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

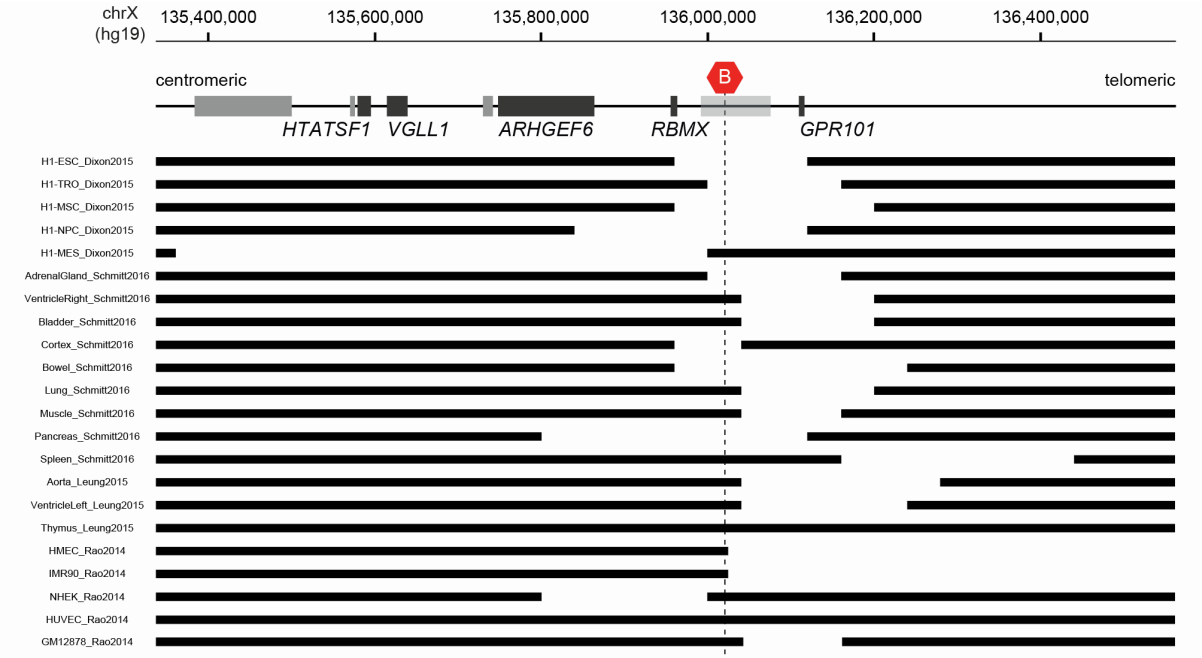

Figure S1. A tissue invariant TAD border separates *GPR101* from further centromeric genomic sequences at the X-LAG locus

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*.

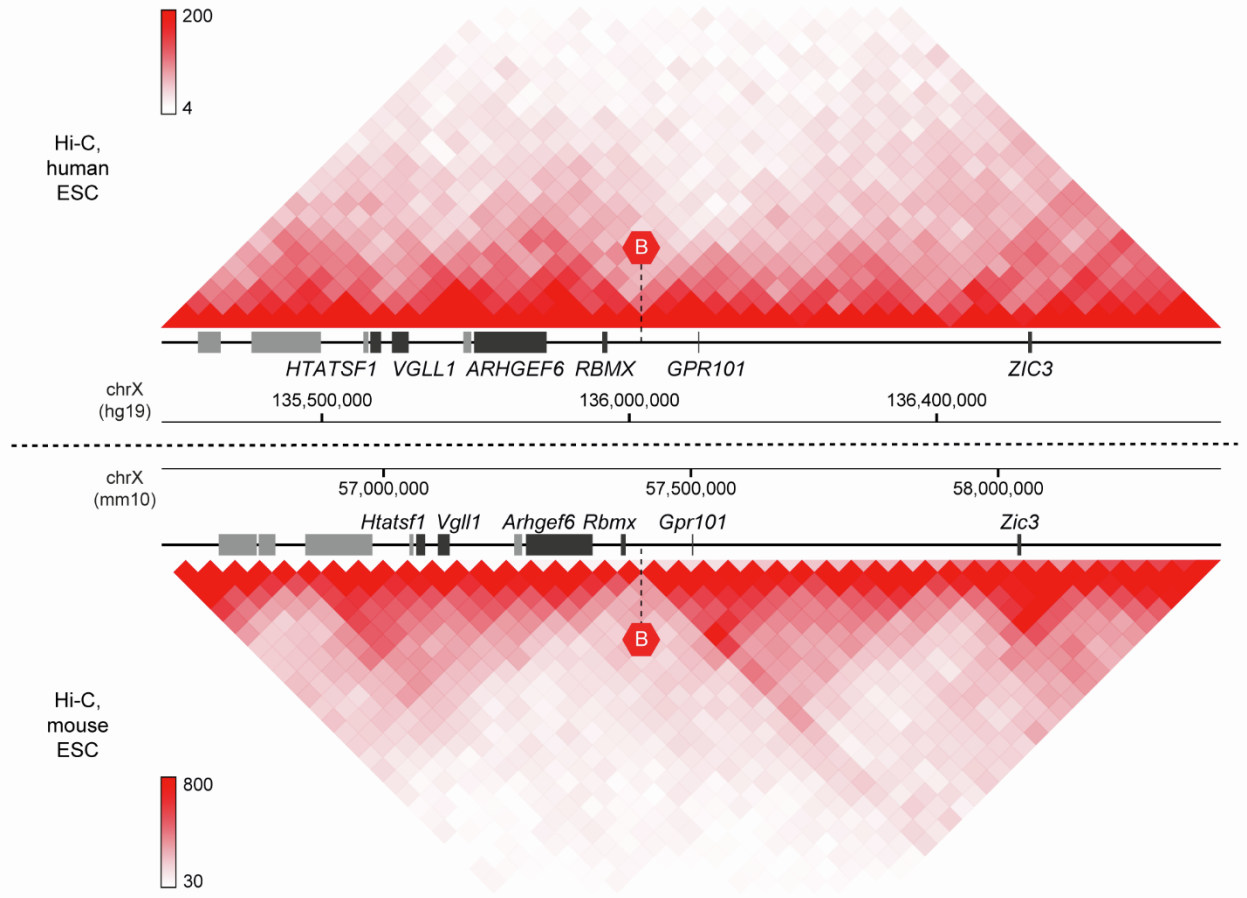

Figure S2. Conserved synteny and chromatin structure around the X-LAG locus between human and mouse

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*.

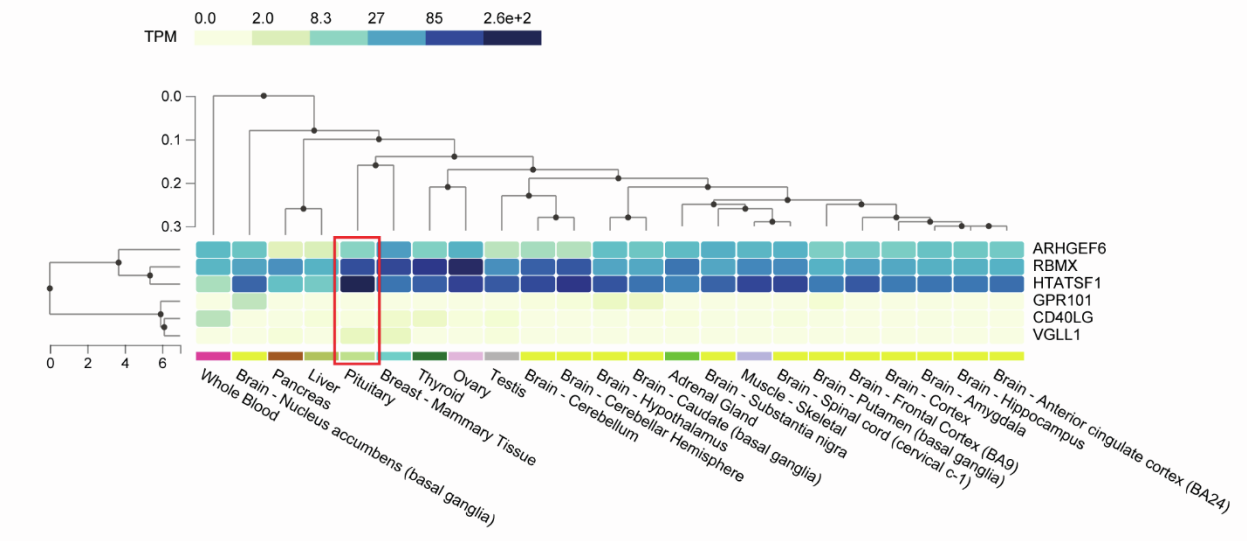

Figure S3. Gene expression patterns at the X-LAG locus

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.

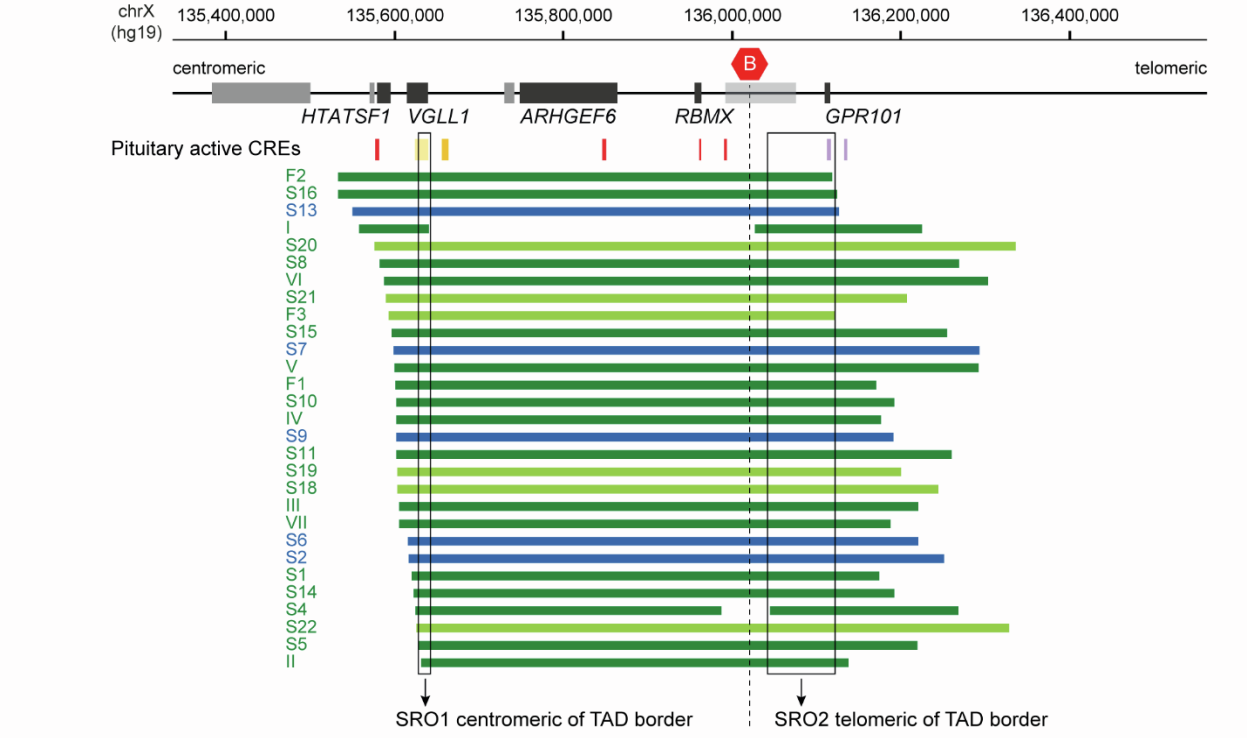

Figure S4. Smallest regions of overlap (SROs) of X-LAG duplications overlap with candidate pituitary enhancers and *GPR101* promoter elements

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.

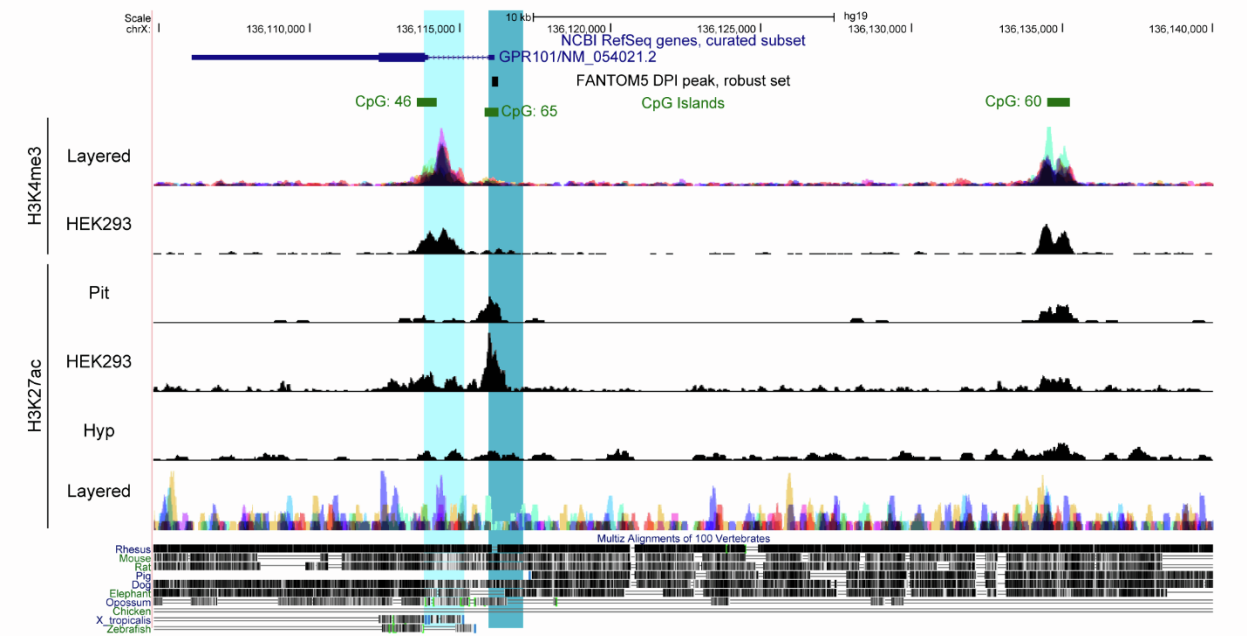

Figure S5. Location and biochemical features of the two CREs located upstream of *GPR101*

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]

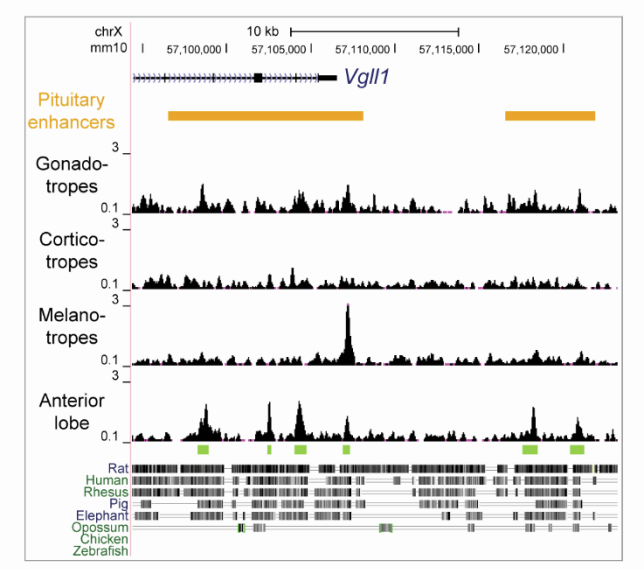

Figure S6. Sequence and functional conservation of candidate pituitary-specific enhancers in mouse pituitary cells

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.