi, S.A., Paterson, A.D., et al. (2020). Enhancer-gene rewiring in the pathogenesis of Quebec platelet disorder. Blood *136*, 2679–2690.
16. Trivellin, G., Daly, A.F., Faucz, F.R., Yuan, B., Rostomyan, L., Larco, D.O., Schernthaner-Reiter, M.H., Szarek, E., Leal, L.F., Caberg, J.H., et al. (2014). Gigantism and acromegaly due to Xq26 microduplications and GPR101 mutation. N. Engl. J. Med. *371*, 2363–2374.
17. Trivellin, G., Faucz, F.R., Daly, A.F., Beckers, A., and Stratakis, C.A. (2020). HEREDITARY ENDOCRINE TUMOURS: CURRENT STATE-OF-THE-ART AND RESEARCH OPPORTUNITIES: GPR101, an orphan GPCR with roles in growth and pituitary tumorigenesis. Endocr. Relat. Cancer *27*, T87–T97.
18. Beckers, A., Lodish, M.B., Trivellin, G., Rostomyan, L., Lee, M., Faucz, F.R., Yuan, B., Choong, C.S., Caberg, J.H., Verrua, E., et al. (2015). X-linked acrogigantism syndrome: clinical profile and therapeutic responses. Endocr. Relat. Cancer *22*, 353–367.
19. Abboud, D., Daly, A.F., Dupuis, N., Bahri, M.A., Inoue, A., Chevigné, A., Ectors, F., Plenevaux, A., Pirotte, B., Beckers, A., and Hanson, J. (2020). GPR101 drives growth hormone hypersecretion and gigantism in mice via constitutive activation of G_s and $G_{q/11}$. Nat. Commun. *11*, 4752.
20. van de Werken, H.J., de Vree, P.J., Splinter, E., Holwerda, S.J., Klous, P., de Wit, E., and de Laat, W. (2012). 4C technology: protocols and data analysis. Methods Enzymol. *513*, 89–112.
21. Splinter, E., de Wit, E., van de Werken, H.J., Klous, P., and de Laat, W. (2012). Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods *58*, 221–230.
22. Noordermeer, D., Leleu, M., Schorderet, P., Joye, E., Chabaud, F., and Duboule, D. (2014). Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. eLife *3*, e02557.
23. Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., Li, D., Choudhary, M.N.K., Li, Y., Hu, M., et al. (2018). The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol. *19*, 151.
24. Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331–336.
25. Haeussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Hinrichs, A.S., Gonzalez, J.N., et al. (2019). The UCSC Genome Browser database: 2019 update. Nucleic Acids Res. *47* (D1), D853–D858.
26. Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans, M., Furey, T.S., Harte, R.A., Hsu, F., et al. (2006). The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. *34*, D590–D598.
27. Leung, D., Jung, I., Rajagopal, N., Schmitt, A., Selvaraj, S., Lee, A.Y., Yen, C.A., Lin, S., Lin, Y., Qiu, Y., et al. (2015). Integrative analysis of haplotype-resolved epigenomes across human tissues. Nature *518*, 350–354.

28. Schmitt, A.D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C.L., Li, Y., Lin, S., Lin, Y., Barr, C.L., and Ren, B. (2016). A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Rep. *17*, 2042–2059.
29. Gupta, R., Bhattacharyya, A., Agosto-Perez, F.J., Wickramasinghe, P., and Davuluri, R.V. (2011). MPromDb update 2010: an integrated resource for annotation and visualization of mammalian gene promoters and ChIP-seq experimental data. Nucleic Acids Res. *39*, D92–D97.
30. Vermunt, M.W., Reinink, P., Korving, J., de Bruijn, E., Creyghton, P.M., Basak, O., Geeven, G., Toonen, P.W., Lansu, N., Meunier, C., et al.; Netherlands Brain Bank (2014). Large-scale identification of coregulated enhancer networks in the adult human brain. Cell Rep. *9*, 767–779.
31. Mayran, A., Khetchoumian, K., Hariri, F., Pastinen, T., Gauthier, Y., Balsalobre, A., and Drouin, J. (2018). Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. Nat. Genet. *50*, 259–269.
32. Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., et al. (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford) *2017*, bax028.
33. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.
34. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
35. Mains, R.E., Blaby-Haas, C., Rheaume, B.A., and Eipper, B.A. (2018). Changes in Corticotrope Gene Expression Upon Increased Expression of Peptidylglycine α-Amidating Monooxygenase. Endocrinology *159*, 2621–2639.
36. Soneson, C., and Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. BMC Bioinformatics *14*, 91.
37. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.
38. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.
39. Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. *10*, 1523.
40. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330.
41. Trivellin, G., Bjelobaba, I., Daly, A.F., Larco, D.O., Palmeira, L., Faucz, F.R., Thiry, A., Leal, L.F., Rostomyan, L., Quezado, M., et al. (2016). Characterization of GPR101 transcript structure and expression patterns. J. Mol. Endocrinol. *57*, 97–111.
42. Pan, J.B., Hu, S.C., Shi, D., Cai, M.C., Li, Y.B., Zou, Q., and Ji, Z.L. (2013). PaGenBase: a pattern gene database for the global and dynamic understanding of gene function. PLoS ONE *8*, e80747.
43. Trivellin, G., Hernández-Ramírez, L.C., Swan, J., and Stratakis, C.A. (2018). An orphan G-protein-coupled receptor causes human gigantism and/or acromegaly: Molecular biology and clinical correlations. Best Pract. Res. Clin. Endocrinol. Metab. *32*, 125–140.
44. Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T., Kulakovskiy, I.V., Lizio, M., Itoh, M., et al.; FANTOM Consortium and the RIKEN PMI and CLST (DGT) (2014). A promoter-level mammalian expression atlas. Nature *507*, 462–470.
45. Beckers, A., Petrossians, P., Hanson, J., and Daly, A.F. (2018). The causes and consequences of pituitary gigantism. Nat. Rev. Endocrinol. *14*, 705–720.
46. Rostomyan, L., Daly, A.F., Petrossians, P., Nachev, E., Lila, A.R., Lecoq, A.L., Lecumberri, B., Trivellin, G., Salvatori, R., Moraitis, A.G., et al. (2015). Clinical and genetic characterization of pituitary gigantism: an international collaborative study in 208 patients. Endocr. Relat. Cancer *22*, 745–757.
47. Weischenfeldt, J., Symmons, O., Spitz, F., and Korbel, J.O. (2013). Phenotypic impact of genomic structural variation: insights from and for human disease. Nat. Rev. Genet. *14*, 125–138.
48. de Bruijn, S.E., Fiorentino, A., Ottaviani, D., Fanucchi, S., Melo, U.S., Corral-Serrano, J.C., Mulders, T., Georgiou, M., Rivolta, C., Pontikos, N., et al. (2020). Structural Variants Create New Topological-Associated Domains and Ectopic Retinal Enhancer-Gene Contact in Dominant Retinitis Pigmentosa. Am. J. Hum. Genet. *107*, 802–814.
49. Melo, U.S., Schöpflin, R., Acuna-Hidalgo, R., Mensah, M.A., Fischer-Zirnsak, B., Holtgrewe, M., Klever, M.K., Türkmen, S., Heinrich, V., Pluym, I.D., et al. (2020). Hi-C Identifies Complex Genomic Rearrangements and TAD-Shuffling in Developmental Diseases. Am. J. Hum. Genet. *106*, 872–884.
50. Aavikko, M., Kaasinen, E., Andersson, N., Pentinmikko, N., Sulo, P., Donner, I., Pihlajamaa, P., Kuosmanen, A., Bramante, S., Katainen, R., et al. (2021). WNT2 activation through proximal germline deletion predisposes to small intestinal neuroendocrine tumors and intestinal adenocarcinomas. Hum. Mol. Genet. *30*, 2429–2440.
51. Trivellin, G., Tirosh, A., Hernández-Ramírez, L.C., Gupta, T., Tsai-Morris, C.H., Faucz, F.R., Burgess, H.A., Feldman, B., and Stratakis, C.A. (2021). The X-linked acrogigantism-associated gene gpr101 is a regulator of early embryonic development and growth in zebrafish. Mol. Cell. Endocrinol. *520*, 111091.
52. Wise-Oringer, B.K., Zanazzi, G.J., Gordon, R.J., Wardlaw, S.L., William, C., Anyane-Yeboa, K., Chung, W.K., Kohn, B., Wisoff, J.H., David, R., and Oberfield, S.E. (2019). Familial X-Linked Acrogigantism: Postnatal Outcomes and Tumor Pathology in a Prenatally Diagnosed Infant and His Mother. J. Clin. Endocrinol. Metab. *104*, 4667–4675.
53. Wu, H.J., Landshammer, A., Stamenova, E.K., Bolondi, A., Kretzmer, H., Meissner, A., and Michor, F. (2021). Topological isolation of developmental regulators in mammalian genomes. Nat. Commun. *12*, 4897.
54. McArthur, E., and Capra, J.A. (2021). Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. Am. J. Hum. Genet. *108*, 269–283.
55. Dichmann, D.S., Fletcher, R.B., and Harland, R.M. (2008). Expression cloning in Xenopus identifies RNA-binding proteins as regulators of embryogenesis and Rbmx as necessary for neural and muscle development. Dev. Dyn. *237*, 1755–1766.
56. Tsend-Ayush, E., O'Sullivan, L.A., Grützner, F.S., Onnebo, S.M., Lewis, R.S., Delbridge, M.L., Marshall Graves, J.A., and

Ward, A.C. (2005). RBMX gene is essential for brain development in zebrafish. Dev. Dyn. *234*, 682–688.

57. Daly, A.Z., Dudley, L.A., Peel, M.T., Liebhaber, S.A., Parker, S.C.J., and Camper, S.A. (2021). Multi-omic profiling of pituitary thyrotropic cells and progenitors. BMC Biol. *19*, 76.
58. Doni Jayavelu, N., Jajodia, A., Mishra, A., and Hawkins, R.D. (2020). Candidate silencer elements for the human and mouse genomes. Nat. Commun. *11*, 1061.
59. Galupa, R., Nora, E.P., Worsley-Hunt, R., Picard, C., Gard, C., van Bemmel, J.G., Servant, N., Zhan, Y., El Marjou, F., Johanneau, C., et al. (2020). A Conserved Noncoding Locus Regulates Random Monoallelic Xist Expression across a Topological Boundary. Mol. Cell *77*, 352–367.e8.
60. Gasperini, M., Tome, J.M., and Shendure, J. (2020). Towards a comprehensive catalogue of validated and target-linked human enhancers. Nat. Rev. Genet. *21*, 292–310.
61. Will, A.J., Cova, G., Osterwalder, M., Chan, W.L., Wittler, L., Brieske, N., Heinrich, V., de Villartay, J.P., Vingron, M., Klopocki, E., et al. (2017). Composition and dosage of a multipartite enhancer cluster control developmental expression of Ihh (Indian hedgehog). Nat. Genet. *49*, 1539–1545.
62. Song, W., Sharan, R., and Ovcharenko, I. (2019). The first enhancer in an enhancer chain safeguards subsequent enhancer-promoter contacts from a distance. Genome Biol. *20*, 197.
63. Zhu, Z., Cui, W., Zhu, D., Gao, N., and Zhu, Y. (2020). Common tools for pituitary adenomas research: cell lines and primary cells. Pituitary *23*, 182–188.
64. Bogdanovic, O., and Irimia, M. (2020). José Luis Gómez-Skarmeta (1966–2020). Nat. Genet. *52*, 1267–1268.
65. Casares, F., Allende, M., de Celis, J.F., Gónzalez-Reyes, A., and Martínez-Morales, J.R. (2020). José Luis Gómez-Skarmeta (1966-2020). Development *147*, 147.

The American Journal of Human Genetics, Volume *109*

Supplemental information

Duplications disrupt chromatin architecture and rewire *GPR101*-enhancer communication in X-linked acrogigantism

Martin Franke, Adrian F. Daly, Leonor Palmeira, Amit Tirosh, Antonio Stigliano, Eszter Trifan, Fabio R. Faucz, Dayana Abboud, Patrick Petrossians, Juan J. Tena, Eleonora Vitali, Andrea G. Lania, José L. Gómez-Skarmeta, Albert Beckers, Constantine A. Stratakis, and Giampaolo Trivellin

Supplemental Information

Supplemental Figures

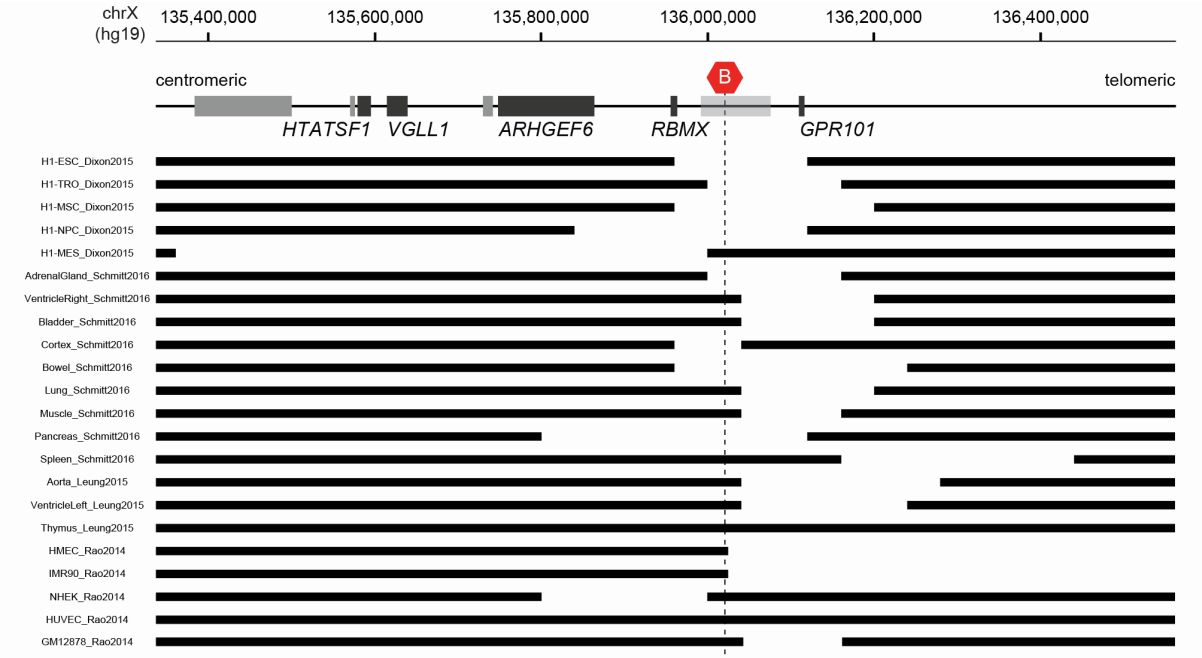

Figure S1. A tissue invariant TAD border separates *GPR101* from further centromeric genomic sequences at the X-LAG locus

TAD position (black bars) in 22 different human primary tissues and cell lines (Table S3). Predicted TAD positions are based on Hi-C data at 40 kb resolution and retrieved from the 3D genome browser.[1] Note that the TAD border (gap between TADs) separating *GPR101* from the centromeric genes is present in most analyzed tissues and cells. Although not computationally predicted in HUVEC cells and thymus, visual inspection of Hi-C data confirms the TAD border position between *RBMX* and *GPR101*.

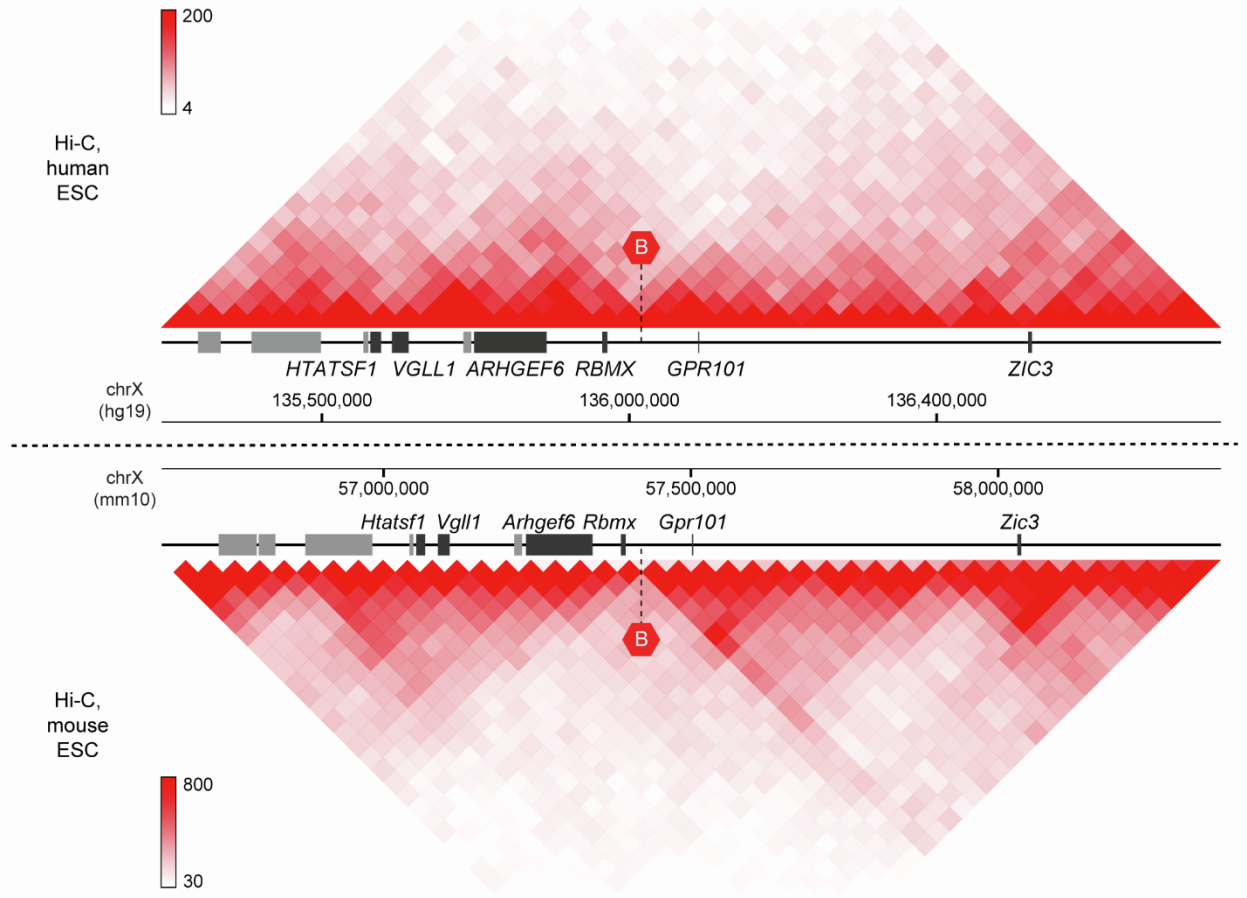

Figure S2. Conserved synteny and chromatin structure around the X-LAG locus between human and mouse

Hi-C data at 40 kb resolution from human H1-ES cells (top, NCBI Gene Expression Omnibus (GEO) GSE52457)[2] and mouse ES cells (bottom, GEO GSE96107),[3] visualized in the 3D genome browser.[1] Displayed genomic regions (human, hg19, chrX:135,240,000-136,960,000 and mouse, mm10, chrX:56,640,000-58,360,000) are centered at *GPR101*/*Gpr101*.

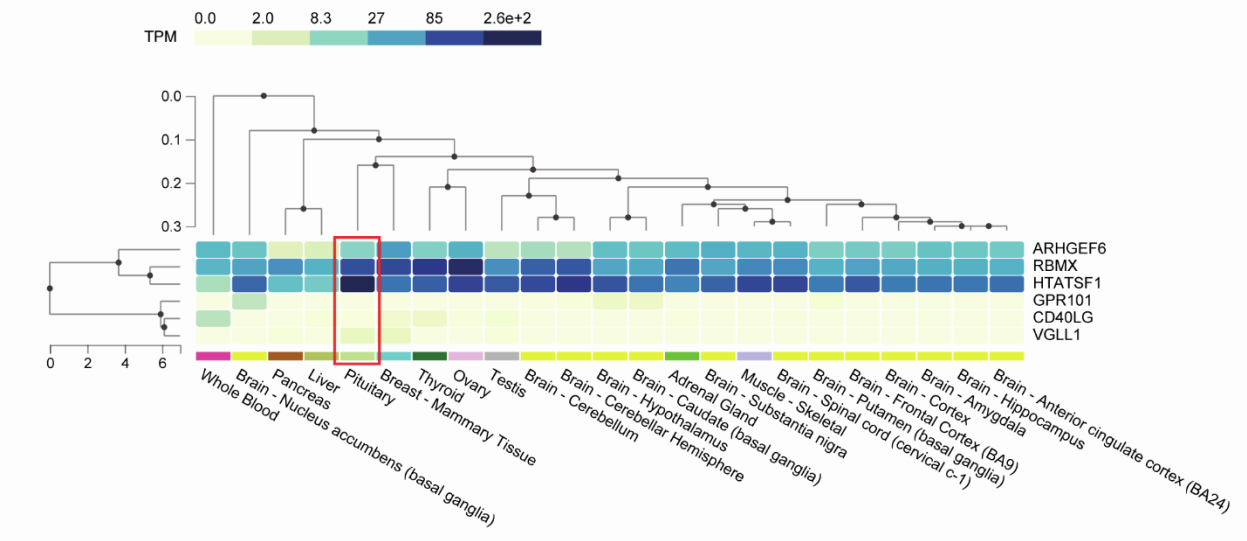

Figure S3. Gene expression patterns at the X-LAG locus

Expression data, gene and tissue clustering were obtained from the GTEx Portal, using the GTEx Multi Gene Query.[4] *ARHGEF6*, *RBMX*, and *HTATSF1* are broadly expressed, with strong expression in the pituitary. In contrast, *GPR101*, *CD40LG*, and *VGLL1* show restricted expression patterns. Note the pituitary expression (red square) of *CD40LG* (0.1 TPM*)* and *VGLL1* (1.1 TPM), but not of *GPR101.* Median TPM (Transcripts Per Million) are displayed.

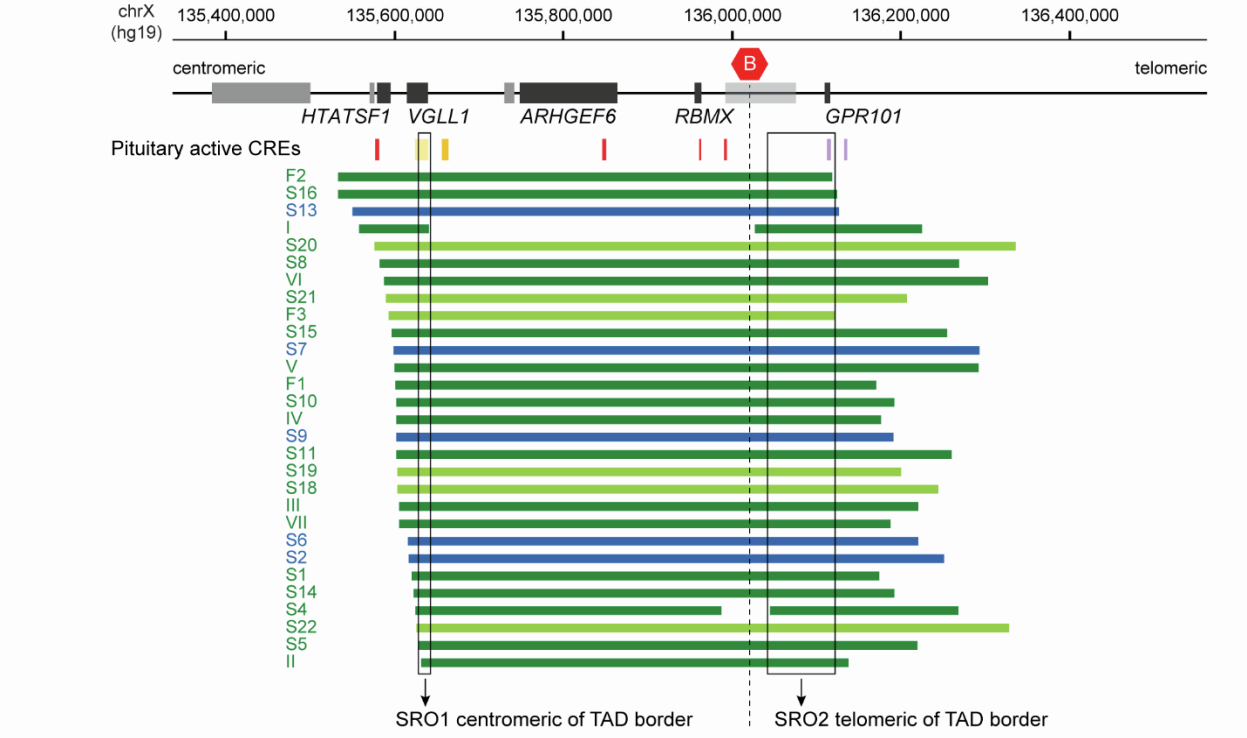

Figure S4. Smallest regions of overlap (SROs) of X-LAG duplications overlap with candidate pituitary enhancers and *GPR101* promoter elements

Genomic position of pituitary active CREs (related to Figure 5A) and published duplications at the X-LAG locus (blue and green). All duplications include genomic regions around *GPR101* and centromeric to the TAD border. SROs (black squares) at the telomeric and centromeric side of the TAD border include *GPR101* and candidate pituitary enhancers (yellow boxes) within *VGLL1*, respectively. Note that, although the duplications reported in subjects *I* and *S4* are interrupted, the rearrangements at the duplication breakpoint favor proximity of *GPR101* and candidate pituitary enhancers. In blue, duplications used for 4C-seq analysis in this study. We herein defined as *S18* and *S19* the two females reported in *Trivellin et al.*,[5] *S20* the female described in *Liang et al.*,[6] and *S21* and *S22* the two females described in *Trarbach et al.*.[7] Note that the duplication breakpoints for individuals *F3*, *S18*, *S19*, *S20*, *S21*, and *S22* (light green) are approximate, as only determined by array CGH.

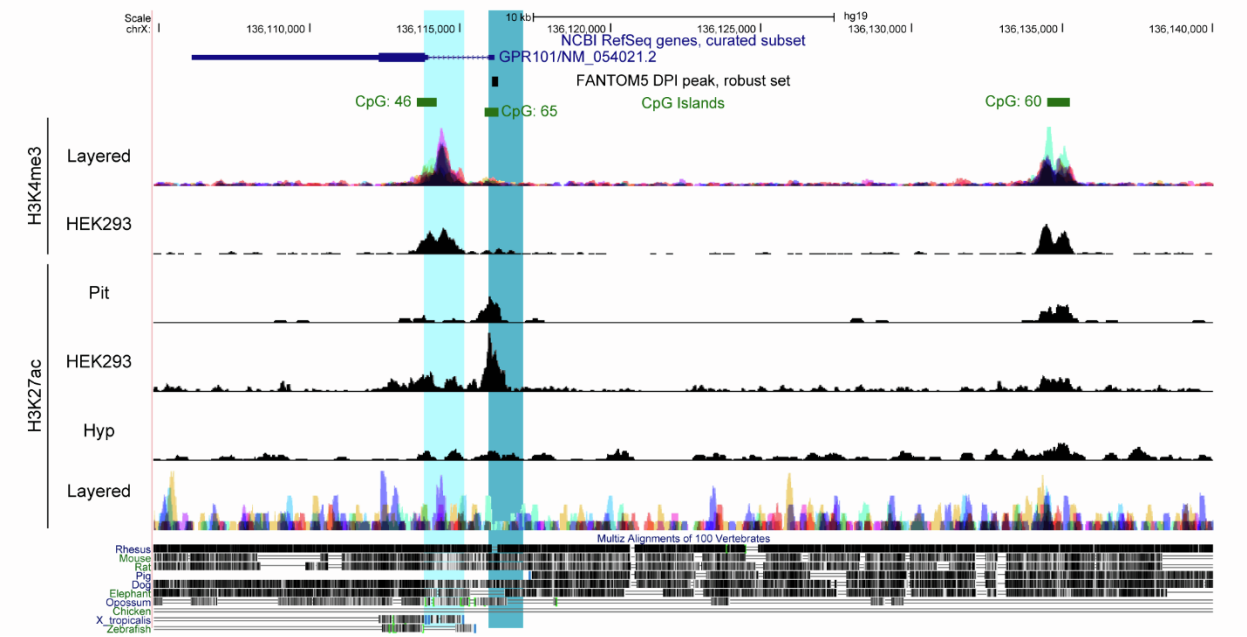

Figure S5. Location and biochemical features of the two CREs located upstream of *GPR101*

The proximal promoter sequence (light blue shading) overlaps with a CpG island (green bar) spanning *GPR101* start codon, and it includes most of the 5'UTR of isoform 1 previously described.[8] It is also marked by H3K4me3 - a promoter-specific histone modification - in several cell types, including HEK293. The distal promoter sequence (dark blue shading) spans the non-coding exon 1 of isoform 2,[8] as well as another CpG island and CAGE-seq defined TSS.[9] It also overlaps with the H3K27ac histone enhancer mark, specifically in the pituitary and in HEK293 cells (GEO GSE81696).[10] Note also the location of the far distal putative *GPR101* promoter on the right side of the figure. The pituitary (Pit) and hypothalamus (Hyp) H3K27ac ChIP-seq tracks were retrieved from GEO GSM1119175 and GSM1119152, respectively.[11] Layered H3K27ac, layered H3K4me3, and H3K4me3 in HEK293 cells tracks were retrieved from the ENCODE consortium.[12]

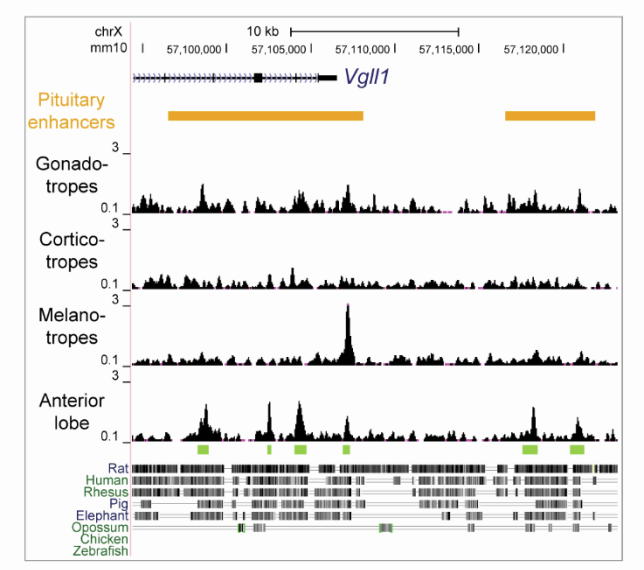

Figure S6. Sequence and functional conservation of candidate pituitary-specific enhancers in mouse pituitary cells

Equivalent position of candidate human pituitary enhancer (yellow boxes) in the mouse genome (mm10). ATAC-seq tracks from purified mouse gonadotropes, corticotropes, and melanotrope cells and anterior lobe pituitary cells (mostly composed of Pit1-dependent somatotropes and lactotropes but excluding gonadotropes) retrieved from NCBI GEO GSM3579920, GSM2324083, GSM2324084, and GSM3579919, respectively.[13] Note that enriched open chromatin regions (indicated by green boxes) in anterior lobe cells overlap with candidate pituitary enhancers identified in the human pituitary.

Supplemental Tables

Table S1. Summary of details of X-LAG individuals studied by 4C-seq and RNA-seq

X-LAG subject code	Sex	Age at disease onset (months)	Tumor classification	Technique	Reference
S2	Female	12	Macroadenoma	4C-seq	[14; 15]
S4	Female	2	Macroadenoma	RNA-seq	[14; 15]
S6	Female	6	Macroadenoma	4C-seq, RNA-seq	[14; 15]
S7	Female	36	Macroadenoma	4C-seq	[14; 15]
S9	Female	3	Macroadenoma	4C-seq	[14; 15]
S11	Male	48	Macroadenoma	RNA-seq	[14; 15]
S13	Female	48	Macroadenoma	4C-seq, RNA-seq	[14; 15]
S17	Female	12	Macroadenoma	4C-seq	na

na: not available.

Table S2. Primers used in this study

Viewpoint	Read Primer 1 (5'-3')	Primer 2 (5'-3')	Ref. Genome	RE (1st/ 2nd)	Fragend coordinates
GPR101	TCTTTCTCCCCCTCTCTCT	TTTCTGACTCTCTCCACCAC	hg19	Csp6I/ DpnII	chrX:136115379-136116771
RBMX	GCTTAAGAACAGCTAAGGGTGA	AGTCCTAACACGGGGAGA	hg19	DpnII/ Csp6I	chrX:135961852-135962755
VGLL1	CAGGGAGTTGCTTGTGAG	CATATCAGTGCCTGGCTTAT	hg19	DpnII/ Csp6I	chrX:135614980-135615359
Breakpoint for patient S6	ACAGTAGGTGTTCTGTAAACTGC	TGAGTCTATTTCTGGCCTCTC	hg19	Csp6I/ DpnII	unique fragend around breakpoint chrX:135615580-136220614

4C-seq primer sequences for inverse PCR and digestion strategy for 4C library preparation. Fragend coordinates for viewpoint (first to second restriction enzyme).

CRE	Forward Primer (5'-3')	Reverse Primer (5'-3')	Ref. Genome	Fragment size (bp)	Genomic coordinates
proximal GPR101 promoter	TCGACCATATGGGAGGGGCCATGGGAAAAAGATGTAGAG	ATCCAACGCGTTGGGGGCAACAGCCAGAGGGCG	hg19	1,300	chrX:136113819-136115148
distal GPR101 promoter	TCGACCATATGGGAGATTTCCCCTTGTGCCCTGCC	ATCCAACGCGTTGGGCTGTTGCTGCTGCAGACG	hg19	1,146	chrX:136115961-136117106
eARHGEF6-intronic	CGATCTAAGTAAGCTTCATGACCGACTGGTCTCTGAAA	CCAGCAAACGAAGCTTTAAGCATTTTTTTAAAAAAGGAAT	hg19	4,811	chrX:135846959-135851769
eVGLL1-intronic	TCTTACGCGTGCTAGCTCATGGTGAAACTAATGCAGGC	GATCGCAGATCTCGAGTGGGGAACACAGGATGCCC	hg19	1,357	chrX:135632564-135633920

Cloning primer sequences for functional evaluation of CREs.

Table S3. TAD maps in BED format for human tissues and cell types retrieved from the 3D Genome Browser

Abbreviation	Data file from 3D Genome Browser	Cellular context	Original Citation	GEO accession number
H1-ESC_Dixon2015	H1-ESC_Dixon2015-raw_TADs.txt	Embryonic stem cells	[2]	GSE52457
H1-TRO_Dixon2015	H1-TRO_Dixon2015-raw_TADs.txt	Trophoblast-like cells	[2]	GSE52457
H1-MSC_Dixon2015	H1-MSC_Dixon2015-raw_TADs.txt	Mesenchymal stem cells	[2]	GSE52457
H1-NPC_Dixon2015	H1-NPC_Dixon2015-raw_TADs.txt	Neural progenitor cells	[2]	GSE52457
H1-MES_Dixon2015	H1-MES_Dixon2015-raw_TADs.txt	Mesendoderm cells	[2]	GSE52457
AdrenalGland_Schmitt2016	AdrenalGland_Donor-AD2-raw_TADs.txt	Adrenal Gland (primary tissue)	[16]	GSE87112
VentricleRight_Schmitt2016	VentricleRight_Donor-RV3-raw_TADs.txt	Right Ventricle (primary tissue)	[16]	GSE87112
Bladder_Schmitt2016	Bladder_Donor-BL1-raw_TADs.txt	Bladder (primary tissue)	[16]	GSE87112
Cortex_Schmitt2016	Cortex_DLPFC_Donor-CO-raw_TADs.txt	Prefrontal Cortex (primary tissue)	[16]	GSE87112
Bowel_Schmitt2016	Bowel_Small_Donor-SB2-raw_TADs.txt	Small bowel (primary tissue)	[16]	GSE87112
Lung_Schmitt2016	Lung_Donor-LG1-raw_TADs.txt	Lung (primary tissue)	[16]	GSE87112
Muscle_Schmitt2016	Muscle_Psoas_Donor-PO1-raw_TADs.txt	Psoas Muscle (primary tissue)	[16]	GSE87112
Pancreas_Schmitt2016	Pancreas_Donor-PA2-raw_TADs.txt	Pancreas (primary tissue)	[16]	GSE87112
Spleen_Schmitt2016	Spleen_Donor-PX1-raw_TADs.txt	Spleen (primary tissue)	[16]	GSE87112
Aorta_Leung2015	Aorta_STL002_Leung2015-raw_TADs.txt	Aorta (primary tissue)	[17]	GSE58752
VentricleLeft_Leung2015	VentricleLeft_STL003_Leung2015-raw_TADs.txt	Left ventricle (primary tissue)	[17]	GSE58752
Thymus_Leung2015	Thymus_STL001_Leung2015-raw_TADs.txt	Thymus (primary tissue)	[17]	GSE58752
HMEC_Rao2014	HMEC_Lieberman-raw_TADs.txt	Human Mammary Epithelial Cells	[18]	GSE63525
IMR90_Rao2014	IMR90_Lieberman-raw_TADs.txt	Foetal lung fibroblast	[18]	GSE63525
NHEK_Rao2014	NHEK_Lieberman-raw_TADs.txt	Normal Human Epidermal Keratinocytes	[18]	GSE63525
GM12878_Rao2014	GM12878_Rao_2014-raw_TADs.txt	Transformed lymphoblastoid cells	[18]	GSE63525
HUVEC_Rao2014	HUVEC_Lieberman-raw_TADs.txt	Umbilical vein endothelial cells	[18]	GSE63525

Table S4. List of differentially expressed genes

Table S5. Mean raw gene counts at the X-LAG locus

Annotation	Mean (logCPM) Normal Pituitary	Mean (logCPM) X-LAG
GPR112	-1.585844278	-1.487361093
BRS3	-1.183253634	-0.424866166
HTATSF1	8.980024092	9.833832722
VGLL1	-0.263426474	0.420912497
MIR934	-4.187293202	-3.65754158
LINC00892	-3.417444506	-2.186648854
CD40LG	-2.068764489	-0.860794746
ARHGEF6	4.58984749	5.231740189
RBMX	6.883316259	7.787668371
SNORD61	-1.439633306	-0.873762095
GPR101	-3.260984837	7.927241678
ZIC3	0.336778929	0.27451518

Both genes found duplicated in at least one X-LAG individual and neighboring genes never duplicated in X-LAG (*GPR112* and *ZIC3*) are reported.

Table S6. Gene fusion analysis

Supplemental Materials and Methods

Fusion gene analysis

The fusion gene analysis was performed on the total RNA-seq data described in the main text Materials and Methods. Following QC assessment and read trimming steps, fusion gene analysis was performed using the TopHat tool and the TopHat-fusion-post function.[19] The annotated results were manually inspected for any fusion of *GPR101* with any other gene. Since no mate was detected, no threshold for fusion score was required.

Supplemental References

1. Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., Li, D., Choudhary, M.N.K., Li, Y., Hu, M., et al. (2018). The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol 19, 151.

2. Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature 518, 331-336.

3. Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.P., Tanay, A., et al. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell 171, 557-572 e524.

4. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science 369, 1318-1330.

5. Trivellin, G., and Stratakis, C.A. (2019). CD40LG duplications in patients with X-LAG syndrome commonly undergo random X-chromosome inactivation. J Allergy Clin Immunol 143, 1659.

6. Liang, H., Gong, F., Liu, Z., Yang, Y., Yao, Y., Wang, R., Wang, L., Chen, M., Pan, H., and Zhu, H. (2020). A Chinese case of X-linked acrogigantism and systemic review. Neuroendocrinology.

7. Trarbach, E.B., Trivellin, G., Grande, I.P.P., Duarte, F.H.G., Jorge, A.A.L., do Nascimento, F.B.P., Garmes, H.M., Nery, M., Mendonca, B.B., Stratakis, C.A., et al. (2021). Genetics, clinical features and outcomes of non-syndromic pituitary gigantism: experience of a single center from Sao Paulo, Brazil. Pituitary 24, 252-261.

8. Trivellin, G., Bjelobaba, I., Daly, A.F., Larco, D.O., Palmeira, L., Faucz, F.R., Thiry, A., Leal, L.F., Rostomyan, L., Quezado, M., et al. (2016). Characterization of GPR101 transcript structure and expression patterns. J Mol Endocrinol 57, 97-111.

9. FANTOM Consortium and the RIKEN PMI and CLST (DGT), Forrest, A.R., Kawaji, H., Rehli, M., Baillie, J.K., de Hoon, M.J., Haberle, V., Lassmann, T., et al. (2014). A promoter-level mammalian expression atlas. Nature 507, 462-470.

10. Savitsky, P., Krojer, T., Fujisawa, T., Lambert, J.P., Picaud, S., Wang, C.Y., Shanle, E.K., Krajewski, K., Friedrichsen, H., Kanapin, A., et al. (2016). Multivalent Histone and DNA Engagement by a PHD/BRD/PWWP Triple Reader Cassette Recruits ZMYND8 to K14ac-Rich Chromatin. Cell Rep 17, 2724-2737.

11. Vermunt, M.W., Reinink, P., Korving, J., de Bruijn, E., Creyghton, P.M., Basak, O., Geeven, G., Toonen, P.W., Lansu, N., Meunier, C., et al. (2014). Large-scale identification of coregulated enhancer networks in the adult human brain. Cell Rep 9, 767-779.

12. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57-74.

13. Mayran, A., Khetchoumian, K., Hariri, F., Pastinen, T., Gauthier, Y., Balsalobre, A., and Drouin, J. (2018). Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. Nat Genet 50, 259-269.

14. Beckers, A., Lodish, M.B., Trivellin, G., Rostomyan, L., Lee, M., Faucz, F.R., Yuan, B., Choong, C.S., Caberg, J.H., Verrua, E., et al. (2015). X-linked acrogigantism syndrome: clinical profile and therapeutic responses. Endocr Relat Cancer 22, 353-367.

15. Trivellin, G., Daly, A.F., Faucz, F.R., Yuan, B., Rostomyan, L., Larco, D.O., Schernthaner-Reiter, M.H., Szarek, E., Leal, L.F., Caberg, J.H., et al. (2014). Gigantism and acromegaly due to Xq26 microduplications and GPR101 mutation. N Engl J Med 371, 2363-2374.

16. Schmitt, A.D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C.L., Li, Y., Lin, S., Lin, Y., Barr, C.L., et al. (2016). A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell Rep 17, 2042-2059.

17. Leung, D., Jung, I., Rajagopal, N., Schmitt, A., Selvaraj, S., Lee, A.Y., Yen, C.A., Lin, S., Lin, Y., Qiu, Y., et al. (2015). Integrative analysis of haplotype-resolved epigenomes across human tissues. Nature 518, 350-354.

18. Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665-1680.

19. Kim, D., and Salzberg, S.L. (2011). TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol 12, R72.