npj Genomic Medicine

https://doi.org/10.1038/s41525-025-00548-7

Article in Press

Distinguishing benign from pathogenic duplications involving *GPR101* and *VGLL1*-adjacent enhancers in the clinical setting with the bioinformatic tool POSTRE

Received: 10 July 2025

Accepted: 27 December 2025

Cite this article as: Trivellin, G., Sánchez-Gaya, V., Grasso, A. *et al.* Distinguishing benign from pathogenic duplications involving *GPR101* and *VGLL1*-adjacent enhancers in the clinical setting with the bioinformatic tool POSTRE. *npj Genom. Med.* (2026). https://doi.org/10.1038/s41525-025-00548-7

Giampaolo Trivellin, Víctor Sánchez-Gaya, Alexia Grasso, Magdalena Pasińska, Constantine A. Stratakis, Di Milnes, Edwin P. Kirk, Albert Beckers, Andrea G. Lania, Patrick Pétrossians, Alvaro Rada-Iglesias, Martin Franke & Adrian F. Daly

Distinguishing benign from pathogenic duplications involving *GPR101* and *VGLL1*-adjacent enhancers in the clinical setting with the bioinformatic tool POSTRE

Giampaolo Trivellin[1,2], Víctor Sánchez-Gaya[3], Alexia Grasso[2], Magdalena Pasińska[4], Constantine A. Stratakis[5], Di Milnes[6], Edwin P. Kirk[7,8], Albert Beckers[9], Andrea G. Lania[1,2], Patrick Pétrossians[9], Alvaro Rada-Iglesias[3], Martin Franke[10], Adrian F. Daly[9]

1. Department of Biomedical Sciences, Humanitas University, via Rita Levi Montalcini 4, 20072 Pieve Emanuele, Milan, Italy.
2. IRCCS Humanitas Research Hospital, Translational Endocrinology and Metabolism Lab, via Manzoni 56, 20089 Rozzano, Milan, Italy.
3. Institute of Biomedicine and Biotechnology of Cantabria (IBBTEC), CSIC/Universidad de Cantabria, Albert Einstein 22, 39011 Santander, Spain.
4. Department of Clinical Genetics, Faculty of Medicine, Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University, Bydgoszcz, Poland.
5. Human Genetics & Precision Medicine, Institute of Molecular Biology and Biotechnology (IMBB), Foundation for Research and Technology Hellas (FORTH), Heraklion, Greece.
6. Genetic Health Queensland, Royal Brisbane Women's Hospital, Brisbane, QLD Australia.
7. NSW Health Pathology East Genomics, Randwick, NSW, Australia.
8. School of Clinical Medicine, University of New South Wales, Randwick NSW, Australia.
9. Department of Endocrinology, Centre Hospitalier Universitaire de Liège, University of Liège, 4000 Liège, Belgium.
10. Andalusian Center for Developmental Biology (CABD), Junta de Andaluciıa–Universidad Pablo de Olavide (UPO) – Consejo Superior de Investigaciones Científicas (CSIC), Seville, Spain.

Correspondence should be addressed to:
Dr. Giampaolo Trivellin, PhD
Department of Biomedical Sciences
Humanitas University
Via Rita Levi Montalcini 4
20072 Pieve Emanuele – Milan, Italy
E-mail: giampaolo.trivellin@hunimed.eu

Short title: Assessment of *GPR101* duplications with POSTRE software

Keywords: topologically associating domains (TADs), chromatin architecture, prenatal diagnostics, copy number variants (CNVs), X-linked acrogigantism (X-LAG)

Abstract

Structural variants (SVs) that disrupt topologically associating domains can cause disease by rewiring enhancer-promoter interactions. Duplications involving *GPR101* are known to cause X-linked acrogigantism (X-LAG) through ectopic GPR101 expression, but not all of these duplications are pathogenic. This presents a diagnostic challenge, especially in the prenatal setting. We evaluated POSTRE, a tool that predicts the regulatory impact of SVs, to distinguish pathogenic from benign *GPR101* duplications. We analyzed seven non-pathogenic duplications and 27 known X-LAG-associated duplications. To enable predictions in an X-LAG-relevant tissue, enhancer maps built using H3K27ac ChIP-seq, ATAC-seq, and RNA-seq data derived from human anterior pituitary samples (NIH research protocol 97-CH-0076, Clinicaltrials.gov Identifier NCT00001595, submitted on 11 March 1999) were integrated into POSTRE. POSTRE correctly classified all 34 duplications as benign or pathogenic. In addition, one X-LAG case with mild clinical features (i.e. severe growth hormone hypersecretion without pituitary tumorigenesis) was found to include only 2/5 *VGLL1* enhancers, whereas all typical X-LAG cases had ≥4 enhancers duplicated. This suggests that partial enhancer hijacking at *VGLL1* could explain the different clinical features in this individual. These findings support the utility of POSTRE to support diagnostic pipelines when interpreting SVs affecting chromatin architecture in pituitary disease and highlight its potential to reduce uncertainty in genetic counselling without requiring chromatin conformation capture assays.

Introduction

The spatial organization of the genome plays a fundamental role in gene regulation. At the sub-mega-base scale, chromatin architecture is partitioned into topologically associating domains (TADs)[1]. These are self-interacting regions that constrain enhancer-promoter communications and insulate genes from regulatory elements outside their TAD[1-3]. Structural variants (SVs) can disrupt TADs to alter the regulatory landscape, resulting in aberrant gene expression. This phenomenon is increasingly recognized as a pathogenic mechanism in human disease[3-7]. So-called "TADopathies"[8] have been implicated in a growing number of disorders, highlighting the importance of chromatin architecture in maintaining the fidelity of gene expression[9].

X-linked acrogigantism (X-LAG) is a TADopathy characterized by early pediatric growth hormone (GH) excess due to pituitary tumors and/or hyperplasia, leading to gigantism[10]. X-LAG is caused by duplications at chromosome Xq26.3 involving the *GPR101* gene. We demonstrated recently that these duplications can disrupt the local *GPR101* TAD structure and lead to the formation of a pathological neo-TAD. The neo-TAD brings *GPR101* into contact with ectopic enhancers, resulting in its pathological overexpression[11,12]. A critical ectopic enhancer element implicated in this mechanism resides within the *VGLL1* gene locus and is active specifically in pituitary cells[12].

Despite the growing understanding of the etiology of X-LAG, distinguishing pathogenic from benign duplications involving *GPR101* remains a major diagnostic challenge. This is especially true in the context of prenatal chromosomal microarray analysis (CMA), where *GPR101* duplications may be detected incidentally, creating the potential for significant stress, due to the challenges inherent in fetal phenotyping. Without functional data, interpreting the clinical significance of these findings is difficult, often leading to uncertainty in genetic counseling and clinical decision-making[13]. We recently showed that 4C-seq and HiC can map chromatin interactions and assess TAD integrity, thereby playing a potential diagnostic role in distinguishing between neutral and pathogenic duplications[14]. Notwithstanding those results, these techniques are highly technically specialized, labor- and time intensive, and unavailable in most clinical laboratories.

In this context, there is a pressing need for computational approaches capable of predicting the regulatory impact of structural variants using only the SV genomic

coordinates and accessible omics data (e.g., HiC, ChIP-seq, ATAC-seq, RNA-seq). One such approach is the tool Prediction Of STRuctural variant Effects (POSTRE), which uses tissue-specific genomic data to model long-range pathological mechanisms of SVs, including enhancer adoption and neo-TAD formation. By integrating TAD maps with chromatin accessibility, histone modifications, and gene expression data, POSTRE enables the prediction of regulatory disruptions caused by SVs in phenotypically-relevant cellular contexts[15]. As originally implemented, POSTRE was applicable to a limited set of congenital defects (i.e. neurodevelopmental, craniofacial, limb, and cardiac) for which genomic data in disease-relevant human tissues was analyzed and integrated. In the present study, we developed an expanded version of POSTRE that is enriched with anterior pituitary-specific data, thus making it compatible with X-LAG and, potentially, with other disorders involving the pituitary gland. We then assessed whether POSTRE could accurately discriminate pathogenic from benign duplications at the *GPR101* locus to potentially inform the clinical interpretation of SVs and their effect on TAD structure.

Results

Overview of SVs analyzed and POSTRE framework

Thirty four duplications involving *GPR101* (Figure 1) were analyzed using the POSTRE tool, with the phenotype set to “Pituitary” and the running mode set to “Standard” (Figure 2A-B). This update maintains the original POSTRE scoring model while extending its applicability to pituitary disorders through the integration of newly generated pituitary-specific functional genomics data. Illustrative examples of POSTRE outputs for the pathogenic assessment of individual X-LAG (pathogenic) and non-pathogenic *GPR101* duplications are provided in Figure 2 (panels C, D, and E) and Figure 3A.

POSTRE identifies non-X-LAG duplications as benign

In all of the non-X-LAG cases, POSTRE predicted the duplications as non-pathogenic (Table 1 and Figures 2B and 2D). The duplication in individual F4A is entirely contained within the *GPR101* TAD (intra-TAD duplication) and, accordingly, is not predicted to result in neo-TAD formation—similar to the non-pathogenic duplications we previously reported (F1A-F3A, Figure 4B)[14]. By contrast, the duplications in F5A, F6A, and F7A spanned the centromeric TAD boundary downstream of *GPR101*, resulting in the formation of a neo-TAD (Figure 4C). However, unlike in X-LAG-associated duplications, neither F5A, F6A, nor F7A included the intronic *VGLL1* enhancer cluster (Figure 1), which is critical for driving aberrant *GPR101* expression in X-LAG. All three duplications encompassed a putative enhancer located near *RBMX* (eRBMX)[12] (Figure 1), but the POSTRE enhancer calling pipeline does not predict this as an enhancer, indicating rather that this *cis*-regulatory element (CRE) alone might not be sufficient to induce *GPR101* misexpression in pituitary cells. As a result, POSTRE predicted these rearrangements to be neutral for *GPR101* regulation despite their inter-TAD configuration. This finding refines the mechanistic interpretation we recently outlined[14], where the pathogenic importance of TAD boundary disruption was emphasized, but the differential impacts of distinct enhancer inclusions were not studied experimentally.

Of note, when benchmarked against findings derived from chromatin conformation capture techniques, POSTRE correctly predicted as neutral the three non-pathogenic *GPR101* duplications that we recently reported (Table 1)[14].

POSTRE correctly classifies all X-LAG duplications as pathogenic

We also analyzed 27 X-LAG-associated contiguous duplications at the *GPR101* locus (Figure 1), and POSTRE predicted all 27 to be pathogenic, with each receiving a *GPR101* score of 0.83 (Table 1 and Figure 2C). Interestingly, one X-LAG case (patient II) had been previously described as having some unique clinical features[12,16]. Typically, X-LAG is associated with pituitary gigantism due to GH hypersecretion from anterior pituitary adenoma/hyperplasia; this is usually accompanied by prolactin excess. Patient II had early-onset GH hypersecretion and overgrowth but, despite many years of detailed clinical follow-up, never developed a pituitary tumor, nor was hyperprolactinemia present[16,17]. When compared with the other X-LAG cases, the duplication in patient II is also unique. As shown in Figure 3B, based on the ATAC peaks enriched in H3K27ac in pituitary cells, POSTRE predicts five enhancers (e1-e5) in the vicinity of *VGLL1*: four in the *VGLL1*-intronic enhancer (located in introns 2, 3, and 4 of the gene) and one at a more distal telomeric peak (Figures 3 and 4). These two enhancer loci were previously identified using publicly available, pituitary-specific, human and mouse H3K27ac ChIP-seq and ATAC-seq datasets, respectively[12,18,19]. However, the detailed dissection and obtaining the exact genomic coordinates of each individual component of the *VGLL1* intronic enhancer cluster was made possible only thanks to the integration of the newly generated human H3K27ac ChIP-seq and ATAC-seq datasets. In all the analyzed X-LAG cases, except for patient II and patient S5, all five of the *VGLL1* enhancers were duplicated (Figure 1). The duplication in patient II only involves e4 and e5, which are also the weakest enhancers based on H3K27ac levels (Figure 3C). Regarding S5, which has no atypical features of X-LAG, the duplication includes enhancers e2-e5, some of which are strong enhancers as assessed by H3K27ac levels (Figure 3C), and only excludes the most centromeric enhancer, e1. These data further emphasize the key importance of the *VGLL1* intronic enhancer cluster and for the first time suggest that the X-LAG phenotype is modulated according to the cumulative strength of these ectopic enhancers acting on the *GPR101* neo-TAD. In the case of patient II, we hypothesize that the duplication might lead to a milder enhancer adoption mechanism, thereby explaining the modified clinical phenotype[16,17].

Discussion

Structural variants altering gene-enhancer communication through TAD disruption–TADopathies– are implicated in an emerging group of genomic disorders[3,5-7]. TADopathies represent a newer diagnostic challenge in clinical genomics, as their investigation can require advanced techniques (e.g., 4C-seq/HiC), which are not optimized for diagnostic workflows. Furthermore, integrating these data with other omics datasets (e.g., ChIP-seq, RNA-seq) is time-consuming and requires expertise for processing and analysis that ultimately limits diagnostic utility[20]. In X-LAG, duplications near the *GPR101* locus are known to cause pathogenic rewiring of enhancer-promoter interactions[11,12]. However, as we showed previously and expand on here, not all *GPR101*-spanning duplications result in disease[14]. Identification of SVs of uncertain significance involving *GPR101* on prenatal genomic screening can be challenging for clinical geneticists when counselling regarding the likelihood of developing a serious condition like X-LAG. In this context, tools to distinguish between benign and pathogenic duplications at the *GPR101* locus without recourse to 4C-seq/HiC would be highly clinically desirable. In the current study, we addressed these challenges by using an *in silico* tool, POSTRE[15], to model the regulatory consequences of SVs at the *GPR101* locus. POSTRE is built upon the integration of several experimentally derived datasets that, when combined with available TAD maps, allows the creation of gene-enhancer maps in disease-relevant tissues and the possibility of predicting the consequences of SVs. By integrating new and existing pituitary-related datasets including ATAC-seq, H3K27ac ChIP-seq, and RNA-seq into POSTRE permits the correct classification of all 34 pathogenic and non-pathogenic *GPR101* duplications studied. Crucially, all seven duplications—including three from this study and four that we and others published recently[14,21]—that were not associated with pituitary dysfunction were classified as benign. This predictive accuracy, coupled with ease of use, highlights its utility for robust diagnostic modeling in the setting of TADopathies like X-LAG thereby obviating the need for complex functional chromatin capture assays.

Our results underline that pathogenicity of duplications involving *GPR101* cause X-LAG when *GPR101* TAD disruption is accompanied by inclusion of regulatory elements, principally the pituitary-active enhancer cluster located near *VGLL1* (Figure 5)[12]. The new pituitary-specific multi-omic dataset used in this study also uncovered important details regarding the *VGLL1*-adjacent enhancers that we identified and studied previously[12]. According to calling criteria in POSTRE, the intronic *VGLL1* enhancer cluster is predicted to consist of four independent enhancers that correspond to four independent ATAC-seq peaks enriched in H3K27ac. In

addition, this intronic *VGLL1* enhancer cluster is accompanied by a single distal enhancer (e5, Figures 3 and 4), that has an overlapping CTCF signal. CTCF-bound enhancers have been described across multiple systems, including elements that modulate chromatin topology or enhancer-promoter communication[22,23], and well-characterized cases such as the ZRS limb enhancer of *Shh*[24]. The presence of CTCF at e5 does not preclude enhancer activity, although the lowest H3K27ac signal observed in comparison with the other *VGLL1*-associated CREs suggests it may be less functionally relevant.

Since the initial characterization of X-LAG, *GPR101* and the *VGLL1* loci formed the smallest regions of overlap (SROs) that were shared by all affected individuals[16,25]. The mechanics of ectopic enhancers acting on the *GPR101* promoter remain to be fully explained, including the relative contributions of components within the *VGLL1* intronic enhancer cluster and elsewhere. We previously identified and tested a series of putative enhancers, including one at *RBMX* (eRBMX) that has experimental activity *in vitro*[12]. In the current study, POSTRE predictions suggest that eRBMX does not play a strong pathogenic role in X-LAG. Notwithstanding the apparent centrality of the *VGLL1*-adjacent enhancer elements in the pathogenesis of X-LAG, we cannot entirely discount that eRBMX could contribute some additive effect to *GPR101* dysregulation *in vivo*. Interestingly, a partial duplication of the *VGLL1* enhancer cluster in patient II led to an atypical form of X-LAG, with severe excess GH hypersecretion in the absence of pituitary tumor formation. No prolactin hypersecretion occurred, which is rare in X-LAG[10,16,26], and GHRH hypersecretion, that is thought to be responsible for pituitary hyperproliferation in many X-LAG cases[27], was absent. This case shares many features with the pituitary somatotrope-specific transgenic mouse model that we described in Abboud et al.[28]. That model had gigantism driven by pituitary GH and insulin-like growth factor 1 (IGF1) excess, that was driven by an up to 30-fold increase in pituitary *Gpr101* expression (in contrast, *GPR101* is increased 1000s-fold in the tumors of X-LAG patients)[12,29]. Importantly, the *Gpr101* transgenic mouse had a normal pituitary morphology and no evidence of increased proliferation, hyperplasia or tumorigenesis. Based on the results of the current study, we hypothesize that partial duplication of the intronic *VGLL1* enhancer cluster could lead to an incomplete form of X-LAG, hypothetically due to modest, somatotrope-specific elevations in *GPR101* expression. Definitive evidence regarding this will require further characterization of the enhancer sequences at *VGLL1* and elsewhere and their functional interactions among themselves (e.g., additive, synergistic, hierarchical or redundant[30,31]), as well as their actions on the *GPR101* promoter. This information would help to further refine the pathogenicity classification of SVs in

this region by tools like POSTRE. These observations demonstrate that POSTRE, when equipped with cell-type specific epigenomic and transcriptomic data, can not only aid in the discrimination between pathogenic and benign SVs, but can also provide insights to guide future functional analyses of TADopathy mechanics.

Like all *in silico* tools, POSTRE has limitations. The resolution of predictions is inherently constrained by the quality and granularity of the enhancer- and TAD maps that POSTRE relies upon. In the current version, TAD boundaries are derived from hESC data and even accounting for strong conservation of TAD boundaries across tissues, may not fully reflect the 3D chromatin architecture of anterior pituitary cells. Future versions will benefit from ongoing efforts to collect and integrate HiC or Micro-C data from pituitary tissues. Additionally, although enhancer annotations were generated for this study using primary adult human pituitary epigenomic datasets, new enhancer maps derived from genomic data at other developmental stages (bulk/single-cell resolution), and improved artificial intelligence strategies to predict enhancer-promoter interactions, could further enhance sensitivity and specificity. Finally, our dataset may be further refined as additional datasets become available. For example, p300 ChIP-seq or related coactivator-based assays could help improve enhancer annotation and functional prioritization in future iterations of POSTRE.

Beyond X-LAG, this study illustrates POSTRE's broader potential in evaluating SVs across TADopathies affecting the pituitary gland, on top of its already proven capabilities to handle this type of alterations in retinal, limb, craniofacial and neurodevelopmental disorders[15,32,33]. Such conditions are identified as the result of disrupted regulatory domains rather than coding mutations[34-37]. Incorporating tissue-relevant data into prediction algorithms allows for more nuanced interpretation of SVs that may otherwise remain unclassified when using *in silico* predictions tools that are agnostic to the cellular context[15,20]. As databases of chromatin and enhancer landscapes expand, tools like POSTRE could eventually be integrated into diagnostic pipelines.

In conclusion, the findings presented here establish POSTRE as an effective, interpretable, and scalable tool for the clinical interpretation of SVs at the *GPR101* locus. Moreover, these results reinforce the mechanistic model whereby duplications need to span both the *GPR101* TAD boundary and include the *VGLL1*-adjacent enhancers to drive *GPR101* misexpression in X-LAG.

Methods

Study population

The study population consisted of 34 individuals, of whom 27 had previously reported pathogenic *GPR101* duplications associated with X-LAG[10,12,16,25,38-40] and four non-pathogenic duplications[14,21]. Three newly identified individuals were also included (F4A-F6A), they harbored microduplications at the *GPR101* locus on chromosome Xq26.3 that were discovered incidentally during prenatal or pediatric genetic testing at sites in three different countries (Figure 1). None of the three individuals exhibited signs of pituitary hyperplasia/tumor, gigantism, or other endocrine abnormalities consistent with X-LAG at the time of assessment. Individual F4A, a female, was identified via prenatal chromosomal microarray analysis (CMA) of DNA extracted from amniotic fluid that was performed for advanced maternal age. Individual F5A, a female, was identified after CMA on chorionic villus sampling for investigation of unrelated fetal ultrasound findings. Individual F6A was diagnosed postnatally as part of a clinical workup for developmental concerns without endocrine pathology. Some details of F6A were previously reported[41]. A recently reported inter-TAD *GPR101* duplication that was incidentally detected in an unaffected kindred (herein identified as F7A)[21], was also included in the analysis.

Ethics approval and consent to participate

Subjects were recruited under the University of Liège Ethics committee approved protocol B707201420418; under the Bioethics Committee of Nicolaus Copernicus University, Toruń, Poland (NCU Committee of Bioethics KB 61/2021); and under the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health protocol 97-CH-0076 (ClinicalTrials.gov: NCT00001595, named “An Investigation of Pituitary Tumors and Related Hypothalamic Disorders”, submitted on 11 March 1999). The study was approved by the Independent Ethics Committee of the IRCCS Humanitas Research Hospital in Rozzano (Milan, Italy) and conformed to the ethical guidelines of the Declaration of Helsinki (approval no. 642/20 obtained on 23/07/2020). Written informed consent was obtained from all the subjects/guardians.

Copy number variant (CNV) analysis in new subjects

CMA was performed in F4A using embryonic DNA derived from amniotic fluid using the Agilent Technologies G3 ISCA V2 8x60K CGH microarray platform. This revealed a 366 kb duplication

at Xq26.3, corresponding to genomic coordinates chrX:136,104,659-136,470,844 (hg19). The duplicated segment included the genes *RBMX* and *GPR101* but did not encompass the *VGLL1* gene or its intronic enhancer cluster, which we previously linked to the pathogenesis of X-LAG[12]. The duplication was inherited from her healthy father, who had no history of growth or other disorders; she was born at term after a normal pregnancy and no evidence of growth disorders was noted.

DNA for F5A was extracted directly from blood and analyzed using the Illumina 850K CytoSNP v1.4 SNP microarray (50 kb mean effective resolution). F5A had a 237 kb duplication involving the region chrX:135,954,223-136,191,468 (hg19). This duplication also included *RBMX* and *GPR101*, but did not reach the *VGLL1* locus. F5A had no history of growth disorders and her hormone levels were normal.

Genomic DNAs from F6A and his healthy mother were analyzed using a clinical-grade 60K CGH array (Agilent ISCA CGH 60K, AMADID 031746), which provides genome-wide coverage with an estimated effective resolution of about 100 kb for gains and losses. The aCGH array analysis identified a 444.8 kb duplication spanning chrX:135,702,382-136,147,166 (hg19). aCGH analysis performed on F6A's mother did not detect the duplication, indicating a *de novo* origin. Confirmation of the CNV was performed using droplet digital PCR (ddPCR), with probes targeting four genes located within the duplication (*CD40LG*, *ARHGEF6*, *RBMX*, and *GPR101*) and two control genes outside the duplicated segment (*HTATSF1* and *ZIC3*). DNA from a previously diagnosed X-LAG patient harboring a constitutional duplication served as a positive control for the ddPCR study. ddPCR showed that the duplication was mosaic, with an average copy number of 1.44 for five probes located within the duplicated region (Supplementary Figure 1). This indicated that approximately 44% of peripheral blood cells carried the duplication. The affected genes included *CD40LG*, *ARHGEF6*, *RBMX*, and *GPR101*, while flanking genes such as *HTATSF1* and *ZIC3* showed a diploid copy number.

Genomic data processing (RNA-seq, ChIP-seq, ATAC-seq and HiC data) and incorporation into POSTRE

To enable accurate modeling of enhancer adoption and regulatory disruption in pituitary tissue, we generated a tissue-specific dataset (NormalPituitary) and supplemented this with publicly available transcriptomic data (NormalPituitary2) for cross-validation.

Normal adult anterior pituitary tissue samples (two males, one female) were obtained post-mortem through Cureline Inc, as previously reported[12]. These were used to generate RNA-seq, ATAC-seq, and H3K27ac ChIP-seq data.

RNA sequencing was performed on total RNA extracted from these tissues (for further details refer to Franke et al.[12]). High-quality reads were aligned to the human reference genome hg19 using Star 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. Next, prior to its incorporation into POSTRE, gene expression levels were converted to RPKMs and aggregated. In this regard, for each gene, the average expression across the three samples was taken as reference. This transcriptomic data was matched with the pituitary enhancer map to establish the NormalPituitary condition in POSTRE. In parallel, RNA-seq data normalized in the form of Transcripts Per Million (TPM)s from normal whole pituitary glands were obtained from the Genotype-Tissue Expression (GTEx) Portal. In particular, the gene_tpm_v10_pituitary.gct.gz file was downloaded in December 2024 from the “Bulk tissue expression” section in the downloads page. Next, the median expression value for each gene across all available samples was taken as reference. The GTEx data formed the basis for the NormalPituitary2 condition in POSTRE, providing an additional independent gene expression match to the enhancer maps.

The TAD coordinates used by POSTRE, in NormalPituitary and NormalPituitary2, were derived from human embryonic cells (hESC). In particular, the hg38 hESC TAD map associated with GSE116862 available in the 3D Genome Browser 2.0[42] was downloaded and lifted over to hg19 with the UCSC liftover tool. The HiC map shown in the POSTRE-associated UCSC genome browser tracks (Figure 3A) is also derived from the hESC data available in GSE116862. Specifically, the file GSM3262956_D00_HiC_Rep1.hic, available in GEO for the Sample GSM3262956 is directly linked to the UCSC sessions.

ChIP-seq profiling of the H3K27ac histone mark was performed on an additional normal adult anterior pituitary sample using the ChIP-IT High Sensitivity kit (Cat. No. V13-53040, Active Motif,) and a polyclonal anti-H3K27ac antibody (Cat. No. 39135, Active Motif). Library preparation was conducted using the Novogene NGS DNA Library Prep Set (Cat. No. PT004), and sequencing was performed on the Illumina NovaSeq X Plus platform with 150 bp paired-end reads. Sequencing reads for the H3K27ac ChIP-seq normal pituitary sample and its

matched input control (Input) were assessed using the internal trimming and mapping pipeline of the sequencing facility. After adapter removal and quality trimming, 43.0 million and 42.2 million raw read pairs were reduced to 42.6 million and 41.2 million clean read pairs for normal pituitary and Input, respectively, corresponding to clean ratios of 97.4-97.2%. Cleaned reads showed high base quality (Q30 ≥ 92–95%) and expected GC-content distributions. Alignment to the reference genome yielded mapping rates of 98.9% (normal pituitary) and 91.8% (Input), with unique-mapping rates of 98.2% and 95.5%, respectively. Duplicate removal resulted in 29.7 million unique deduplicated reads for NP (69.8% of unique reads) and 18.9 million for Input (46.0%) (Supplementary Table 1). Peak calling was performed using MACS2 v 2.2.7.1, yielding 147937 broad peaks.

ATAC-seq was performed on two anterior pituitary samples using about 50,000 cells per preparation. The ATAC-seq Kit (Cat. No. C01080006, Diagenode) was used according to the manufacturer's protocol, including Illumina-compatible library generation and indexing (UDI Index Set I, Diagenode). Sequencing was carried out at Azenta Life Sciences using the Illumina NovaSeq platform (2x150 bp), yielding about 350 million paired-end reads (~105 GB) per sample. ATAC-seq libraries were processed following standard ENCODE pipelines. Reads were trimmed with Trimmomatic v0.38 and aligned to hg38 using bowtie2. PCR/optical duplicates were marked and removed using Picard v2.18.26, and low-quality and non-primary alignments were filtered using samtools (MAPQ ≥30). Mitochondrial and unplaced contigs were removed prior to peak calling. Peaks were called using MACS2 v2.1.2 with ENCODE blacklist filtering. QC metrics, including alignment rate, duplication rate, mitochondrial content, number of peaks, and FriP, were extracted from the sample reports (Supplementary Table 2). Quality control of the ATAC-seq libraries followed the internal guidelines of the sequencing facility. Mapping statistics were evaluated from the alignment reports, requiring an acceptable proportion of read pairs to align concordantly to the reference genome. Read pairs aligning concordantly more than once were expected to represent >1% and <40% of total aligned pairs. Library complexity was assessed by the proportion of read pairs remaining after duplicate removal, with 60–80% (or higher) retention considered indicative of high-quality libraries. Mitochondrial contamination was evaluated after deduplication, with <10% of deduplicated read pairs mapping to the mitochondrial genome. Peak calls were required to yield >30,000 filtered peaks (for human or mouse samples), and signal enrichment was evaluated using the number of fragments in peaks (Fip).

The resulting H3K27ac and ATAC-seq datasets were integrated to define tissue-specific active enhancers. First, only ATAC peaks from each sample with a fold change > 4 and q value < 0.05 were kept. Next, the remaining peaks from both samples were combined and overlaping peaks were merged with the help of the reduce() function from the GenomicRanges R package (min.gapwidth=1). The remaining ATAC peak coordinates were predicted to be enhancers when located more than 5 kb from any protein-coding transcription start site (TSS) and within 500 bp of H3K27ac peaks (fold change >2 and p value <0.05). Finally, the enhancer maps were matched, in parallel, with the NormalPituitary and NormalPituitary2 gene expression datasets, giving rise to the two different conditions evaluated in POSTRE in relation to pituitary abnormalities. Given that the current version of POSTRE only accepts hg19 reference genome coordinates and the ChIP-seq and ATAC-seq available data/coordinates were from hg38, the coordinates of the predicted active enhancers were lifted over from hg38 to hg19 employing the UCSC liftover tool. The resulting coordinates were the ones used in POSTRE.

Selection of disease-relevant genes for Pituitary Disorders in POSTRE

POSTRE considers, by default, as disease-relevant genes only those that have been previously associated with the selected phenotypic category of interest (e.g., Head-Neck, Limb, Pituitary, etc.). This default criteria can be altered with the Advanced Parameters option to consider all genes related with any disease as important (i.e., Gene-PatientPheno option set to “No”).

In order to define a relevant set of pituitary disease genes, following POSTRE’s previously described guidelines[15], pituitary disease-relevant genes were defined to be those already associated in OMIM with any of the following Human Phenotype Ontology (HPO) terms: HP:0012503 (Abnormality of the pituitary gland - disorder of pituitary gland), HP:0040086 (Abnormal prolactin level), HP:0040277 (Neoplasm of the pituitary gland), HP:0032367 (Abnormal growth hormone level), HP:0030338 (Abnormal circulating gonadotropin concentration), HP:0031097 (Abnormal thyroid-stimulating hormone level), HP:0011043 (Abnormal circulating adrenocorticotropin concentration), and HP:0032481 (Abnormal pituitary glycoprotein hormone alpha subunit level). In addition, given that POSTRE also considers as relevant the phenotypical data from mice available in the Mouse Genome Informatics (MGI) website, we included orthologous genes with the following Mammalian Phenotype (MP) Ontology terms: MP:0000633 (abnormal pituitary gland morphology), MP:0005646 (abnormal pituitary gland physiology), MP:0003965 (Abnormal pituitary hormone level), MP:0003968

(Abnormal growth hormone level), MP:0003361(Abnormal circulating gonadotropin level), MP:0003971(Abnormal thyroid-stimulating hormone level) and MP:0003966 (Abnormal adrenocorticotropin level).

In silico modeling with POSTRE

The pathogenic potential of multiple CNVs was assessed using the POSTRE software available at https://github.com/vicsanga/Postre[15]. The expanded version, which includes the newly incorporated pituitary data, is currently available on request and will be directly available through POSTRE GitHub page with the final version of this manuscript.

POSTRE was run specifying the “Pituitary" phenotype category and with the rest of default parameters. A potential limitation of the default running mode is that it only considers as disease-relevant those genes already associated with the specified phenotypical category (in this case pituitary-related diseases) in established databases, such as OMIM as noted above. All duplications were analysed in POSTRE through the Single SV and Multiple SV Submission interfaces. Regarding the results provided by each mode, the Single SV analysis report included: details of the predicted enhancer adoption events, redefined TAD boundaries (neo-TADs), and additional information for the impacted target genes (illustrated in Figures 2 and 3). With respect to the Multiple SV analysis, after uploading all the SVs information through a single txt file (Supplementary Table 3), a set of tables with the aggregated results of the predictions was obtained.

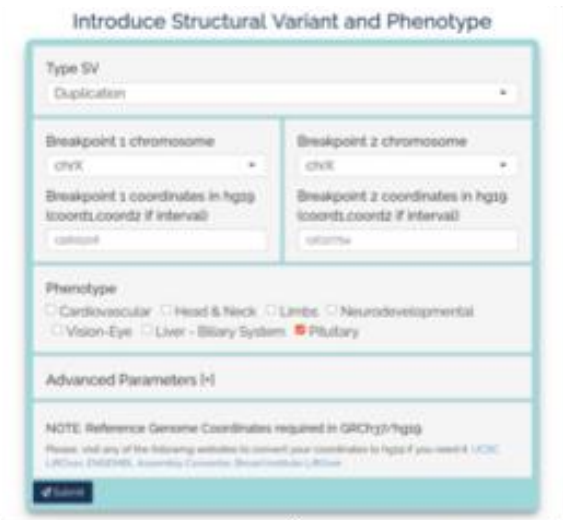
Figures

Figure 1. Overview of duplications at the Xq26.3 locus involving *GPR101*, classified by pathogenicity.

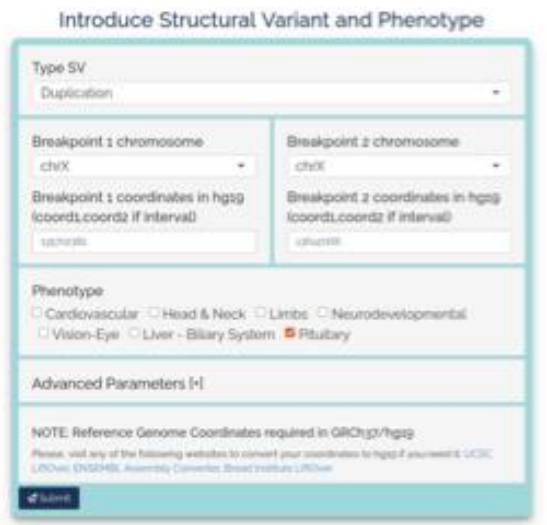
The genomic region surrounding *GPR101* (chrX:135,500,000-136,700,000, hg19) is shown with annotated protein-coding genes (blue track), CTCF binding sites from hES cells (black ChIP-seq track downloaded from GEO, accession code GSE29611), and putative CREs (six we previously described[12] based on publicly available data[18,19] shown as light green bars—the "Prior enhancers" track—and a subset predicted by POSTRE located within or distal to *VGLL1* shown as pink bars). The centromeric, tissue-invariant, *GPR101*-TAD boundary is marked by

a red vertical bar. Colored bars below represent the extent of individual duplications: yellow for three non-pathogenic intra-TAD duplications reported in Daly et al.[14] (F1A-F3A), orange for the newly identified non-pathogenic duplications presented in this study (F4A-F6A) and one from Hilditch et al.[21] (F7A), and blue for 27 continuous X-LAG-associated pathogenic duplications. Only pathogenic duplications span the *VGLL1*-adjacent enhancers. All non-pathogenic duplications, despite partial TAD disruption or inclusion of putative CREs such as the previously analyzed *RBMX* enhancer[12], which corresponds to the green bar overlapping *RBMX*, were predicted by POSTRE to be neutral. The image was derived from the UCSC genome browser (http://genome.ucsc.edu)[43].

A **X-LAG (pathogenic) variant case S22**



B **Non-pathogenic variant case F6A**



Prediction Results

C **SV predicted to be pathogenic**

Gene	NormalPituitary	NormalPituitary2	Pathomechanism
GPR101	GOF: 0.83	GOF: 0.83	Gene Duplication and Long-Range
ARHGEF6	-	-	Gene Duplication and Long-Range
RBMX	-	-	Gene Duplication and Long-Range
CD40LG	-	-	Gene Duplication and Long-Range

...

D **SV predicted to be NOT pathogenic**

Gene	NormalPituitary	NormalPituitary2	Pathomechanism
GPR101	-	-	Gene Duplication
ARHGEF6	-	-	Gene Duplication
RBMX	-	-	Gene Duplication
HTATSF1	-	-	Long-Range
FHL1	-	-	Long-Range
SLC9A6	-	-	Long-Range

...

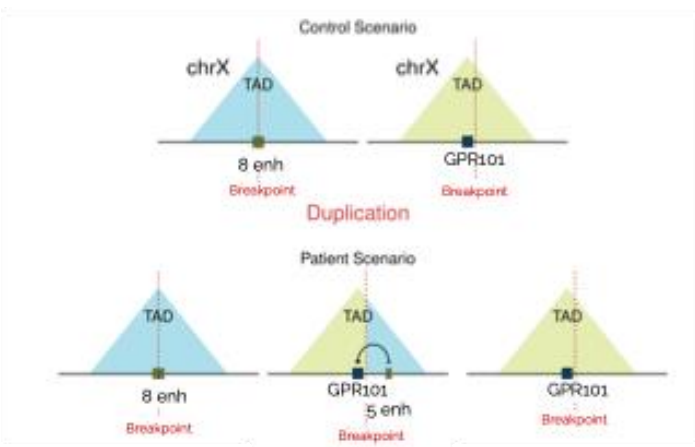
Predictions table legend:

- **NOT pathogenic.** No pathogenic event detected for the associated gene (row) and cell type (column)

GOF **GOF pathogenic.** Pathogenic Gain of Function predicted for the associated gene (row) and cell type (column)

E **Additional information is provided for the detected pathogenic events to clarify the prediction basis (accessible upon click on GOF cell)**

- **Simplified graphical abstract of SV impact**



- **Visualization of Genomic Data in UCSC genome browser**

See Figure 3 A

- **Link to gene-phenotype annotation databases OMIM-MGI**

OMIM MGI

for + info visit
github.com/vicsanga/Postre

Figure 2. Example of POSTRE output for a pathogenic (X-LAG case S22) and a non-pathogenic (F6A) duplication at the *GPR101* locus.

Panels A-E show part of the results obtained with POSTRE for the analysis of two representative duplications. A) and B) Screenshots of POSTRE submission menu for the Single SV analysis mode with the genomic coordinates introduced for X-LAG case S22 (left) and F6A (right). The duplications were evaluated using POSTRE in Standard mode with the "Pituitary" phenotype selected. This phenotype selection triggers the consideration of anterior pituitary-specific enhancer and expression data. C) and D) Main POSTRE prediction summary tables for S22 (left) and F6A (right). E) Overview of the additional information provided by POSTRE for the analyzed pathogenic event. This includes a graphical representation of the impact of S22 duplication with respect to the *GPR101* TAD. In this regard, it is represented that in the control scenario (upper part), there are no enhancers on the *GPR101* TAD, while there are eight predicted enhancers (enh) on its neighbor centromeric TAD. Due to the duplication (lower part), a new TAD is formed containing an extra copy of *GPR101* and five of the enhancers normally present in the neighbor TAD. These five enhancers represent the *VGLL1*-adjacent enhancers (Figure 3). The remaining three enhancers, which are not duplicated, can be also appreciated in Figure 3A. These enhancers (located >250Kb away from the closest duplication breakpoint) were excluded from downstream considerations in this manuscript as they are not duplicated (nor were they duplicated in any of the other reported X-LAG duplications). On the right side of the graphical abstract it is highlighted additional information provided by POSTRE, such as links to the UCSC genome browser visualization (Figure 3A) and to OMIM and MGI.

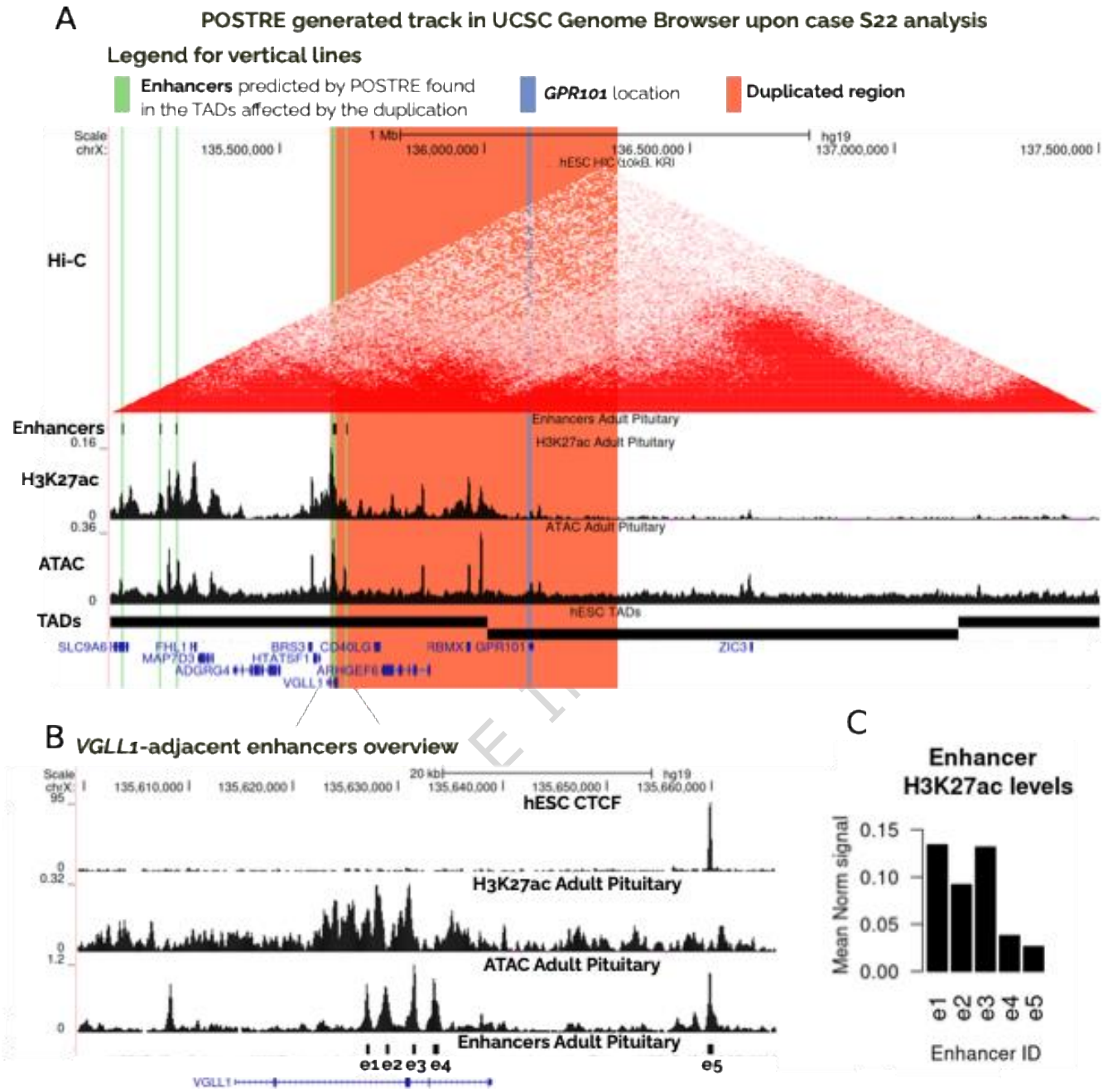


Figure 3. *VGLL1*-adjacent enhancers overview.

A) Visualization of the S22 duplication using the UCSC genome browser. The image was obtained through the adjustment (zoom in) of a UCSC link available from the POSTRE output. The duplicated area is highlighted in orange. The session also depicts bigwigs of H3K27ac ChIP-seq and ATAC-seq data from anterior pituitary, TAD maps from hESC, and highlights POSTRE predicted active enhancers with green vertical lines. It can be appreciated that the entire *VGLL1*-TAD—particularly the region surrounding the gene, including enhancer e5—exhibits strong H3K27ac enrichment, in clear contrast to the *GPR101*-TAD, where acetylation levels remain uniformly low across the locus. B) Overview of the *VGLL1*-adjacent enhancers.

According to the POSTRE enhancer calling pipeline, five enhancers, that we name e1-e5 (e1: chrX:135,626,976-135,627,240, e2: chrX:135,628,829-135,629,196, e3: chrX:135,631,401-135,631,758, e4: chrX:135,633,336-135,633,961, e5: chrX:135,659,670-135,660,230), are predicted. Four are located within *VGLL1* and one (e5) is 25 kb telomeric to the gene. C) Barplot of the mean H3K27ac levels at the e1-e5 enhancers (calculated from H3K27ac ChIP-seq bigwigs considering a window of +-500bp with respect to the enhancer coordinates). Higher H3K27ac levels are indicators of stronger enhancer activity[44].

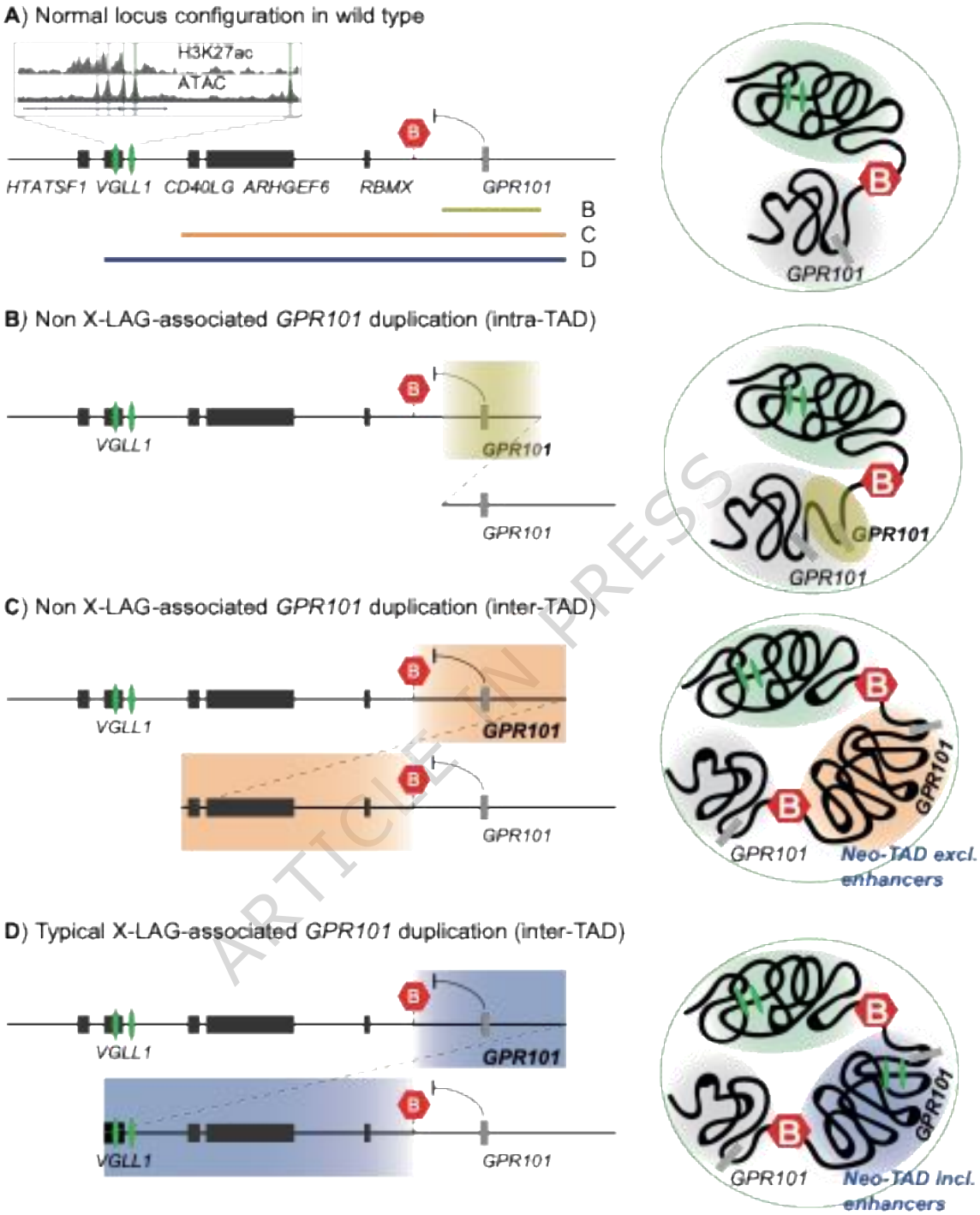


Figure 4. *GPR101* duplications reshape TADs and enhancer interactions with variable outcomes at the X-LAG locus.

Panels A–D show a schematic representation of genomic configurations at the X-LAG locus showing TAD boundaries (red hexagons), pituitary-active enhancers (green ovals), and *GPR101*-associated duplications (adapted from Daly et al.[14], with permission). For each configuration, a linear genomic view (left) and the corresponding TAD model (right) are shown.

A) Normal locus configuration with a zoom-in on the *VGLL1* region showing the *VGLL1*-intronic enhancer (a cluster of four elements), and the *VGLL1* distal enhancer, marked by ATAC-seq and H3K27ac peaks. *GPR101* and the pituitary-active *VGLL1* enhancers are separated by a TAD boundary, forming a distinct *GPR101*-TAD (grey) adjacent to a pituitary enhancer-containing TAD (green). The size and position of *GPR101*-associated duplications are indicated below.

B) Intra-TAD duplications involving *GPR101* that preserve the TAD boundary incorporate an additional *GPR101* copy within the original TAD (yellow), maintaining its separation from pituitary enhancers.

C) Inter-TAD duplications spanning *GPR101* and the TAD boundary, but not the pituitary enhancers, result in neo-TAD formation (orange) that preserves the insulation of *GPR101* from pituitary enhancers.

D) Pathogenic inter-TAD duplications spanning *GPR101* and encompassing both the TAD boundary and pituitary enhancers form a neo-TAD (blue) that establishes ectopic chromatin contacts between *GPR101* and pituitary enhancers, resulting in *GPR101* misexpression and X-LAG.

SV ID	Phenotype	Pathogenic Score	Pathogenic (Yes/No)	Causative genes	Candidate genes (Pathogenic Score)
F1A	pituitary	0.5	No		**GPR101(0.5)**,ZIC3(0.25)
F2A	pituitary	0.5	No		**GPR101(0.5)**,ZIC3(0.25)
F3A	pituitary	0.5	No		**GPR101(0.5)**,ZIC3(0.25)
F4A	pituitary	0.5	No		**GPR101(0.5)**,ZIC3(0.25)
F5A	pituitary	0.5	No		**GPR101(0.5)**,RBMX(0.49),ARHGEF6(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),VGLL1(0.19),MAP7D3(0.16),CD40LG(0.13)
F6A	pituitary	0.5	No		**GPR101(0.5)**,ARHGEF6(0.5),RBMX(0.49),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),VGLL1(0.19),CD40LG(0.17),MAP7D3(0.16)
F7A	pituitary	0.5	No		**GPR101(0.5)**,RBMX(0.49),ARHGEF6(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),VGLL1(0.19),MAP7D3(0.16),CD40LG(0.13)
F1	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),RBMX(0.62),VGLL1(0.44),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
F2	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),HTATSF1(0.62),RBMX(0.62),BRS3(0.55),VGLL1(0.44),CD40LG(0.38),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
F3	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),RBMX(0.62),HTATSF1(0.5),VGLL1(0.44),CD40LG(0.38),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S1	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),RBMX(0.62),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),VGLL1(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S2	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),RBMX(0.62),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),VGLL1(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S5	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),RBMX(0.62),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),VGLL1(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S6	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),RBMX(0.62),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),VGLL1(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S7	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),RBMX(0.62),VGLL1(0.44),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S8	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),RBMX(0.62),HTATSF1(0.5),VGLL1(0.44),CD40LG(0.38),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S9	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),RBMX(0.62),VGLL1(0.44),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S10	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),RBMX(0.62),VGLL1(0.44),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S11	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),RBMX(0.62),VGLL1(0.44),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S13	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),HTATSF1(0.62),RBMX(0.62),BRS3(0.55),VGLL1(0.44),CD40LG(0.38),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S14	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),RBMX(0.62),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),VGLL1(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S15	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),RBMX(0.62),VGLL1(0.44),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S16	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),HTATSF1(0.62),RBMX(0.62),BRS3(0.55),VGLL1(0.44),CD40LG(0.38),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S18	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),RBMX(0.62),VGLL1(0.44),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S19	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),RBMX(0.62),VGLL1(0.44),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S20	pituitary	0.83	Yes	GPR101	**GPR101(0.83)**,ARHGEF6(0.63),HTATSF1(0.62),RBMX(0.62),VGLL1(0.44),BRS3(0.4),CD40LG(0.38),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)

SV ID	Phenotype	Pathogenic Score	Pathogenic (Yes/No)	Causative genes	Candidate genes (Pathogenic Score)
S21	pituitary	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),HTATSF1(0.5),VGLL1(0.44),CD40LG(0.38),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
S22	pituitary	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),VGLL1(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
II	pituitary	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),VGLL1(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
III	pituitary	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),VGLL1(0.44),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
IV	pituitary	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),VGLL1(0.44),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
V	pituitary	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),VGLL1(0.44),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
VI	pituitary	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),HTATSF1(0.5),VGLL1(0.44),CD40LG(0.38),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)
VII	pituitary	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),VGLL1(0.44),CD40LG(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),BRS3(0.3),ZIC3(0.25),MMGT1(0.25),SMIM10(0.25),MOSPD1(0.24),ZNF449(0.22),ZNF75D(0.22),MAP7D3(0.16)

Table 1. POSTRE output for the Multiple SV analysis. Note: pathogenicity scores are shown only when above 0. Genes with score = 0 were omitted to improve readability.

Data availability

The datasets generated during this study are available at GEO, accession code GSE193113 and at the Humanitas Research Hospital and Humanitas University data repository in Zenodo (https://doi.org/10.5281/zenodo.15854045).

Acknowledgements

The work was supported in part by the following funding sources: Fondazione Telethon, Italy grant no. GGP20130 (to GT, supporting AG); grants from the Fonds d'Investissment pour la Recherche Scientifique 2018-2023 of the Centre Hospitalier Universitaire de Liège and grant number FSR-F-2023-FM from the Faculty of Medicine, University of Liège; Intramural Research Program, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH) Research project Z1A HD008920 (to CAS), USA. The project that gave rise to these results received the support of a fellowship from "La Caixa" Foundation (ID 100010434). The fellowship code is LCF/BQ/PR22/11920006 (to MF). VS-G is hired under the Generation D initiative, promoted by Red.es, an organisation attached to the Ministry for Digital Transformation and the Civil Service, for the attraction and retention of talent through grants and training contracts, financed by the Recovery, Transformation and Resilience Plan through the European Union's Next Generation funds. GT would like to acknowledge financial support by the Italian Ministry of University and Research (grant #MSCA_0000055).

The authors would like to thank the patients and families involved for their interest, generosity and patience. GT and AG also acknowledge Fondazione Humanitas per la Ricerca, the institutional recipient of the Fondazione Telethon's research grant.

Authors' contributions

GT: Conceptualization, Investigation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Funding acquisition; VS-G: Methodology, Software, Formal analysis, Data Curation, Visualization, Writing - Review & Editing; AG: Investigation, Writing - Review & Editing; MP: Resources, Writing - Review & Editing; CAS: MP: Resources, Writing - Review & Editing; DM: Resources, Writing - Review & Editing; EK: Writing - Review & Editing; AB: Writing - Review & Editing; AGL: Writing - Review &

Editing; PP: Resources, Writing - Review & Editing; AR-I: Resources, Writing - Review & Editing; MF: Conceptualization, Writing - Original Draft, Writing - Review & Editing; AFD: Conceptualization, Investigation, Writing - Original Draft, Writing - Review & Editing, Supervision, Funding acquisition.

Competing interests

AFD, CAS, and GT hold a patent on GPR101 and its function (US Patent No. 10,350,273, Treatment of Hormonal Disorders of Growth). The authors declare no other competing interests.

References

1 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012). https://doi.org:10.1038/nature11082

2 Nora, E. P., Dekker, J. & Heard, E. Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods? *Bioessays* **35**, 818-828 (2013). https://doi.org:10.1002/bies.201300040

3 Lupianez, D. G., Spielmann, M. & Mundlos, S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet* **32**, 225-237 (2016). https://doi.org:10.1016/j.tig.2016.01.003

4 Weischenfeldt, J. & Ibrahim, D. M. When 3D genome changes cause disease: the impact of structural variations in congenital disease and cancer. *Curr Opin Genet Dev* **80**, 102048 (2023). https://doi.org:10.1016/j.gde.2023.102048

5 Spielmann, M., Lupianez, D. G. & Mundlos, S. Structural variation in the 3D genome. *Nat Rev Genet* **19**, 453-467 (2018). https://doi.org:10.1038/s41576-018-0007-0

6 D'Haene, E. & Vergult, S. Interpreting the impact of noncoding structural variation in neurodevelopmental disorders. *Genet Med* **23**, 34-46 (2021). https://doi.org:10.1038/s41436-020-00974-1

7 Valton, A. L. & Dekker, J. TAD disruption as oncogenic driver. *Curr Opin Genet Dev* **36**, 34-40 (2016). https://doi.org:10.1016/j.gde.2016.03.008

8 Matharu, N. & Ahituv, N. Minor Loops in Major Folds: Enhancer-Promoter Looping, Chromatin Restructuring, and Their Association with Transcriptional Regulation and Disease. *PLoS Genet* **11**, e1005640 (2015). https://doi.org:10.1371/journal.pgen.1005640

9 Rajderkar, S. *et al.* Topologically associating domain boundaries are required for normal genome function. *Commun Biol* **6**, 435 (2023). https://doi.org:10.1038/s42003-023-04819-w

10 Trivellin, G. *et al.* Gigantism and acromegaly due to Xq26 microduplications and GPR101 mutation. *N Engl J Med* **371**, 2363-2374 (2014). https://doi.org:10.1056/NEJMoa1408028

11 Caruso, M. *et al.* Case report: Management of pediatric gigantism caused by the TADopathy, X-linked acrogigantism. *Front Endocrinol (Lausanne)* **15**, 1345363 (2024). https://doi.org:10.3389/fendo.2024.1345363

12 Franke, M. *et al.* Duplications disrupt chromatin architecture and rewire GPR101-enhancer communication in X-linked acrogigantism. *Am J Hum Genet* **109**, 553-570 (2022). https://doi.org:10.1016/j.ajhg.2022.02.002

13 Dimartino, P. *et al.* Structural Variants at the LMNB1 Locus: Deciphering Pathomechanisms in Autosomal Dominant Adult-Onset Demyelinating Leukodystrophy. *Ann Neurol* **96**, 855-870 (2024). https://doi.org:10.1002/ana.27038

14 Daly, A. F. *et al.* Chromatin conformation capture in the clinic: 4C-seq/HiC distinguishes pathogenic from neutral duplications at the GPR101 locus. *Genome Med* **16**, 112 (2024). https://doi.org:10.1186/s13073-024-01378-5

15 Sanchez-Gaya, V. & Rada-Iglesias, A. POSTRE: a tool to predict the pathological effects of human structural variants. *Nucleic Acids Res* **51**, e54 (2023). https://doi.org:10.1093/nar/gkad225

16 Iacovazzo, D. *et al.* Germline or somatic GPR101 duplication leads to X-linked acrogigantism: a clinico-pathological and genetic study. *Acta Neuropathol Commun* **4**, 56 (2016). https://doi.org:10.1186/s40478-016-0328-1

17 Burren, C. P., Williams, G., Coxson, E. & Korbonits, M. Effective Long-term Pediatric Pegvisomant Monotherapy to Final Height in X-linked Acrogigantism. *JCEM Case Rep* **1**, luad028 (2023). https://doi.org:10.1210/jcemcr/luad028

18 Mayran, A. *et al.* Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nat Genet* **50**, 259-269 (2018). https://doi.org:10.1038/s41588-017-0035-2

19 Vermunt, M. W. *et al.* Large-scale identification of coregulated enhancer networks in the adult human brain. *Cell Rep* **9**, 767-779 (2014). https://doi.org:10.1016/j.celrep.2014.09.023

20 Sanchez-Gaya, V., Mariner-Fauli, M. & Rada-Iglesias, A. Rare or Overlooked? Structural Disruption of Regulatory Domains in Human Neurocristopathies. *Front Genet* **11**, 688 (2020). https://doi.org:10.3389/fgene.2020.00688

21 Hilditch, C., Curtis, S., Cotton, S., LeBlanc, S. & De Sousa, S. Non-penetrant Xq26.3 duplication involving the invariant TAD border: clinical evidence for the VGLL1 region as the GPR101 pituitary enhancer of X-linked acrogigantism. *Pituitary* **28**, 85 (2025). https://doi.org:10.1007/s11102-025-01559-4

22 Ong, C. T. & Corces, V. G. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* **15**, 234-246 (2014). https://doi.org:10.1038/nrg3663

23 Ren, G. *et al.* CTCF-Mediated Enhancer-Promoter Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression. *Mol Cell* **67**, 1049-1058 e1046 (2017). https://doi.org:10.1016/j.molcel.2017.08.026

24 Paliou, C. *et al.* Preformed chromatin topology assists transcriptional robustness of Shh during limb development. *Proc Natl Acad Sci U S A* **116**, 12390-12399 (2019). https://doi.org:10.1073/pnas.1900672116

25 Trivellin, G., Hernandez-Ramirez, L. C., Swan, J. & Stratakis, C. A. 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 (2018). https://doi.org:10.1016/j.beem.2018.02.004

26 Daly, A. F. & Beckers, A. The Genetic Pathophysiology and Clinical Management of the TADopathy, X-Linked Acrogigantism. *Endocr Rev* **45**, 737-754 (2024). https://doi.org:10.1210/endrev/bnae014

27 Daly, A. F. *et al.* GHRH excess and blockade in X-LAG syndrome. *Endocr Relat Cancer* **23**, 161-170 (2016). https://doi.org:10.1530/ERC-15-0478

28 Abboud, D. *et al.* GPR101 drives growth hormone hypersecretion and gigantism in mice via constitutive activation of Gs and Gq/11. *Nat Commun* **11**, 4752 (2020). https://doi.org:10.1038/s41467-020-18500-x

29 Trivellin, G. *et al.* Characterization of GPR101 transcript structure and expression patterns. *J Mol Endocrinol* **57**, 97-111 (2016). https://doi.org:10.1530/JME-16-0045

30 Blayney, J. W. *et al.* Super-enhancers include classical enhancers and facilitators to fully activate gene expression. *Cell* **186**, 5826-5839 e5818 (2023). https://doi.org:10.1016/j.cell.2023.11.030

31 Peng, Y. & Zhang, Y. Enhancer and super-enhancer: Positive regulators in gene transcription. *Animal Model Exp Med* **1**, 169-179 (2018). https://doi.org:10.1002/ame2.12032

32 Plaisancie, J. *et al.* Structural Variant Disrupting the Expression of the Remote FOXC1 Gene in a Patient with Syndromic Complex Microphthalmia. *Int J Mol Sci* **25** (2024). https://doi.org:10.3390/ijms25052669

33 Hamerlinck, L. *et al.* Non-coding structural variants identify a commonly affected regulatory region steering FOXG1 transcription in early neurodevelopment. *medRxiv*, 2025.2003.2010.25323301 (2025). https://doi.org:10.1101/2025.03.10.25323301

34 Carballo-Pacoret, P., Carracedo, A. & Rodriguez-Fontenla, C. Unraveling the three-dimensional (3D) genome architecture in Neurodevelopmental Disorders (NDDs). *Neurogenetics* **25**, 293-305 (2024). https://doi.org:10.1007/s10048-024-00774-8

35 Zhang, L. *et al.* Three-dimensional genome landscape comprehensively reveals patterns of spatial gene regulation in papillary and anaplastic thyroid cancers: a study using representative cell lines for each cancer type. *Cell Mol Biol Lett* **28**, 1 (2023). https://doi.org:10.1186/s11658-022-00409-6

36 Lima, A. C. *et al.* Deletion of an evolutionarily conserved TAD boundary impacts spermatogenesis in mice. *Biol Reprod* **112**, 767-779 (2025). https://doi.org:10.1093/biolre/ioaf017

37 de Bruijn, S. E. *et al.* Structural Variants Create New Topological-Associated Domains and Ectopic Retinal Enhancer-Gene Contact in Dominant Retinitis Pigmentosa. *Am J Hum Genet* **107**, 802-814 (2020). https://doi.org:10.1016/j.ajhg.2020.09.002

38 Beckers, A. *et al.* X-linked acrogigantism syndrome: clinical profile and therapeutic responses. *Endocr Relat Cancer* **22**, 353-367 (2015). https://doi.org:10.1530/ERC-15-0038

39 Naves, L. A. *et al.* Aggressive tumor growth and clinical evolution in a patient with X-linked acro-gigantism syndrome. *Endocrine* **51**, 236-244 (2016). https://doi.org:10.1007/s12020-015-0804-6

40 Trarbach, E. B. *et al.* Genetics, clinical features and outcomes of non-syndromic pituitary gigantism: experience of a single center from Sao Paulo, Brazil. *Pituitary* **24**, 252-261 (2021). https://doi.org:10.1007/s11102-020-01105-4

41 Pasinska, M. *et al.* Clinical Importance of aCGH in Genetic Counselling of Children with Psychomotor Retardation. *Appl Clin Genet* **15**, 27-38 (2022). https://doi.org:10.2147/TACG.S357136

42 Yu, S. *et al.* The 3D Genome Browser 2.0: an enhanced online platform for visualizing and analyzing 3D genome architecture. *Nucleic Acids Res* (2025). https://doi.org:10.1093/nar/gkaf1109

43 Perez, G. *et al.* The UCSC Genome Browser database: 2025 update. *Nucleic Acids Res* **53**, D1243-D1249 (2025). https://doi.org:10.1093/nar/gkae974

44 Fulco, C. P. *et al.* Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet* **51**, 1664-1669 (2019). https://doi.org:10.1038/s41588-019-0538-0