1 **Distinguishing benign from pathogenic duplications involving *GPR101* and *VGLL1-***

2 **adjacent enhancers in the clinical setting with the bioinformatic tool POSTRE**

3

4 Giampaolo Trivellin[1, 2], Víctor Sánchez-Gaya[3], Alexia Grasso[2], Magdalena Pasińska[4],

5 Constantine A. Stratakis[5], Di Milnes[6], Edwin P. Kirk[7, 8], Albert Beckers[9], Andrea G. Lania[1,

6 2], Patrick Pétrossians[9], Alvaro Rada-Iglesias[3], Martin Franke[10], Adrian F. Daly[9]

7

8 1. Department of Biomedical Sciences, Humanitas University, via Rita Levi Montalcini 4,

9 20072 Pieve Emanuele, Milan, Italy.

10 2. IRCCS Humanitas Research Hospital, Translational Endocrinology and Metabolism

11 Lab, via Manzoni 56, 20089 Rozzano, Milan, Italy.

12 3. Institute of Biomedicine and Biotechnology of Cantabria (IBBTEC), CSIC/Universidad

13 de Cantabria, Albert Einstein 22, 39011 Santander, Spain.

14 4. Department of Clinical Genetics, Faculty of Medicine, Collegium Medicum in

15 Bydgoszcz, Nicolaus Copernicus University, Bydgoszcz, Poland.

16 5. Human Genetics & Precision Medicine, Institute of Molecular Biology and

17 Biotechnology (IMBB), Foundation for Research and Technology Hellas (FORTH),

18 Heraklion, Greece.

19 6. Genetic Health Queensland, Royal Brisbane Women's Hospital, Brisbane, QLD

20 Australia.

21 7. NSW Health Pathology East Genomics, Randwick, NSW, Australia.

22 8. School of Clinical Medicine, University of New South Wales, Randwick NSW, Australia.

23 9. Department of Endocrinology, Centre Hospitalier Universitaire de Liège, University of

24 Liège, 4000 Liège, Belgium.

25 10. Andalusian Center for Developmental Biology (CABD), Junta de Andalucıa–

26 Universidad Pablo de Olavide (UPO) – Consejo Superior de Investigaciones Científicas

27 (CSIC), Seville, Spain.

28

29   **Correspondence should be addressed to:**

30   Dr. Giampaolo Trivellin, PhD

31   Department of Biomedical Sciences

32   Humanitas University

33   Via Rita Levi Montalcini 4

34   20072 Pieve Emanuele – Milan, Italy

35   E-mail: giampaolo.trivellin@hunimed.eu

36

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

38

39   **Keywords:** topologically associating domains (TADs), chromatin architecture, prenatal

40   diagnostics, copy number variants (CNVs), X-linked acrogigantism (X-LAG)

41

42   **Abstract**

43   **Background:** Structural variants (SVs) that disrupt topologically associating domains

44   (TADs) can cause disease by rewiring enhancer-promoter interactions. Duplications

45   involving *GPR101* are known to cause X-linked acrogigantism (X-LAG) by enabling

46   aberrant expression of *GPR101* through hijacking of enhancers at *VGLL1*. However, not

47   all *GPR101*-containing duplications are pathogenic, presenting a diagnostic challenge,

48   especially in the prenatal setting.

49

50   **Methods:** We evaluated POSTRE, a tool designed to predict the regulatory impact of

51   SVs, to distinguish pathogenic from benign *GPR101* duplications. We analyzed six non-

52   pathogenic duplications, and 27 known X-LAG associated pathogenic duplications.

53   Tissue-specific enhancer maps built using H3K27ac ChIP-seq and ATAC-seq data as well

54   as gene expression data derived from human anterior pituitary samples were integrated

55   into POSTRE to enable predictions in a X-LAG relevant tissue context.

56

57   **Results:** POSTRE correctly classified all 33 duplications as benign or pathogenic. In

58   addition, one X-LAG case with mild clinical features (e.g., severe GH hypersecretion in

59   the absence of pituitary tumorigenesis) was found to include only 2/5 *VGLL1* enhancers

60   (also predicted to be the weakest enhancers), whereas all 26 typical X-LAG cases had

61   ≥4 enhancers duplicated. This suggests that milder enhancer hijacking at *VGLL1* could

62   explain the different clinical features of X-LAG in this individual.

63

64   **Conclusions:** These findings support the utility of POSTRE to support diagnostic

65   pipelines when interpreting SVs affecting chromatin architecture in pituitary disease. By

66   accurately modelling enhancer adoption in a cell type-specific context, POSTRE could

67   help to reduce uncertainty in genetic counselling and offers a rapid alternative to

68   performing chromatin conformation capture experiments.

69

70   **Abstract word count:** 248

71

Introduction

73 The spatial organization of the genome plays a fundamental role in gene regulation. At
74 the sub-mega-base scale, chromatin architecture is partitioned into topologically
75 associating domains (TADs)[1]. These are self-interacting regions that constrain enhancer-
76 promoter communications and insulate genes from regulatory elements outside their
77 TAD[1-3]. Disruption of TADs by structural variants (SVs), such as duplications, deletions,
78 inversions, or translocations can result in aberrant gene expression by altering the
79 regulatory landscape, a phenomenon increasingly recognized as a common mechanism
80 in human disease[3-7]. These so-called "TADopathies"[8] have been implicated in a growing
81 number of disorders, highlighting the importance of chromatin architecture in maintaining
82 gene expression fidelity[9].

83 One TADopathy is X-linked acrogigantism (X-LAG), a very rare condition characterized
84 by early pediatric growth hormone (GH) excess due to pituitary tumors and/or
85 hyperplasia[10]. The disease is caused by duplications at chromosome Xq26.3 involving
86 the *GPR101* gene. We demonstrated recently that these duplications can disrupt the local
87 *GPR101* TAD structure and lead to the formation of a pathological neo-TAD. The neo-
88 TAD brings *GPR101* into contact with ectopic enhancers, resulting in its pathological
89 overexpression[11,12]. A critical ectopic enhancer cluster implicated in this mechanism
90 resides within the *VGLL1* gene locus and is active specifically in pituitary cells[12]. The
91 action of this element on *GPR101* likely underlies the pathogenicity of duplications in X-
92 LAG.

93 Despite the growing understanding of these regulatory mechanisms, distinguishing
94 pathogenic from benign duplications involving *GPR101* remains a major diagnostic
95 challenge. This is especially true in the context of prenatal chromosomal microarray
96 analysis (CMA), where duplications may be detected incidentally, creating the potential
97 for significant distress, due to the challenges inherent in fetal phenotyping. Without
98 functional data, interpreting the clinical significance of these findings is difficult, often
99 leading to uncertainty in genetic counseling and clinical decision-making[13]. We recently
100 showed that 4C-seq and HiC can map chromatin interactions and assess TAD integrity
101 in these, thereby playing a diagnostic role[14]. Notwithstanding their clinical utility, we are

102   aware that these techniques are highly technically specialized, labor-intensive, and not

103   readily available in most clinical laboratories.

104   In this context, there is a pressing need for computational approaches capable of

105   predicting the regulatory impact of structural variants using only the SV genomic

106   coordinates and accessible omics data (e.g., HiC, ChIP-Seq, ATAC-seq, RNA-seq). One

107   such approach is POSTRE (Prediction Of STRuctural variant Effects), a recently

108   developed tool designed to model SV long-range pathomechanisms, including enhancer

109   adoption and neo-TAD formation, based on tissue-specific genomic data. By integrating

110   TAD maps with chromatin accessibility, histone modifications, and gene expression data,

111   POSTRE enables the prediction of regulatory disruptions caused by SVs in

112   phenotypically-relevant cellular contexts[15]. As originally implemented, POSTRE was only

113   applicable to a limited set of congenital defects (i.e. neurodevelopmental, craniofacial,

114   limb, heart) for which genomic data in disease relevant human tissues was analyzed and

115   integrated. In the present study, we developed an expanded version of POSTRE that is

116   enriched with anterior pituitary-specific data, thus making it compatible with X-LAG and,

117   potentially, with other disorders involving the pituitary gland. We then assessed whether

118   POSTRE could accurately discriminate pathogenic from benign duplications at the

119   *GPR101* locus and thereby serve as a tool to inform the clinical interpretation of TAD-

120   disrupting SVs.

121

Materials and Methods

Study population

The study population consisted of 33 individuals, of whom 30 had previously reported pathogenic *GPR101* duplications associated with X-LAG[10,12,16-20] or non-pathogenic duplications[14]. Three newly identified individuals were also included, 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. 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[21].

Copy number variant (CNV) analysis in new subjects

CMA was performed in F4A on 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 a healthy parent, 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.

152 Genomic DNAs from F6A and his healthy parent were analyzed using a clinical-grade
153 60K CGH array (Agilent ISCA CGH 60K, AMADID 031746), which provides genome-wide
154 coverage with an estimated effective resolution of about 100 kb for gains and losses. The
155 aCGH array analysis identified a 444.8 kb duplication spanning chrX:135,702,382-
156 136,147,166 (hg19). aCGH analysis performed on F6A's parent did not detect the
157 duplication, indicating a *de novo* origin. Confirmation of the CNV was performed using
158 droplet digital PCR (ddPCR), with probes targeting four genes located within the
159 duplication (*CD40LG*, *ARHGEF6*, *RBMX*, and *GPR101*) and two control genes outside
160 the duplicated segment (*HTATSF1* and *ZIC3*). DNA from a previously diagnosed X-LAG
161 patient harboring a constitutional duplication served as a positive control for the ddPCR
162 study. ddPCR showed that the duplication was mosaic, with an average copy number of
163 1.44 for five probes located within the duplicated region (Supplementary Figure 1). This
164 indicated that approximately 44% of peripheral blood cells carried the duplication. The
165 affected genes included *CD40LG*, *ARHGEF6*, *RBMX*, and *GPR101*, while flanking genes
166 such as *HTATSF1* and *ZIC3* showed a diploid copy number.

167

168 *Genomic data processing (RNA-seq, ChIP-seq, ATAC-seq and HiC data) and*
169 *incorporation into POSTRE*
170 To enable accurate modeling of enhancer adoption and regulatory disruption in pituitary
171 tissue, we generated a tissue-specific dataset (NormalPituitary) and supplemented this
172 with publicly available transcriptomic data (NormalPituitary2) for cross-validation.
173 Normal adult anterior pituitary tissue samples (two males, one female) were obtained
174 post-mortem through Cureline Inc, as previously reported[12]. These were used to generate
175 RNA-seq, ATAC-seq, and H3K27ac ChIP-seq data. RNA sequencing was performed on
176 total RNA extracted from these tissues (for further details refer to Franke et al.[12]). Prior to
177 its incorporation into POSTRE, gene expression levels were converted to RPKMs and
178 aggregated. In this regard, for each gene, the average expression across the three
179 samples was taken as reference. In parallel, RNA-seq data normalized in the form of
180 Transcripts Per Million (TPM)s, from normal whole pituitary glands were obtained from
181 the Genotype-Tissue Expression (GTEx) Portal. In particular, the
182 gene_tpm_v10_pituitary.gct.gz file was downloaded in December 2024 from the "Bulk

183   tissue expression" section in the downloads page. Next, the median expression value for

184   each gene across all available samples was taken as reference. The GTEx data formed

185   the basis for the NormalPituitary2 condition in POSTRE, providing an additional

186   independent gene expression match to the enhancer maps.

187   The TAD coordinates used by POSTRE, in NormalPituitary and NormalPituitary2, were

188   derived from the brain prefrontal cortex boundary map provided from Schmitt et al.[22] as

189   previously done for other tissues (for more details please visit the POSTRE user guide).

190   ChIP-seq profiling of the H3K27ac histone mark was performed on an additional normal

191   adult anterior pituitary sample using the ChIP-IT High Sensitivity kit (Cat. No. V13-53040,

192   Active Motif,) and a polyclonal anti-H3K27ac antibody (Cat. No. 39135, Active Motif).

193   Library preparation was conducted using the Novogene NGS DNA Library Prep Set (Cat.

194   No. PT004), and sequencing was performed on the Illumina NovaSeq X Plus platform

195   with 150 bp paired-end reads.

196   ATAC-seq was performed on two anterior pituitary samples using about 50,000 cells per

197   preparation. The ATAC-seq Kit (Cat. No. C01080006, Diagenode) was used according to

198   the manufacturer's protocol, including Illumina-compatible library generation and

199   indexing (UDI Index Set I, Diagenode). Sequencing was carried out at Azenta Life

200   Sciences using the Illumina NovaSeq platform (2x150 bp), yielding about 350 million

201   paired-end reads (~105 GB) per sample.

202   The resulting H3K27ac and ATAC-seq datasets were integrated to define tissue-specific

203   active enhancers. In this regard, enhancers were defined as ATAC peaks (fold change>4

204   and q value <0.05) located more than 5 kb from any protein coding transcription start site

205   (TSS) and within 500 bp of H3K27ac peaks (fold change >2 and p value <0.05). Finally,

206   the enhancer maps were matched, in parallel, with the NormalPituitary and

207   NormalPituitary2 gene expression datasets, giving rise to the two different conditions

208   evaluated in POSTRE in relation with Endocrine abnormalities.

209

210   *In silico modeling with POSTRE*

211   The pathogenic potential of multiple CNVs was assessed using the POSTRE software, a

212   computational tool, available at https://github.com/vicsanga/Postre, designed to predict

213   the regulatory impact of SVs that alter enhancer-promoter interactions[15]. To do so,

214 POSTRE utilizes gene expression and enhancer maps from disease relevant cell types
215 and combines them with available TAD maps. Since the first version of POSTRE did not
216 present genomic data relevant for the interpretation of pituitary related diseases, an
217 expanded version has been created in this project, as described in the previous Method
218 section. The expanded version, which includes the newly incorporated pituitary data, is
219 currently available on request and will be directly available through POSTRE GitHub page
220 upon peer-reviewed publication of this manuscript.

221 Regarding the details of POSTRE usage, it was run specifying the "Endocrine" phenotype
222 category and with the rest of default parameters. A potential limitation of the default
223 running mode is that it only considers as disease relevant those genes already associated
224 with the specified phenotypical category (in this case endocrine-related diseases) in
225 established databases, such as OMIM. In this regard, higher sensitivity analyses allowing
226 broader gene inclusion were also explored by disabling the patient phenotype filter (i.e.,
227 Gene-PatientPheno option set to No).

228 All duplications were analysed in POSTRE through the Single SV and Multiple SV
229 Submission interfaces. Regarding the results provided by each mode, the Single SV
230 analysis report included: details of the predicted enhancer adoption events, redefined
231 TAD boundaries (neo-TADs), and additional information for the impacted target genes
232 (illustrated in Figure 2). With respect to the Multiple SV analysis, after uploading all the
233 SVs information through a single txt file (Supplementary Table 1), a set of tables with the
234 aggregated results of the predictions were obtained. POSTRE also allows users to upload
235 their own TAD map to perform the interpretation of the SVs analyzed but this feature was
236 not used, since the TAD map already selected was valid for the locus of interest.

237

Results

Thirty three duplications involving *GPR101* (Figure 1) were analyzed using the POSTRE tool, with the phenotype set to "Endocrine" and the running mode set to "Standard" (Figure 2). This configuration integrates enhancer maps derived from ATAC-seq and H3K27ac ChIP-seq data generated from human anterior pituitary cells, alongside RNA-seq profiles specific to this tissue (see Methods). The analysis also relies on prefrontal cortex-derived TAD maps and disease-gene annotations obtained from different databases, such as OMIM[23] and the Mouse Genome Database[24]. As we established previously, owing to the overall conservation of TADs across cell types, TAD maps from other human tissues can provide informative data on gene regulatory domains and the prediction of long range pathological mechanisms related to SVs[15].

In all three of the new non-X-LAG cases, POSTRE predicted the duplications as non-pathogenic (Table 1 and Figure 3). The duplication in individual F4A is entirely contained within the *GPR101* TAD and, accordingly, is not predicted to result in neo-TAD formation—similar to the non-pathogenic duplications we previously reported[14]. By contrast, the duplications in F5A and F6A spanned the centromeric TAD boundary downstream of *GPR101*, resulting in the formation of a neo-TAD. However, unlike in X-LAG associated duplications neither F5A nor F6A included the intronic *VGLL1* enhancer cluster, which is thought to be critical for driving aberrant *GPR101* expression. Both duplications encompassed a putative enhancer located near *RBMX* (eRBMX)[12]*,* but the POSTRE enhancer calling pipeline does not predict this as one, indicating that this *cis-*regulatory element (CRE) alone might not be sufficient to induce *GPR101* misexpression. 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 impact of distinct enhancer inclusions was not examined in detail. POSTRE predictions were also benchmarked against the three non-pathogenic *GPR101* duplications that we recently reported (Table 1)[14]. All three were correctly predicted to be non-pathogenic, in line with the original experimental observations using chromatin conformation capture techniques.

268   We analyzed 27 X-LAG-associated contiguous duplications at the *GPR101* locus, and

269   POSTRE predicted all 27 to be pathogenic, with each receiving a *GPR101* score of 0.83

270   (Table 1 and Figure 2). Interestingly, one X-LAG case (patient II) had been previously

271   described as having some unique clinical features[12,17]. Typically, X-LAG is associated

272   with pituitary gigantism due to GH hypersecretion from anterior pituitary

273   adenoma/hyperplasia. Patient II had early-onset GH hypersecretion and overgrowth, but

274   despite many years of detailed clinical follow-up never developed a pituitary tumor. When

275   compared with the other X-LAG cases, the duplication in patient II is also unique. As

276   shown in Figure 4, based on the ATAC peaks enriched in H3K27ac in pituitary cells,

277   POSTRE predicts five enhancers (e1-e5) in the vicinity of *VGLL1*: four in the VGLL1-

278   intronic enhancer (located in introns 2, 3, and 4 of the gene) and one in a more distal

279   telomeric peak (Figure 4 A). These two enhancer loci were previously identified using

280   publicly available, pituitary-specific, human and mouse H3K27ac ChIP-seq and ATAC-

281   seq datasets, respectively[12]. In all the analyzed X-LAG cases, except for patient II and

282   patient S5, all five of the *VGLL1* enhancers were duplicated. The duplication in patient II

283   only involves e4 and e5, which are also the weakest enhancers based on H3K27ac levels

284   (Figure 4B). Regarding S5, which has no atypical features of X-LAG, the duplication

285   includes enhancers e2-e5, some of which are strong by H3K27ac levels, and only

286   excludes the most centromeric enhancer, e1 (Figure 4). These data further emphasize

287   the key importance of the *VGLL1* intronic enhancer cluster and for the first time suggests

288   that the X-LAG phenotype is modulated according to the cumulative strength of these

289   ectopic enhancers acting on the *GPR101* neo-TAD. In the case of patient II, the

290   duplication might lead to a milder enhancer adoption mechanism and, thereby explain

291   the modified clinical phenotype[17,25].

292

Discussion

293

294 Structural variants altering gene-enhancer communication through TAD disruption are
295 implicated in an emerging group of genomic disorders[3,5-7]. These disorders, which alter
296 regulatory landscapes, present a significant and growing diagnostic challenge in clinical
297 genomics. Their investigation often requires advanced techniques like 4C-seq and HiC,
298 which are not yet optimized for routine diagnostic workflows. Furthermore, integrating
299 these data with other omics datasets (e.g., ChIP-Seq, RNA-Seq) demands specialized
300 genomic and bioinformatics expertise for processing and analysis. This complexity
301 ultimately limits diagnostic yields[26]. In the case of X-LAG, duplications near the *GPR101*
302 locus are known to cause pathogenic rewiring of enhancer-promoter interactions[11,12].
303 However, as shown previously and expanded on in this study, not all *GPR101*-spanning
304 duplications result in disease, and data on the frequency of 3D genome rearrangements
305 in the general population is unavailable[14]. In parallel, the use of genomic techniques for
306 prenatal screening leads to the identification of SVs of uncertain significance. As not all
307 *GPR101*-related duplications lead to X-LAG, this creates a challenge for clinical
308 geneticists when counselling patients, and misclassification can lead to anxiety and
309 potential misdiagnosis. The ability to reliably distinguish benign from pathogenic
310 duplications at the *GPR101* locus is, therefore, highly clinically relevant, especially in
311 prenatal and pediatric contexts. Our data provide further support for the concept that
312 pathogenicity is not solely determined by TAD disruption or duplication size, but critically
313 depends on inclusion of specific regulatory elements, particularly the pituitary-active
314 enhancer cluster located in the vicinity of the *VGLL1* gene[12].

315 In this study, we demonstrate the utility of POSTRE[15], an *in silico* tool designed to model
316 the regulatory consequences of SVs, at the *GPR101* locus. POSTRE is built upon the
317 integration of several experimentally derived datasets, including ATAC-seq, H3K27ac
318 ChIP-seq, and RNA-seq. These datasets, when combined with available TAD maps,
319 allow the creation of gene-enhancer maps in disease relevant tissues and the possibility
320 of predicting the consequences of their alteration through SVs. While chromatin
321 conformation capture techniques such as 4C-seq and HiC provide direct evidence of
322 three-dimensional genome organization, they are resource-intensive and impractical in
323 routine diagnostic workflows, particularly those that are time-sensitive like prenatal

324   testing. POSTRE complements these experimental approaches by enabling rapid,

325   reproducible predictions of pathogenic enhancer adoption requiring only, from the user,

326   the SV coordinates. When equipped with anterior pituitary-specific datasets, POSTRE

327   correctly classified all known pathogenic X-LAG and non-pathogenic *GPR101*

328   duplications. Crucially, all six duplications—including three from this study and three from

329   our recent work[14]—that were not associated with pituitary dysfunction were classified as

330   benign. This predictive accuracy, coupled with ease of use, highlights its potential for

331   robust diagnostic modeling, particularly when functional assays may be unfeasible or

332   unavailable.

333   The clinical implications of these findings are especially relevant in time-sensitive

334   contexts such as prenatal diagnosis. In multiple individuals described here and

335   previously, chromosome Xq26.3 duplications were detected through prenatal microarray

336   testing, and were flagged for review based on their proximity to *GPR101* and its

337   association with X-LAG. However, absence of the *VGLL1*-adjacent enhancers, as

338   confirmed by POSTRE modeling, provides strong evidence against a pathogenic

339   interpretation. This not only could aid genetic counseling but also could help to avoid

340   unnecessary monitoring or interventions. In this context, tools like POSTRE provide an

341   objective framework to move toward mechanistically informed variant classification of

342   SVs.

343   The new pituitary-specific multi-omic dataset used in this study also uncovered important

344   details regarding the *VGLL1*-adjacent enhancers that we identified previously[12].

345   According to POSTRE enhancer calling criteria, the enhancer cluster identified within

346   *VGLL1* introns is predicted to be formed by four independent enhancers, given the

347   existence of four independent ATAC-seq peaks enriched in H3K27ac. Further research

348   will be needed to assess their independent contributions to *GPR101* expression and their

349   cooperation mode (e.g., additive, synergistic, hierarchical or redundant)[27,28]. In addition,

350   this cluster of *VGLL1* intronic enhancers is accompanied by a single distal enhancer

351   (Figure 4). This information supports the crucial importance of duplication of these

352   enhancers for the generation of the X-LAG phenotype. Since early studies on the genetics

353   of X-LAG, the *VGLL1* locus, along with *GPR101*, formed the smallest regions of overlap

354   (SROs) that were shared by all affected individuals[17,20] (Figure 1). Since then, we have

13

355 also identified and tested a series of putative enhancers including one at *RBMX*

356 (eRBMX)[12]. Using POSTRE, the current study indicates the centrality of *VGLL1*-adjacent

357 enhancer elements in the pathogenicity of duplications in X-LAG patients. Interestingly, a

358 partial duplication of a subset of enhancers in patient II led to an atypical form of X-LAG,

359 with severe excess GH hypersecretion in the absence of pituitary tumor formation. No

360 prolactin hypersecretion occurred, which is rare in X-LAG[10,17,29], and GHRH

361 hypersecretion, that is thought to be responsible for pituitary hyperproliferation in many

362 X-LAG cases[30], was absent. This case shares many features with the pituitary

363 somatotrope-specific transgenic mouse model that we described in Abboud et al.[31]. That

364 model had gigantism driven by pituitary GH and IGF1 excess, that was driven by an up

365 to 30-fold increase in pituitary *Gpr101* expression (in contrast, *GPR101* is increased

366 1000s-fold in the tumors of X-LAG patients)[12,32]. Importantly, the *Gpr101* transgenic

367 mouse had a normal pituitary morphology and no evidence of increased proliferation,

368 hyperplasia or tumorigenesis. Taken together, partial duplication of the intronic *VGLL1*

369 enhancer cluster could lead to an incomplete form of X-LAG, hypothetically due to

370 modest, somatotrope-specific elevations in *GPR101* expression. Definitive proof will

371 require further characterization of the enhancer sequences and their functional

372 interactions among themselves (i.e. enhancer cooperation mode), as well as their actions

373 on the *GPR101* promoter. This information would help to further refine the pathogenicity

374 classification of SVs in this region by tools like POSTRE. These observations

375 demonstrate that POSTRE, when equipped with cell-type specific epigenomic and

376 transcriptomic data, can effectively discriminate between pathogenic and benign SVs,

377 even among inter-TAD duplications. Moreover, they reinforce the functional significance

378 of the *VGLL1*-adjacent enhancers in X-LAG and argue against a pathogenic role for the

379 eRBMX CRE, despite its duplication and transcriptional enhancing activity shown in

380 HEK293 cells[12].

381 Despite its clear advantages, POSTRE does have limitations. The resolution of

382 predictions is inherently constrained by the quality and granularity of the enhancer and

383 TAD maps it relies upon. In the current version, TAD boundaries are derived from brain

384 prefrontal cortex data and even accounting for strong conservation of TAD boundaries

385 across tissues, may not fully reflect the three-dimensional chromatin architecture of

386   anterior pituitary cells. Future versions will benefit from ongoing efforts to collect and

387   integrate HiC or Micro-C data from pituitary tissues. Additionally, although enhancer

388   annotations were generated for this study using primary adult human pituitary epigenomic

389   datasets, new enhancer maps derived from genomic data at other developmental stages

390   (bulk/single-cell resolution), and improved artificial intelligence strategies to predict

391   enhancer-promoter interactions, could further enhance sensitivity and specificity.

392   Beyond X-LAG, this study illustrates POSTRE's broader potential in evaluating SVs

393   across TADopathies affecting the endocrine system, on top of its already proven

394   capabilities to handle this type of alterations in retinal, limb, craniofacial and

395   neurodevelopmental disorders[15,33,34]. Many such conditions are increasingly recognized

396   as the result of disrupted regulatory domains rather than coding mutations[35-38].

397   Incorporating tissue-relevant data into prediction algorithms allows for more nuanced

398   interpretation of SVs that may otherwise remain unclassified when using *in silico*

399   predictions tools that are agnostic to the cellular context[15,26]. As databases of chromatin

400   and enhancer landscapes expand, and as tools like POSTRE could eventually be

401   integrated into diagnostic pipelines, this will help the field move closer to precision

402   genomic medicine.

403   In conclusion, the findings presented here establish POSTRE as an effective,

404   interpretable, and scalable tool for the clinical interpretation of SVs at the *GPR101* locus.

405   Moreover, they reinforce the mechanistic model whereby duplications need to span both

406   the *GPR101* TAD boundary and include the *VGLL1*-adjacent enhancers to drive *GPR101*

407   misexpression in X-LAG.

408

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). 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). Written informed consent was obtained from all the subjects/guardians.

Availability of data and materials

The datasets generated during this study are available on request.

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.

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.

Acknowledgements

17
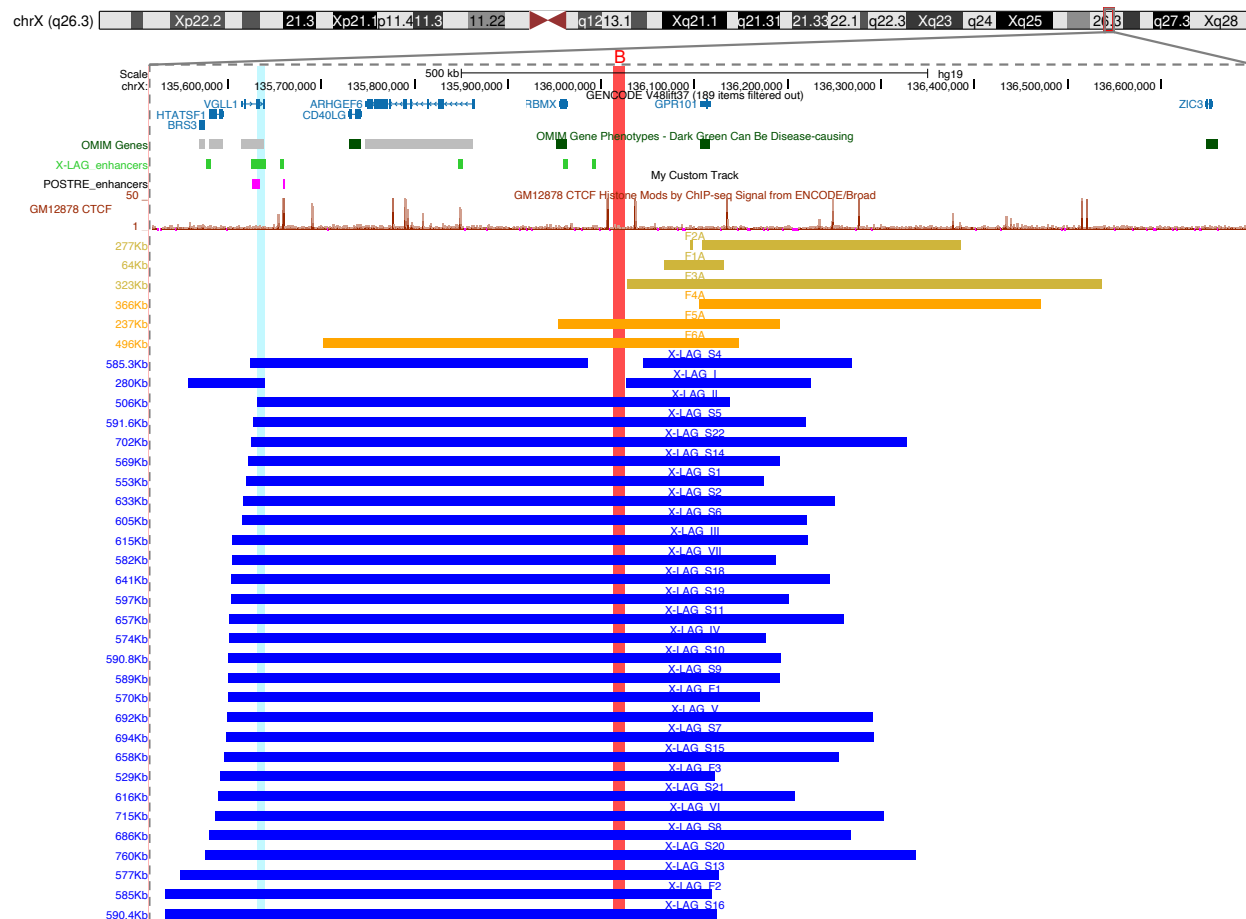
Figure 1. Overview of duplications at the Xq26.3 locus involving *GPR101*, classified by pathogenicity.

The genomic region surrounding *GPR101* (chrX:135,500,000-136,700,000, hg19) is shown with annotated protein-coding and OMIM genes (blue and dark green, respectively), CTCF binding sites (orange ChIP-seq track from GM12878 cells), and putative CREs (six we previously described[12] based on publicly available data[39,40] given as light green bars and a subset predicted by POSTRE located within or distal to *VGLL1* given as purple bars). The centromeric *GPR101*-TAD boundary is marked by a red vertical bar. Colored bars below represent the extent of individual duplications: yellow for three non-pathogenic duplications reported by Daly et al.[14] (F1A-F3A), orange for the three newly identified non-pathogenic duplications presented in this study (F4A-F6A), and blue for X-LAG-associated pathogenic duplications, including both continuous and discontinuous (patients S4 and I) rearrangements. Only pathogenic duplications span the *VGLL1* intronic enhancer cluster. The light blue vertical bar highlights the smallest region

18

470   of overlap of all duplications partially encompassing that CRE. All non-pathogenic

471   duplications, despite partial TAD disruption or inclusion of other CREs such as the *RBMX*

472   enhancer[12], were predicted by POSTRE to be neutral.

473

A — Introduce Structural Variant and Phenotype

B

C — Results Overview endocrine

This structural variant is predicted to be pathogenic in relation with endocrine abnormalities. You can find more information below in the predictions table.

Predictions table legend:

- **NOT Pathogenic.** If a grey rectangle with '-' appears, this indicates that the pathogenic score computed for the associated gene (row) and cell type (column) is below the pathogenic threshold (0.8). Thus, considered as NOT pathogenic.
- **LOF - Pathogenic.** If a red rectangle with 'LOF' appears, this indicates that a pathogenic Loss Of Function has been predicted for the associated gene (row) and cell type (column). Click on the rectangle to know more.
- **GOF - Pathogenic.** If a green rectangle with 'GOF' appears, this indicates that a pathogenic Gain Of Function has been predicted for the associated gene (row) and cell type (column). Click on the rectangle to know more.

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
HTATSF1	-	-	Long-Range
FHL1	-	-	Long-Range
SLC9A6	-	-	Long-Range
ZIC3	-	-	Long-Range
VGLL1	-	-	Gene Truncation
MMGT1	-	-	Long-Range
BRS3	-	-	Long-Range
MAP7D3	-	-	Long-Range
ADGRG4	-	-	Long-Range

D — Graphical summary of GPR101 regulatory domain in NormalPituitary Simplification
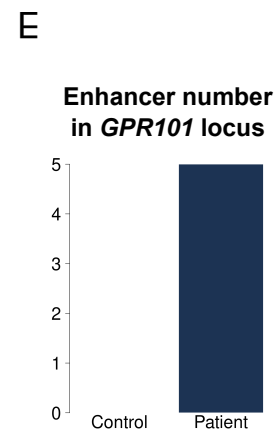
E — Enhancer number in *GPR101* locus

474

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

Panels A-E show part of the results obtained with POSTRE for the analysis of the pathogenic duplication S22. A) Screenshot of POSTRE submission menu for the Single SV analysis mode with the X-LAG case S22 SV coordinates introduced. The duplication was evaluated using POSTRE in Standard mode with the "Endocrine" phenotype selected. This phenotype selection triggers the consideration of anterior pituitary-specific enhancer and expression data. B) Visualization of S22 duplication at 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 red. The session also depicts bigwigs of H3K27ac ChIP-seq and ATAC-seq data from anterior pituitary, TAD maps from the brain prefrontal cortex, and highlights POSTRE predicted active enhancers with green vertical lines. C) Main POSTRE prediction summary table. D) Graphical representation of the SV impact with respect to the *GPR101* regulatory domain. In the control scenario, the existence of eight enhancers can be observed at the TAD containing the *VGLL1*-adjacent enhancers. These eight enhancers include the five *VGLL1*-adjacent enhancers (Figure 4) and three additional ones located at the centromeric side of that TAD (according to the prefrontal cortex TAD map used as reference). However, these latter three enhancers (located >250Kb away from the closest duplication breakpoint) were excluded from downstream considerations as they are never duplicated in any of the reported X-LAG duplications (Figure 1). E) Barplot highlighting the changes in the enhancer number surrounding *GPR101* in the control versus patient alleles. Overall, for S22, POSTRE predicts enhancer adoption and neo-TAD formation involving the *VGLL1*-adjacent enhancers, resulting in a high pathogenicity score and a likely mechanism of *GPR101* misexpression.

Figure 3. POSTRE output for a non-pathogenic (F6A) duplication at the *GPR101* locus.

A) Screenshot of POSTRE submission menu for the Single SV analysis mode with the F6A SV coordinates introduced. The duplication was evaluated using POSTRE in Standard mode with the "Endocrine" phenotype selected. This phenotype selection triggers the consideration of anterior pituitary-specific enhancer and expression data. B) Main POSTRE prediction summary table. It predicts the SV as non-pathogenic.

A

B

509 .

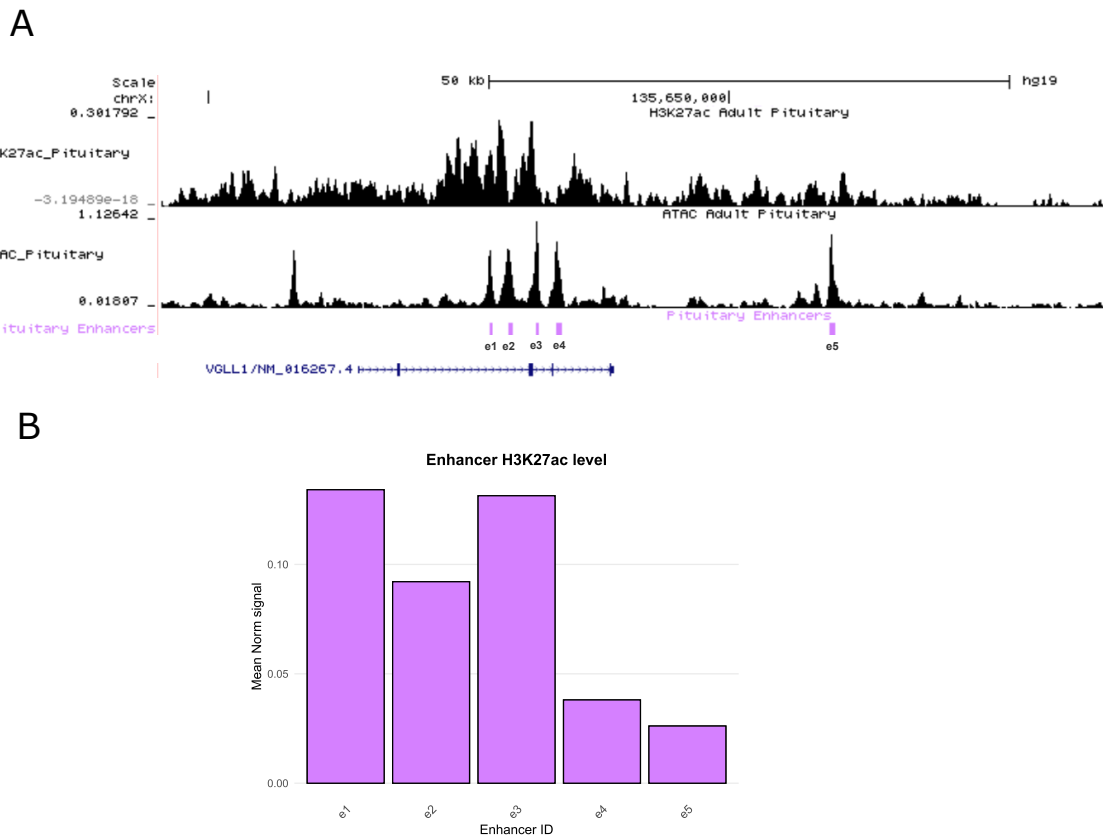510 **Figure 4. *VGLL1*-adjacent enhancers overview.**

511 A) Overview of the *VGLL1*-adjacent enhancers. According to the POSTRE enhancer

512 calling pipeline, five enhancers, that we named here e1-e5, are predicted. B) Barplot of

513 the mean H3K27ac levels (calculated from H3K27ac ChIP-Seq bigwigs) at e1-e5

514 enhancers. Higher H3K27ac levels are indicators of stronger enhancer activity[41].

SV ID	Phenotype	Pathogenic Score	Pathogenic (Yes/No)	Causative genes	Candidate genes (Pathogenic Score)
F1A	endocrine	0.5	No		GPR101(0.5),ZIC3(0.25)
F2A	endocrine	0.5	No		GPR101(0.5),ZIC3(0.25)
F3A	endocrine	0.5	No		GPR101(0.5),ARHGEF6(0.38),HTATSF1(0.37),RBMX(0.37),FHL1(0.37),SLC9A6(0.33),CD40LG(0.25),ZIC3(0.25),MMGT1(0.25),VGLL1(0.19),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
F4A	endocrine	0.5	No		GPR101(0.5),ZIC3(0.25)
F5A	endocrine	0.5	No		GPR101(0.5),RBMX(0.49),ARHGEF6(0.38),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),CD40LG(0.25),ZIC3(0.25),MMGT1(0.25),VGLL1(0.19),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
F6A	endocrine	0.5	No		GPR101(0.5),ARHGEF6(0.5),RBMX(0.49),HTATSF1(0.37),FHL1(0.37),CD40LG(0.33),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),VGLL1(0.19),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
F1	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
F2	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),HTATSF1(0.62),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),BRS3(0.43),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),MAP7D3(0.16),ADGRG4(0)
F3	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),HTATSF1(0.5),VGLL1(0.44),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
S1	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),VGLL1(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
S2	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),VGLL1(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
S5	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),VGLL1(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
S6	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),VGLL1(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
S7	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
S8	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),HTATSF1(0.5),VGLL1(0.44),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
S9	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
S10	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
S11	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
S13	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),HTATSF1(0.62),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),BRS3(0.43),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),MAP7D3(0.16),ADGRG4(0)
S14	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),VGLL1(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
S15	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
S16	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),HTATSF1(0.62),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),BRS3(0.43),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),MAP7D3(0.16),ADGRG4(0)
S18	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
S19	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
S20	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),HTATSF1(0.62),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.23),MAP7D3(0.16),ADGRG4(0)
S21	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),HTATSF1(0.5),VGLL1(0.44),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)

SV ID	Phenotype	Pathogenic Score	Pathogenic (Yes/No)	Causative genes	Candidate genes (Pathogenic Score)
S22	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),VGLL1(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
II	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),VGLL1(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
III	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
IV	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
V	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
VI	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),HTATSF1(0.5),VGLL1(0.44),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)
VII	endocrine	0.83	Yes	GPR101	GPR101(0.83),ARHGEF6(0.63),RBMX(0.62),CD40LG(0.5),VGLL1(0.44),HTATSF1(0.37),FHL1(0.37),SLC9A6(0.33),ZIC3(0.25),MMGT1(0.25),BRS3(0.18),MAP7D3(0.16),ADGRG4(0)

Table 1. POSTRE output for the Multiple SV analysis. Note: pathogenicity scores for the F3A duplication were also computed by POSTRE for the genes located in the predicted TAD centromeric to *GPR101*. Although the duplication is reported to lie entirely within the *GPR101*-TAD (Figure 1), the TAD map used by POSTRE considers that it encroaches on this centromeric TAD but does not overlap the enhancers located within. This discrepancy reflects a limitation of TAD-calling softwares, as precise annotation of TAD boundaries is often lacking.

References

520

521

522   1.   Dixon JR, Selvaraj S, Yue F, et al. Topological domains in mammalian genomes
523   identified by analysis of chromatin interactions. *Nature*. Apr 11 2012;485(7398):376-80.
524   doi:10.1038/nature11082
525   2.   Nora EP, Dekker J, Heard E. Segmental folding of chromosomes: a basis for
526   structural and regulatory chromosomal neighborhoods? *Bioessays*. Sep 2013;35(9):818-
527   28. doi:10.1002/bies.201300040
528   3.   Lupianez DG, Spielmann M, Mundlos S. Breaking TADs: How Alterations of
529   Chromatin Domains Result in Disease. *Trends Genet*. Apr 2016;32(4):225-237.
530   doi:10.1016/j.tig.2016.01.003
531   4.   Weischenfeldt J, Ibrahim DM. When 3D genome changes cause disease: the
532   impact of structural variations in congenital disease and cancer. *Curr Opin Genet Dev*.
533   Jun 2023;80:102048. doi:10.1016/j.gde.2023.102048
534   5.   Spielmann M, Lupianez DG, Mundlos S. Structural variation in the 3D genome.
535   *Nat Rev Genet*. Jul 2018;19(7):453-467. doi:10.1038/s41576-018-0007-0
536   6.   D'Haene E, Vergult S. Interpreting the impact of noncoding structural variation in
537   neurodevelopmental disorders. *Genet Med*. Jan 2021;23(1):34-46. doi:10.1038/s41436-
538   020-00974-1
539   7.   Valton AL, Dekker J. TAD disruption as oncogenic driver. *Curr Opin Genet Dev*.
540   Feb 2016;36:34-40. doi:10.1016/j.gde.2016.03.008
541   8.   Matharu N, Ahituv N. Minor Loops in Major Folds: Enhancer-Promoter Looping,
542   Chromatin Restructuring, and Their Association with Transcriptional Regulation and
543   Disease. *PLoS Genet*. Dec 2015;11(12):e1005640. doi:10.1371/journal.pgen.1005640
544   9.   Rajderkar S, Barozzi I, Zhu Y, et al. Topologically associating domain boundaries
545   are required for normal genome function. *Commun Biol*. Apr 20 2023;6(1):435.
546   doi:10.1038/s42003-023-04819-w
547   10.   Trivellin G, Daly AF, Faucz FR, et al. Gigantism and acromegaly due to Xq26
548   microduplications and GPR101 mutation. *N Engl J Med*. Dec 18 2014;371(25):2363-74.
549   doi:10.1056/NEJMoa1408028
550   11.   Caruso M, Mazzatenta D, Asioli S, et al. Case report: Management of pediatric
551   gigantism caused by the TADopathy, X-linked acrogigantism. *Front Endocrinol*
552   *(Lausanne)*. 2024;15:1345363. doi:10.3389/fendo.2024.1345363
553   12.   Franke M, Daly AF, Palmeira L, et al. Duplications disrupt chromatin architecture
554   and rewire GPR101-enhancer communication in X-linked acrogigantism. *Am J Hum*
555   *Genet*. Apr 7 2022;109(4):553-570. doi:10.1016/j.ajhg.2022.02.002
556   13.   Dimartino P, Zadorozhna M, Yumiceba V, et al. Structural Variants at the LMNB1
557   Locus: Deciphering   Pathomechanisms   in   Autosomal   Dominant   Adult-Onset
558   Demyelinating   Leukodystrophy.   *Ann   Neurol*.   Nov   2024;96(5):855-870.
559   doi:10.1002/ana.27038
560   14.   Daly AF, Dunnington LA, Rodriguez-Buritica DF, et al. Chromatin conformation
561   capture in the clinic: 4C-seq/HiC distinguishes pathogenic from neutral duplications at the
562   GPR101 locus. *Genome Med*. Sep 13 2024;16(1):112. doi:10.1186/s13073-024-01378-5
563   15.   Sanchez-Gaya V, Rada-Iglesias A. POSTRE: a tool to predict the pathological
564   effects of human structural variants. *Nucleic Acids Res*. May 22 2023;51(9):e54.
565   doi:10.1093/nar/gkad225

566 16.   Beckers A, Lodish MB, Trivellin G, et al. X-linked acrogigantism syndrome: clinical
567 profile and therapeutic responses. *Endocr Relat Cancer*. Jun 2015;22(3):353-67.
568 doi:10.1530/ERC-15-0038
569 17.   Iacovazzo D, Caswell R, Bunce B, et al. Germline or somatic GPR101 duplication
570 leads to X-linked acrogigantism: a clinico-pathological and genetic study. *Acta
571 Neuropathol Commun*. Jun 1 2016;4(1):56. doi:10.1186/s40478-016-0328-1
572 18.   Naves LA, Daly AF, Dias LA, et al. Aggressive tumor growth and clinical evolution
573 in a patient with X-linked acro-gigantism syndrome. *Endocrine*. Feb 2016;51(2):236-44.
574 doi:10.1007/s12020-015-0804-6
575 19.   Trarbach EB, Trivellin G, Grande IPP, et al. Genetics, clinical features and
576 outcomes of non-syndromic pituitary gigantism: experience of a single center from Sao
577 Paulo, Brazil. *Pituitary*. Apr 2021;24(2):252-261. doi:10.1007/s11102-020-01105-4
578 20.   Trivellin G, Hernandez-Ramirez LC, Swan J, Stratakis CA. An orphan G-protein-
579 coupled receptor causes human gigantism and/or acromegaly: Molecular biology and
580 clinical correlations. *Best Pract Res Clin Endocrinol Metab*. Apr 2018;32(2):125-140.
581 doi:10.1016/j.beem.2018.02.004
582 21.   Pasinska M, Lazarczyk E, Repczynska A, et al. Clinical Importance of aCGH in
583 Genetic Counselling of Children with Psychomotor Retardation. *Appl Clin Genet*.
584 2022;15:27-38. doi:10.2147/TACG.S357136
585 22.   Schmitt AD, Hu M, Jung I, et al. A Compendium of Chromatin Contact Maps
586 Reveals Spatially Active Regions in the Human Genome. *Cell Rep*. Nov 15
587 2016;17(8):2042-2059. doi:10.1016/j.celrep.2016.10.061
588 23.   Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org:
589 Online Mendelian Inheritance in Man (OMIM(R)), an online catalog of human genes and
590 genetic disorders. *Nucleic Acids Res*. Jan 2015;43(Database issue):D789-98.
591 doi:10.1093/nar/gku1205
592 24.   Bult CJ, Blake JA, Smith CL, Kadin JA, Richardson JE, Mouse Genome Database
593 G. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res*. Jan 8 2019;47(D1):D801-
594 D806. doi:10.1093/nar/gky1056
595 25.   Burren CP, Williams G, Coxson E, Korbonits M. Effective Long-term Pediatric
596 Pegvisomant Monotherapy to Final Height in X-linked Acrogigantism. *JCEM Case Rep*.
597 May 2023;1(3):luad028. doi:10.1210/jcemcr/luad028
598 26.   Sanchez-Gaya V, Mariner-Fauli M, Rada-Iglesias A. Rare or Overlooked?
599 Structural Disruption of Regulatory Domains in Human Neurocristopathies. *Front Genet*.
600 2020;11:688. doi:10.3389/fgene.2020.00688
601 27.   Blayney JW, Francis H, Rampasekova A, et al. Super-enhancers include classical
602 enhancers and facilitators to fully activate gene expression. *Cell*. Dec 21
603 2023;186(26):5826-5839 e18. doi:10.1016/j.cell.2023.11.030
604 28.   Peng Y, Zhang Y. Enhancer and super-enhancer: Positive regulators in gene
605 transcription. *Animal Model Exp Med*. Sep 2018;1(3):169-179. doi:10.1002/ame2.12032
606 29.   Daly AF, Beckers A. The Genetic Pathophysiology and Clinical Management of the
607 TADopathy, X-Linked Acrogigantism. *Endocr Rev*. Sep 12 2024;45(5):737-754.
608 doi:10.1210/endrev/bnae014
609 30.   Daly AF, Lysy PA, Desfilles C, et al. GHRH excess and blockade in X-LAG
610 syndrome. *Endocr Relat Cancer*. Mar 2016;23(3):161-70. doi:10.1530/ERC-15-0478

611   31.   Abboud D, Daly AF, Dupuis N, et al. GPR101 drives growth hormone
612   hypersecretion and gigantism in mice via constitutive activation of Gs and Gq/11. *Nat*
613   *Commun*. Sep 21 2020;11(1):4752. doi:10.1038/s41467-020-18500-x

614   32.   Trivellin G, Bjelobaba I, Daly AF, et al. Characterization of GPR101 transcript
615   structure and expression patterns. *J Mol Endocrinol*. Aug 2016;57(2):97-111.
616   doi:10.1530/JME-16-0045

617   33.   Plaisancie J, Chesneau B, Fares-Taie L, et al. Structural Variant Disrupting the
618   Expression of the Remote FOXC1 Gene in a Patient with Syndromic Complex
619   Microphthalmia. *Int J Mol Sci*. Feb 25 2024;25(5)doi:10.3390/ijms25052669

620   34.   Hamerlinck L, D'haene E, Van Loon N, et al. Non-coding structural variants identify
621   a commonly affected regulatory region steering FOXG1 transcription in early
622   neurodevelopment.   *medRxiv*.   2025:2025.03.10.25323301.
623   doi:10.1101/2025.03.10.25323301

624   35.   Carballo-Pacoret P, Carracedo A, Rodriguez-Fontenla C. Unraveling the three-
625   dimensional (3D) genome architecture in Neurodevelopmental Disorders (NDDs).
626   *Neurogenetics*. Oct 2024;25(4):293-305. doi:10.1007/s10048-024-00774-8

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

631   37.   Lima AC, Okhovat M, Stendahl AM, et al. Deletion of an evolutionarily conserved
632   TAD boundary impacts spermatogenesis in mice. *Biol Reprod*. Apr 13 2025;112(4):767-
633   779. doi:10.1093/biolre/ioaf017

634   38.   de Bruijn SE, Fiorentino A, Ottaviani D, et al. Structural Variants Create New
635   Topological-Associated Domains and Ectopic Retinal Enhancer-Gene Contact in
636   Dominant Retinitis Pigmentosa. *Am J Hum Genet*. Nov 5 2020;107(5):802-814.
637   doi:10.1016/j.ajhg.2020.09.002

638   39.   Mayran A, Khetchoumian K, Hariri F, et al. Pioneer factor Pax7 deploys a stable
639   enhancer repertoire for specification of cell fate. *Nat Genet*. Feb 2018;50(2):259-269.
640   doi:10.1038/s41588-017-0035-2

641   40.   Vermunt MW, Reinink P, Korving J, et al. Large-scale identification of coregulated
642   enhancer networks in the adult human brain. *Cell Rep*. Oct 23 2014;9(2):767-79.
643   doi:10.1016/j.celrep.2014.09.023

644   41.   Fulco CP, Nasser J, Jones TR, et al. Activity-by-contact model of enhancer-
645   promoter regulation from thousands of CRISPR perturbations. *Nat Genet*. Dec
646   2019;51(12):1664-1669. doi:10.1038/s41588-019-0538-0
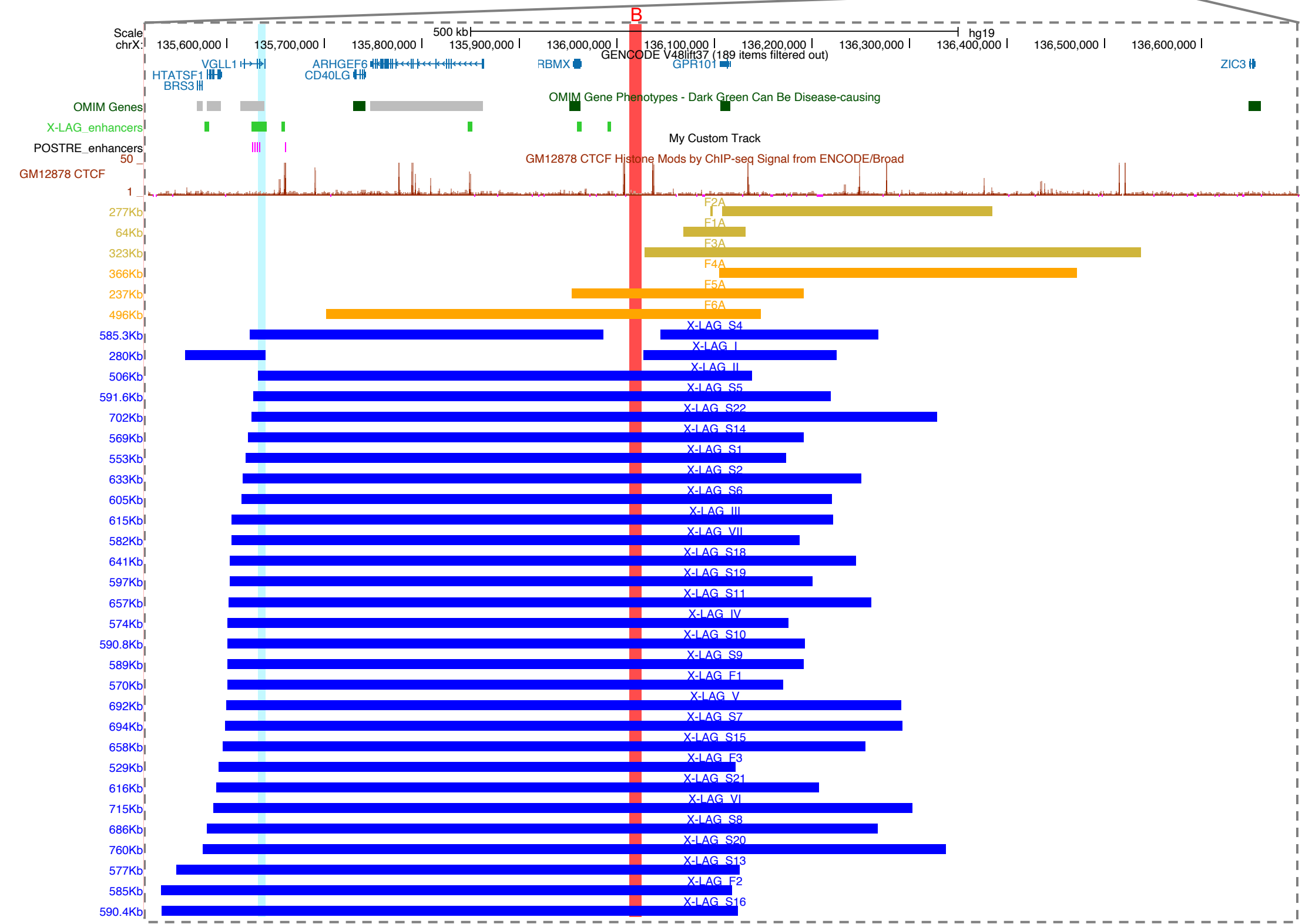
647

A Introduce Structural Variant and Phenotype

Type SV

Duplication

Breakpoint 1 chromosome   Breakpoint 2 chromosome

chrX   chrX

Breakpoint 1 coordinates in hg19   Breakpoint 2 coordinates in hg19
(coord1,coord2 if interval)   (coord1,coord2 if interval)

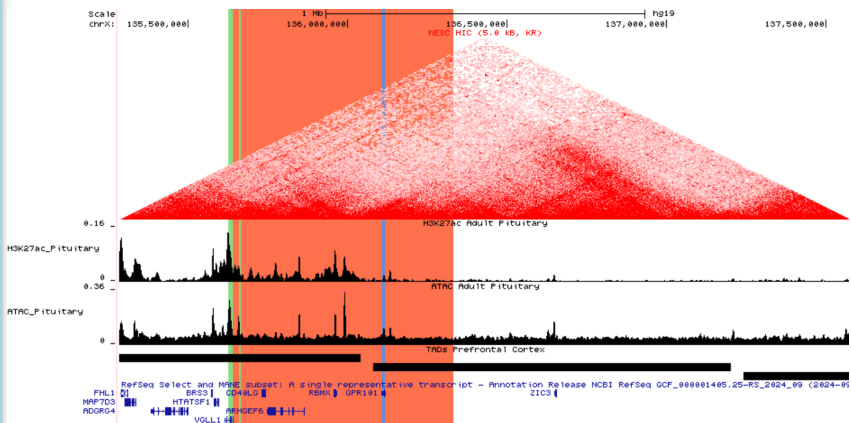135625218   136327754

Phenotype

☐ Cardiovascular ☐ Head & Neck ☐ Limbs ☐ Neurodevelopmental ☐ Vision-Eye
☐ Liver - Biliary System ☑ Endocrine

Advanced Parameters [+]

NOTE: Reference Genome Coordinates required in GRCh37/hg19
Please, visit any of the following websites to convert your coordinates to hg19 if you need it: UCSC LiftOver, ENSEMBL Assembly Converter, Broad Institute LiftOver

Submit

B

C
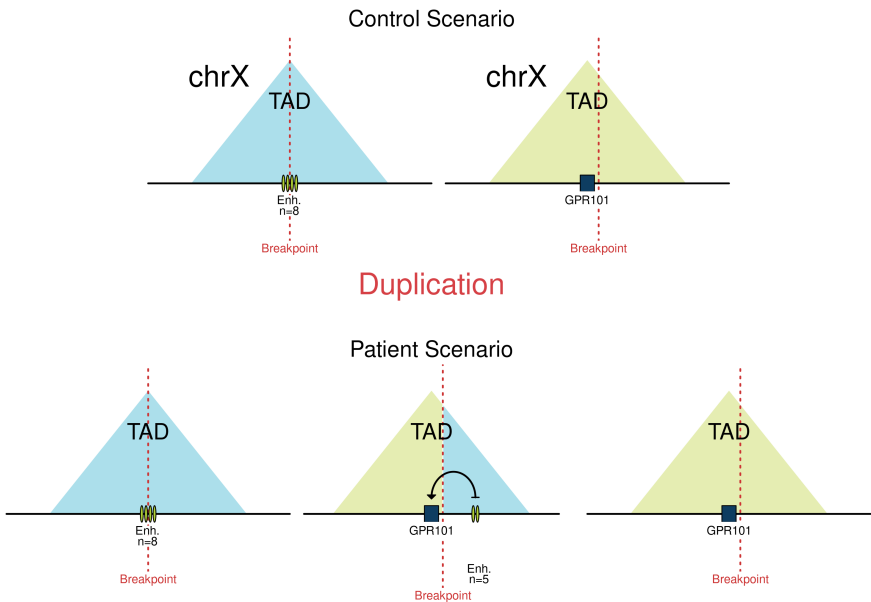
Results Overview endocrine

This structural variant is predicted to be **pathogenic** in relation with endocrine abnormalities. You can find more information below in the predictions table.

Predictions table legend:

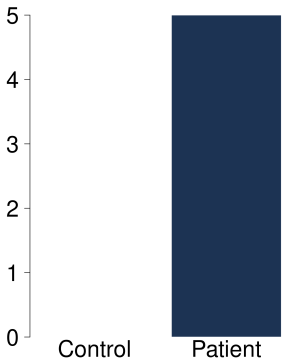|	
-	**NOT Pathogenic.** If a grey rectangle with '-' appears, this indicates that the pathogenic score computed for the associated gene (row) and cell type (column) is below the pathogenic threshold (0.8). Thus, considered as NOT pathogenic.
LOF	**LOF - Pathogenic.** If a red rectangle with 'LOF' appears, this indicates that a pathogenic Loss Of Function has been predicted for the associated gene (row) and cell type (column). Click on the rectangle to know more.
GOF	**GOF - Pathogenic.** If a green rectangle with 'GOF' appears, this indicates that a pathogenic Gain Of Function has been predicted for the associated gene (row) and cell type (column). Click on the rectangle to know more.

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
HTATSF1	-	-	Long-Range
FHL1	-	-	Long-Range
SLC9A6	-	-	Long-Range
ZIC3	-	-	Long-Range
VGLL1	-	-	Gene Truncation
MMGT1	-	-	Long-Range
BRS3	-	-	Long-Range
MAP7D3	-	-	Long-Range
ADGRG4	-	-	Long-Range

D Graphical summary of GPR101 regulatory domain in NormalPituitary Simplification

E

Enhancer number in *GPR101* locus

A

Introduce Structural Variant and Phenotype

Type SV

Duplication ▾

Breakpoint 1 chromosome

chrX ▾

Breakpoint 1 coordinates in hg19 (coord1,coord2 if interval)

135702382

Breakpoint 2 chromosome

chrX ▾

Breakpoint 2 coordinates in hg19 (coord1,coord2 if interval)

136147166

Phenotype

☐ Cardiovascular ☐ Head & Neck ☐ Limbs ☐ Neurodevelopmental ☐ Vision-Eye
☐ Liver - Biliary System ☑ Endocrine

Advanced Parameters [+]

NOTE: Reference Genome Coordinates required in GRCh37/hg19

Please, visit any of the following websites to convert your coordinates to hg19 if you need it: UCSC LiftOver, ENSEMBL Assembly Converter, Broad Institute LiftOver

⊿ Submit

B

Results Overview endocrine

This structural variant is NOT predicted to be pathogenic in relation with endocrine abnormalities.

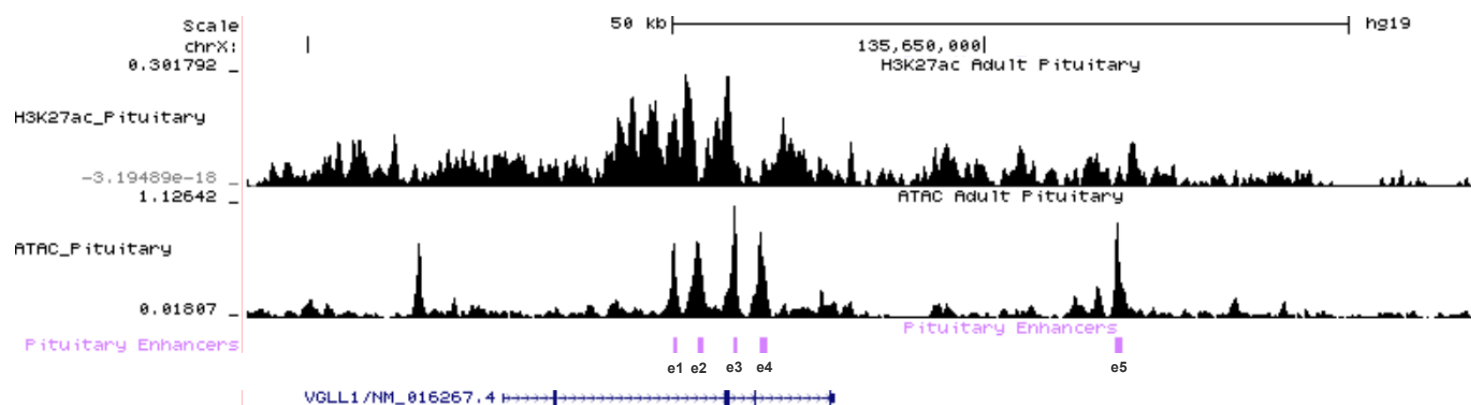**Predictions table legend:**

-	**NOT Pathogenic.** If a grey rectangle with '-' appears, this indicates that the pathogenic score computed for the associated gene (row) and cell type (column) is below the pathogenic threshold (0.8). Thus, considered as NOT pathogenic.
LOF	**LOF - Pathogenic.** If a red rectangle with 'LOF' appears, this indicates that a pathogenic Loss Of Function has been predicted for the associated gene (row) and cell type (column). Click on the rectangle to know more.
GOF	**GOF - Pathogenic.** If a green rectangle with 'GOF' appears, this indicates that a pathogenic Gain Of Function has been predicted for the associated gene (row) and cell type (column). Click on the rectangle to know more.

Gene	NormalPituitary	NormalPituitary2	Pathomechanism
GPR101	-	-	Gene Duplication
ARHGEF6	-	-	Gene Duplication
RBMX	-	-	Gene Duplication
HTATSF1	-	-	Long-Range
FHL1	-	-	Long-Range
CD40LG	-	-	Gene Duplication
SLC9A6	-	-	Long-Range
ZIC3	-	-	Long-Range
MMGT1	-	-	Long-Range
VGLL1	-	-	Long-Range
BRS3	-	-	Long-Range
MAP7D3	-	-	Long-Range
ADGRG4	-	-	Long-Range

A

B

Enhancer H3K27ac level

Mean Norm signal

Enhancer ID