1   **Characterization of *GPR101* transcript structure and expression patterns**

2

3   Giampaolo Trivellin[1], Ivana Bjelobaba[2], Adrian F. Daly[3], Darwin O. Larco[4], Leonor Palmeira[3], Fabio R.

4   Faucz[1], Albert Thiry[5], Letícia F. Leal[1,6], Liliya Rostomyan[3], Martha Quezado[7], Marie Helene

5   Schernthaner-Reiter[1], Marija M. Janjic[2], Chiara Villa[3,8], T. John Wu[4], Stanko S. Stojilkovic[2], Albert

6   Beckers[3], Benjamin Feldman[9], Constantine A. Stratakis[1]

7

8   [1]Section on Endocrinology and Genetics, *Eunice Kennedy Shriver* National Institute of Child Health and

9   Human Development (NICHD), National Institutes of Health (NIH), Bethesda, MD, 20892, USA.

10   [2]Section on Cellular Signaling, *Eunice Kennedy Shriver* National Institute of Child Health and Human

11   Development (NICHD), National Institutes of Health (NIH), Bethesda, MD, 20892, USA.

12   [3]Department of Endocrinology, Centre Hospitalier Universitaire de Liège, University of Liège, Domaine

13   Universitaire du Sart-Tilman, 4000 Liège, Belgium.

14   [4]Department of Obstetrics and Gynecology, Uniformed Services University of the Health Sciences,

15   Bethesda, MD 20814, USA.

16   [5]Department of Pathology, Centre Hospitalier Universitaire de Liège, University of Liège, Domaine

17   Universitaire du Sart-Tilman, 4000 Liège, Belgium.

18   [6]Department of Pediatrics, Ribeirao Preto Medical School, University of Sao Paulo, Ribeirao Preto, SP,

19   Brazil.

20   [7]Laboratory of Pathology, National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda,

21   MD, 20892, USA.

22   [8]Hopital Foch, Service d'Anatomie et Cytologie Pathologiques, 40, rue Worth - B.P 36 - 92151 Suresnes

23   Cedex, France.

24   [9]Division of Developmental Biology, *Eunice Kennedy Shriver* National Institute of Child Health and

25   Human Development (NICHD), National Institutes of Health (NIH), Bethesda, MD, 20892, USA.

26

1

27   **Manuscript type:** Original article

28   **Abbreviated title:** *GPR101* expression

29   **Key terms:** GPR101, pituitary, hypothalamus, expression analysis, transcripts structure

30   **Main Text word count:** 5328

31   **Abstract word count:** 250

32   **Figure number:** 8

33   **References:** 32

34

35

36

37   **Correspondence should be addressed to:**

38   Dr. Constantine A. Stratakis, M.D., D(Med)Sc.

39   SEGEN, NICHD, NIH

40   10 Center Drive, Building 10, NIH-Clinical Research Center, Room 1-3330, MSC1103

41   Bethesda, MD, 20892-1862, USA

42   Tel: 001-301-496-4686; 001-301-402-0574

43   E-mail: stratakc@mail.nih.gov

44

45   **Disclosure Statement:** The authors have nothing to disclose.

46

47 **Abstract**

48

49 We recently showed that Xq26.3 microduplications cause X-linked acrogigantism (X-LAG). X-LAG

50 patients mainly present with growth hormone and prolactin-secreting adenomas and share a minimal

51 duplicated region containing at least four genes. *GPR101* was the only gene highly expressed in their

52 pituitary lesions, but little is known about its expression patterns. *GPR101* transcripts were characterized

53 in human tissues by 5'-RACE and RNAseq, while the putative promoter was bioinformatically predicted.

54 We investigated *GPR101* mRNA and protein expression by RT-qPCR, whole-mount *in situ* hybridization,

55 and immunostaining, in human, rhesus monkey, rat, and zebrafish. We identified four *GPR101* isoforms

56 characterized by different 5' untranslated regions (UTRs) and a common 6.1 kb-long 3'UTR. *GPR101*

57 expression was very low or absent in almost all adult human tissues examined, except for specific brain

58 regions. Strong GPR101 staining was observed in human fetal pituitary and during adolescence, whereas

59 very weak/absent expression was detected during childhood and adult life. In contrast to humans, adult

60 pituitaries of monkey and rat expressed GPR101, but in different cell types. *Gpr101* is expressed in the

61 brain and pituitary during rat and zebrafish development; in rat pituitary *Gpr101* is expressed only after

62 birth and showed sexual dimorphism. This study shows that different *GPR101* transcripts exist and that

63 the brain is the major site of *GPR101* expression across different species, although divergent species- and

64 temporal-specific expression patterns are evident. These findings suggest an important role for GPR101 in

65 brain and pituitary development and likely reflect the very different growth, development and maturation

66 patterns among species.

67

68   **Introduction**

69

70   X-linked acrogigantism (X-LAG) is a recently discovered syndrome characterized by early childhood-

71   onset gigantism due to Xq26.3 microduplications (Trivellin, et al. 2014). We have shown that

72   microduplications in X-LAG patients can occur constitutionally or as somatic mosaicism (Daly, et al.

73   2016b). Patients with X-LAG present with mixed growth hormone (GH) and prolactin (PRL)-secreting

74   pituitary macroadenomas and/or hyperplasia before five years of age (Beckers, et al. 2015). We have

75   previously shown that the smallest common duplicated region shared by all patients is about 500 kb long

76   and contains four OMIM genes, of which only an orphan G protein-coupled receptor (GPCR) named

77   *GPR101* (OMIM# 300393) is highly expressed in pituitary tumor tissue. Functional studies showed that

78   *GPR101* duplication is the likely cause of X-LAG (Trivellin et al. 2014).

79

80   GPR101 belongs to the class A Rhodopsin superfamily of GPCRs (Kakarala and Jamil 2014).

81   Phylogenetic analysis showed that it is most similar to another orphan GPCR named GPR161

82   (Mukhopadhyay, et al. 2013). Both GPCRs are closely related to the adrenergic and serotonin receptors,

83   but share only about 30% sequence identity in their transmembrane regions (Lee, et al. 2001). So far very

84   little is known regarding which tissues express *GPR101*, especially in humans. We and others have

85   previously shown that *Gpr101* is highly expressed in the hypothalamus in rodents, where it seems to play

86   a role in hypothalamic control of energy homeostasis (Bates, et al. 2006; Lee et al. 2001; Nilaweera, et al.

87   2007; Nilaweera, et al. 2008; Regard, et al. 2008; Trivellin et al. 2014). Interestingly, an increase in

88   *Gpr101* mRNA expression in the rat hypothalamus was seen after puberty and from late pregnancy to late

89   lactation suggesting that the receptor may also play a role in physiological processes associated with

90   pregnancy, lactation, and reproduction (Nilaweera et al. 2008) (E Beadle, D O Larco, A Vana, M Cho-

91   Clark, T J Wu, personal communication). On the contrary, few cells seem to express GPR101 in the adult

92   human hypothalamus. Normal pituitary and GH-secreting tumors without the Xq26.3 duplication also

93   show relatively scarce or negligible *GPR101* expression (Trivellin et al. 2014).

94

95   The complete structure of the human *GPR101* gene is unknown as only the coding sequence (CDS) has

96   been characterized to date (NM_054021.1 / ENST00000298110.1). The putative *GPR101* transcript is

97   1527 base-pairs (bp) long and consists of one protein-coding exon. The structure of the *Gpr101* gene has

98   been better characterized in mouse (NM_001033360.3) and rat (NM_001108258.1) and in both species it

99   consists of two exons, of which only exon 2 is protein-coding. In zebrafish a *gpr101* orthologue has been

100 predicted and annotated (ENSDARG00000039218). The human *GPR101* gene shares 80%, 79%, and

101 55%   similarity   with   the   mouse,   rat,   and   zebrafish   orthologues,   respectively

102 (http://www.ncbi.nlm.nih.gov/homologene/).

103

104 We undertook the present study to better characterize GPR101 expression across different species, and to

105 define the structure of the human gene.

106

107   **Materials and Methods**

108

109   *In silico analyses*

110

111   Alignment of the mouse Gpr101 mRNA (NM_001033360.3) to the human genome (GRCh37/h19

112   assembly) and of the human putative *GPR101* promoter region to the mouse genome (GRCm38/mm10)

113   was performed with the BLAST tool (http://blast.ncbi.nlm.nih.gov/Blast.cgi) using the Discontiguous

114   megablast algorithm. Analysis of the human putative *GPR101* promoter region was also performed with

115   other three software packages: MPromDb (http://mpromdb.wistar.upenn.edu/search), the UCSC genome

116   browser track annotating transcription factors (TFs) binding sites assayed by ChIP-seq [TF ChIP-seq (161

117   factors) from ENCODE with Factorbook Motifs], and the MatInspector tool from the Genomatix

118   Software Suite (http://www.genomatix.de/). Genomatix results were filtered by vertebrate elements,

119   location (only sites located within or spanning the conserved human-mouse regions were considered), a

120   matrix similarity threshold of 0.85, and strand (only for TATA-box elements, since true positives are

121   considered those in the same orientation of the gene).

122

123   *RNAseq and Rapid Amplification of cDNA Ends (RACE)*

124

125   Total RNA was extracted from two female patients with X-LAG (both from lymphocytes and pituitary

126   tumors) (Beckers et al. 2015; Trivellin et al. 2014), and from HEK293 cells, as previously described

127   (Trivellin et al. 2014). RNAseq analysis was performed by Zymo Research in pituitary tumor RNA

128   samples from both X-LAG patients. Briefly, HiSeq 50 x 2 bp paired-end reads from RNA-Seq samples

129   first went through quality control steps by trimming adaptors and low quality bases (q score < 20) and

130   were then analyzed using the Cufflinks software (v2.2.0) for transcript assembly and differential

131   expression. Default parameters were used unless otherwise specified. BAM files were visualized with

132   IGV v.2.3.61.

133

134   To precisely characterize the 5'UTR of *GPR101* the SMARTer RACE cDNA Amplification Kit

135   (Clontech Laboratories, Inc., Mountain View, CA, USA) was employed, following the manufacturer's

136   instructions. The following reverse gene-specific primer (GSP) located within the CDS of *GPR101* was

137   used for 5'-RACE: 5'-CCCCAGCCGTAGAGTGGAGGAGTGC-3'. Samples that were analyzed by 5'-

138   RACE include matched RNA samples isolated from lymphocytes and the pituitary tumor of an X-LAG

139   patient, and RNA isolated from HEK293 cells. The identity of the 5'-RACE products was confirmed by

140   sequencing using the following primers: Isoform1F 5'-GGGCCATGGGAAAAAGATGTAGAGA-3',

141   Isoform2F 5'-CGCTGCCTCTCAGAGCTT-3', Isoform3F 5'-TTTTGCCTTTCTCCCCAGCT-3',

142   Isoform4F 5'-CCAACGAGAGAGTGCACACA-3', Isoforms1-4commonR 5'-

143   GCACGGTTGAGCGGATGAT-3'. The identity of the entire 3'UTR of *GPR101* determined by RNAseq

144   was confirmed by using the following primer pairs: 3'UTR_1F 5'-GGTGGGACTGAAGGCAAGAT-3'

145   and 3'UTR_1R 5'-ATTCTGCCTTCCTTGCTCCC-3', 3'UTR_2F 5'-

146   AGAAGTGGCCTAAGTCCTACAA-3' and 3'UTR_2R 5'-GGGAATCTCAGTGCCCTTGA-3',

147   3'UTR_3F 5'-CCCCTTCAAAGTGTATCAGAGA-3' and 3'UTR_3R 5'-

148   CTGGCACTTTTCCTTACCTGT-3', 3'UTR_4F 5'-TGGAAGCAGGGAAGGAGAGA-3' and

149   3'UTR_4R 5'-AGTGACCTGGCAATCTGTGG-3', 3'UTR_5F 5'-GTCAAATGCCCAGCTTGCTT-3'

150   and 3'UTR_5R 5'-ACCGTAGCTGTGACCCCATA-3', 3'UTR_6F 5'-CCCCTTTGGTTGGTTGTCTG-

151   3' and 3'UTR_6R 5'-TCAATTTCTCCGCATTTCCCA-3'.

152

153   There are some intrinsic limitations in the techniques employed: RNAseq experiments are limited by the

154   number of reads covering specific regions in each sample making thus difficult to exactly determine the

155   boundaries of transcripts; RACE products can be incomplete due to the premature termination of first-

156   strand cDNA synthesis caused by pausing of the reverse transcriptase enzyme, especially for long RNAs

157   (SMARTer® RACE 5'/3' Kit User Manual, Clontech). To describe the beginning of each alternative

158   5'UTR exon 1 and the end of the 3'UTR we decided to adopt a conservative approach and to report in the

159 Results section genomic coordinates that were verified by Sanger sequencing. Based on the location of

160 the primers used for sequencing it is therefore possible that actual transcripts may vary by a few bp.

161

162 *mRNA expression analysis*

163

164 The expression levels of human *GPR101* were measured by RT-qPCR in a Human Normal cDNA Array

165 containing 48 samples covering all major human normal tissues (HMRT103, Origene, Rockville, MD,

166 USA). The TaqMan system using a ready-made probe-primer kit supplied from Applied Biosystems was

167 used (Assay ID: Hs00369662_s1). This assay is located within GPR101 CDS and can detect all four

168 isoforms. To determine the expression levels of each individual transcript, isoform-specific TaqMan

169 assays were designed using the Primers Express v3.0 software (Applied Biosystems, Foster City, CA,

170 USA):

TaqMan Assay	Primer/Probe	Sequence (5'-3')	Assay length (bp)
Isoform 1	Forward	AAGAATCAGGAAACGGTGAGACA	
	Reverse	CTGTTTACGACCGTCAGTTTCTGT	112
	Probe	CGAGAGACACTGGAGAAG	
Isoform 2	Forward	TGCCTCTCAGAGCTTGCAG	
	Reverse	GCACAGACCGCACTCAGT	227
	Probe	CAGCAGCAACAGCATAGGC	
Isoform 3	Forward	CGGCAAAATAAAGTCCACAGGG	
	Reverse	GAGACAGGTTGCAGGCTCAG	141
	Probe	ACCAAGCATAGGCACTGAGT	
Isoform 4	Forward	CAGAGACACAGCGAACGACA	
	Reverse	GAGACAGGTTGCAGGCTCAG	101
	Probe	GAACAGCATAGGCACTGAGTG	

8

171

172   Total RNA was reverse transcribed to cDNA using the Superscript III Kit (Applied Biosystems)

173   according to the manufacturer's protocols. The following samples were analyzed: pituitary tumors and

174   lymphocytes of both X-LAG patients, lymphocytes of the unaffected parents of one X-LAG patient, two

175   GH-secreting tumors without Xq26.3 microduplication, four normal pituitary glands, and HEK293 cells.

176

177   Human *ACTB* (Beta Actin, Applied Biosystems, 4326315E) was used as endogenous control in all

178   reactions. TaqMan assays were performed according to manufacturer's protocol (Applied Biosystems)

179   and analyzed using the $\Delta\Delta Ct$ method, as previously described (Trivellin et al. 2014).

180

181   Total RNA was extracted from individual rat anterior pituitary glands using an RNeasy Plus Mini Kit

182   (Qiagen, Valencia, CA, USA) and reverse transcribed with a Transcriptor First Stand cDNA Synthesis Kit

183   (Roche Applied Sciences). The expression levels of rat *Gpr101* were measured by RT-qPCR using the

184   TaqMan system (Assay ID: Rn02084654_s1, Applied Biosystems) and normalized on *Gapdh* expression,

185   as previously described (Bjelobaba, et al. 2015).

186

187   Total RNA extracted from WT EK zebrafish embryos at different developmental stages was a generous

188   gift of Dr. Sadie Bergeron (NIH, Bethesda, USA). 1 μg of RNA was treated with DNase I (Applied

189   Biosystems) to remove genomic DNA contamination and then reverse transcribed to cDNA using the

190   Superscript III Kit (Applied Biosystems) according to the manufacturer's protocols. The expression levels

191   of zebrafish *gpr101* were measured by RT-PCR using the following primers: Forward:

192   TTCGTGCTCAACTTGCTCCT, Reverse: CGCAGACACTAGGGCAGAAT. Beta-actin (*bactin1*) was

193   used as endogenous control (McCurley and Callard 2008).

194

195   *Zebrafish embryo production*

196

197   WT EK zebrafish were used for all embryo production. All embryos were staged according to (Kimmel,

198   et al. 1995).

199

200   *Whole-mount in situ hybridization (WISH)*

201

202   A 1438 bp genomic sequence corresponding to the entire zebrafish *gpr101* gene

203   (ENSDARG00000039218) was cloned into the pCS2 vector and a DIG-labeled RNA antisense probe was

204   synthesized by *in vitro* transcription (DIG RNA Labeling Kit (SP6/T7), Sigma-Aldrich, St. Louis, MO,

205   USA) following manufacturer's protocol. The *gh1* antisense probe was a generous gift of Dr. Alberto

206   Rissone (NIH, Bethesda, USA). WISH was then performed in WT EK embryos collected at different

207   stages of development as previously described (Thisse and Thisse 2008). Staining with sense probes for

208   both *gpr101* and *gh1* showed no staining, as expected (data not shown). Double

209   WISH/immunofluorescence was performed as previously described (Barresi, et al. 2005). The following

210   antibodies were used: rabbit polyclonal anti-PRL (1:500, generous gift of Dr. Akiyoshi Takahashi,

211   Kitasato University, Japan), mouse monoclonal anti-tyrosine hydroxylase (1:400, MAB318, clone LNC1,

212   EMD Millipore, Billerica, MA), and appropriate rabbit/mouse secondary Alexa Fluor 488 (1:400,

213   ThermoFisher Scientific, Waltham, MA USA). Stained embryos were cleared in glycerol and

214   photographed.

215

216   *Tissue collection and immunostaining*

217

218   Normal human pituitary tissues were collected during autopsy. Changes in cell density in the fetal

219   pituitary were studied according to gestational age and location in the lateral or medial portions of the

220   gland. There were 18 fetal pituitary samples from gestational ages ranging from 12 to 38 weeks. Of these,

221   seven pituitary samples could be analyzed completely in lateral and medial portions on both sides of the

222   developing gland. Formalin-fixed paraffin embedded samples of whole fetal pituitaries preserved at

223   autopsy were sectioned using standard techniques. All paraffin embedded slides were submitted to de-

224   parafinization, rehydration and antigen retrieval for 30 min in citrate buffer solution (pH 6.0). All slides

225   were incubated with 10% normal donkey serum (NDS) for one hour for blocking of nonspecific binding,

226   and afterwards incubated with the following primary antibodies: rabbit anti-GPR101 (dilution 1:500;

227   SAB4503289, Sigma-Aldrich, St. Louis, MO, USA), goat anti-GH (dilution 1:100, sc-10364; Santa Cruz

228   Biotechnology, Santa Cruz, CA, USA) overnight at 4C. All slides were incubated for 1-2 hours with the

229   following secondary antibodies: donkey anti-rabbit 555 (A-31572, Applied Biosystems, Foster City, CA)

230   and donkey anti-goat 488 (A-11055, Applied Biosystems), both at 1:500 dilution. Prolong gold mounting

231   media with DAPI (P36934, Applied Biosystems) was used to set the slides. As negative control, a

232   specimen's section was incubated under identical conditions without primary antibody. Fluorescence was

233   analyzed with a Leica AF6000 microscope (Leica, Allendale, NJ, USA) at a 63x magnification with fixed

234   time of exposure for all samples. Subsequently, images were deconvoluted with the supplied Leica image

235   processing software (Leica).

236

237   Male and female rat and female Rhesus monkey pituitaries were isolated and fixed in Bouin's solution for

238   48 hours. Tissue was then processed for paraffin embedding. Five-micrometer-thick coronal sections of

239   the pituitary tissues were deparaffinized and the heat induced antigen retrieval was done in a microwave

240   for 5 min with 10 mM citrate buffer pH 6. The rabbit polyclonal anti-GPR101 (SAB4503289, Sigma-

241   Aldrich, St. Louis, MO, USA) was applied in 2 μg/ml dilution overnight at 4 ºC. Immunofluorescence

242   staining was performed using a Tyramide signal amplification kit (T20922, Invitrogen, Carlsbad, CA,

243   USA) with horseradish peroxidase-goat anti-rabbit IgG and Alexa fluor 488 tyramide, according to the

244   manufacturer's instructions. Tissue was further processed for the second staining with another rabbit

245   polyclonal antibody; to avoid false positive overlapping of signals microwave treatment was used (Toth

246   and Mezey 2007) as previously described (Li, et al. 2011). Rabbit anti-rat PRL, LHβ, ACTH, TSHβ, GH

247   antisera were obtained from Dr. A. F. Parlow (National Institute of Diabetes and Digestive and Kidney

248   Diseases (NIDDK), National Hormone and Peptide Program, Torrance, CA, USA) and used at a 1:400

249   dilution. For monkey tissue, rabbit anti-GH was obtained from DAKO (Dako North America, Inc.,

250   Carpinteria, CA, USA), 1:400 dilution. Donkey anti-rabbit Alexa fluor 647 was then applied at a 1:400

251   dilution for 2 hours. The sections were mounted in Mowiol and visualized under an inverted confocal

252   microscope (Zeiss LSM 510). The same linear adjustments for brightness, contrast and color balance have

253   been applied with Adobe Photoshop CS6 to each entire image.

254

255   *Statistical analysis*

256

257   For the mRNA expression experiments in rat pituitary tissue a statistical analysis was performed with

258   StatsDirect software (Addison-Wesley-Longman, Cambridge, UK). Data are presented as the mean ±

259   standard error of the mean (SEM) values from 6-37 pituitaries per group (Bjelobaba et al. 2015).

260   Comparisons were calculated using a two-tailed Student's t test for unpaired data. The data were

261   considered to be significant when $P < 0.05$.

262

263   **Results**

264

265   *In silico analysis to predict the promoter and untranslated regions (UTRs) of human GPR101*

266

267   For both mouse and rat *Gpr101* orthologues the untranslated regions (UTRs) have been characterized,

268   whereas for the human *GPR101* gene only the CDS, located on the reverse strand of chromosome X at

269   136,112,307-136,113,833 (hg19 assembly), has been characterized. Since the mouse *Gpr101* RefSeq is

270   reported as "Validated", while the rat RefSeq status is "Provisional", we decided to BLAST the mouse

271   *Gpr101* mRNA (NM_001033360.3) to the human genome. This alignment resulted in 74% identity and

272   83% coverage. Three regions were aligned with the human genome, predicting the UTRs of *GPR101*:

273   region 1 (215 bp from 136,115,961 to 136,116,175), which includes the putative 5'UTR; region 2 (2380

274   bp from 136,111,456 to 136,113,835), which includes the CDS and part of the putative 3'UTR; and

275   region 3 (1901 bp from 136,106,090 to 136,107,990), which includes most of the putative 3'UTR (Figure

276   1A). This analysis suggests that a splicing event occurs within the 3'UTR. The prediction of a long

277   3'UTR is also supported by the conservation of the DNA sequence in vertebrates (Supplemental Figure

278   1A).

279   We then analyzed the putative human *GPR101* promoter region by selecting 5 kb upstream of the

280   predicted transcription start site (TSS) and aligning it to the mouse genome to look for conserved regions

281   that are likely to harbor important transcriptional sites. This alignment resulted in 78% identity and 21%

282   coverage. From this analysis four regions were aligned with the human genome (Figure 1B). We then

283   further analyzed this 5 kb region by employing three different software programs: MPromDb (a

284   mammalian promoter database), the UCSC genome browser track annotating TF binding sites assayed by

285   ChIP-seq, and Genomatix. MPromDb predicted a 1097 bp promoter region located at 136,116,010–

286   136,117,106. The location of the TF binding sites predicted with the other two software packages are

287   shown in Figure 1B and Supplemental Figure 1B and 1C. Taken together, these *in silico* analyses showed

288   that the putative *GPR101* promoter is likely located within 2 kb upstream of the predicted TSS, with the

289   most important sites present within the first 1000 bp.

290

291   *In vitro characterization of GPR101 UTRs*

292

293   To confirm the *in silico* analysis we characterized *in vitro* the UTRs of *GPR101*. Four *GPR101* isoforms

294   were identified in different human mRNA samples by the combination of 5'-RACE and RNAseq (Figure

295   2A and Supplemental Figure 2). The four isoforms are the result of alternative splicing at the 5'UTR,

296   whereas they share a common 3'UTR that extends for 6.1 kb after the stop codon, until position

297   136,106,207.

298

299   Isoform 1 consists of a single exon. The 5'UTR extends for 1300 bp upstream of the Start Codon

300   (136,113,834-136,115,136). This isoform was observed in HEK293 cells and in all pituitary samples,

301   albeit at different expression levels. In particular, the expression of this isoform in HEK293 cells was at

302   the limit of detection by RT-qPCR, but we were nonetheless able to identify it in these cells by 5'-RACE

303   PCR (Figure 2C).

304

305   Isoform 2 consists of two exons and was observed by RNAseq and 5'-RACE PCR in the pituitary tumors

306   of the two X-LAG patients and in HEK293 cells. The first exon is entirely non-coding and extends for

307   215 bp (136,115,961-136,116,175), while the second exon consists of 92 non-coding bp upstream of the

308   Start Codon (136,113,834-136,113,925) and the 1527 bp CDS. The two exons are separated by a 2035

309   bp-long intron.

310

311   Isoforms 3 and 4 also consist of two exons and differ from isoform 2 at exon 1. In isoform 3, exon 1

312   extends for 302 bp (136,115,139-136,115,440) and is separated from exon 2 by a 1213 bp-long intron; in

313   isoform 4, exon 1 extends for 295 bp (136,114,730-136,115,024) and is separated from exon 2 by an 804

314   bp-long intron. Both isoforms were found in the pituitary tumors of the two X-LAG patients; isoform 4

315   was also observed in HEK293 cells but only by 5'-RACE PCR.

316

317   While we were able to detect isoforms 2, 3, and 4 by 5'-RACE PCR in different cells, their expression

318   was detectable by RT-qPCR only in the pituitary tumors of X-LAG patients, albeit at very low levels

319   (Figure 2B). Both intronic sequences of isoforms 2, 3, and 4 present classic donor and acceptor splice

320   sites.

321

322   *GPR101 mRNA expression in adult human tissues*

323

324   To determine which tissues express *GPR101* in humans, we performed RT-qPCR in a panel of 48 normal

325   adult tissues. *GPR101* was generally expressed at low levels in all tissues analyzed (Figure 3A). Since

326   *Gpr101* was previously shown to be expressed predominantly in the brain in rodents, we analyzed

327   *GPR101* messenger levels in a panel of 24 different human brain regions. In adult humans, the highest

328   *GPR101* levels were observed in the nucleus accumbens. Other areas that showed relatively high

329   expression were the medulla and the occipital lobe (Figure 3B).

330

331   *GPR101 expression in the adult anterior pituitary gland of other species*

332

333   We previously reported that GPR101 expression in the normal adult human pituitary is very low or absent

334   (Trivellin et al. 2014). We confirm here our previous data by extending the analysis to further samples

335   (Supplemental Figure 3). To determine if this expression pattern is similar in other species, we assessed

336   *GPR101* expression in the adult pituitary of rhesus monkey and rat. In rhesus monkey pituitary tissue

337   sections, GPR101 immunoreactivity was observed in gonadotroph cells, whereas all the other hormone-

338   secreting cells stained negative (Figure 4A-F). Immunohistochemical analysis in the rat pituitary showed

15

339   a different expression pattern, with only GH-secreting cells expressing GPR101 (Figure 4G-I and

340   Supplemental Figure 4).

341

342   *GPR101 expression during development*

343

344   In order to determine if the expression of GPR101 can be regulated in a time-dependent manner during

345   different stages of human growth, we performed GPR101 immunostaining in anterior pituitary tissues

346   collected at autopsy during fetal development, early childhood and adolescence. GPR101-positive cells

347   were absent or seen at very low levels in pituitary samples from before 19 weeks of gestation (Figure 5A,

348   5B). From 19-25 weeks, about 20-30% of cells were positive for GPR101 (Figure 5C, 5D). Thereafter

349   positivity for GPR101 rose steadily from nearly 40% positive at 26 weeks to >65% of cells positive at 38

350   weeks of gestation (Figure 5E). However, studying medial and lateral locations separately, we found that

351   GPR101 positivity in the medial portions of the pituitary remained markedly lower than that of the lateral

352   portions throughout the fetal pituitary sample series.

353

354   In the pituitary tissue from a two-year-old male child very few if any GPR101-positive cells were seen

355   (Figure 5F). In contrast, we observed several GPR101-expressing cells in two male adolescent subjects

356   aged 10 and 17 years old (Figure 5G, 5H). As previously noted (Trivellin et al. 2014) the positive cells

357   were not GH-positive. GPR101 expression was virtually absent in an adult 21-year-old male (Figure 5I).

358

359   We also evaluated GPR101 expression during development in other species. GPR101 immunostaining in

360   the forebrain of E20 female rat embryos was observed in several regions, including accumbens

361   subventricular zone, rhinencephalon, striatal neuroepithelium, striatal subventricular zone, septum,

362   hippocampus, and anterior hypothalamus (Figure 6). No sex differences were observed. We did not

363   observe overlapping of GPR101 staining with vimentin (data not shown). In the prepubertal rat pituitary,

364   *Gpr101* mRNA expression showed sexual dimorphism during the first three weeks of life, with higher

16

365   expression levels in females than males (Supplemental Figure 5). An expression pattern similar to that

366   observed in adult rat pituitary was also seen in prepubertal pituitary tissue by immunostaining (data not

367   shown).

368

369   In the developing zebrafish embryo we performed whole-mount in situ hybridization (WISH) using a

370   *gpr101*-specific probe to determine at what stages and in which anatomical regions *gpr101* is expressed.

371   Transcripts of *gpr101* showed a bimodal temporal expression pattern, with early, and likely maternal,

372   messenger RNA levels waning during very early stages, then rising from the bud stage through 2 dpf, and

373   remaining constant through the remainder of embryonic development (Figure 7). This expression pattern

374   was recapitulated when we measured *gpr101* mRNA expression by RT-CR (Supplemental Figure 6).

375   Starting at 2dpf, a strong and brain-specific staining was observed (Figure 7 and 8), overlapping with the

376   location of pituitary and hypothalamus (Figure 8).

377

378 **Discussion**

379

380 GPR101 is an orphan GPCR that is duplicated in patients with X-LAG. High GPR101 expression was

381 observed in patients' pituitary tumor tissues, while low expression levels were seen in their blood and in

382 normal pituitaries, suggesting that *GPR101* might be the cause of the disease (Trivellin et al. 2014).

383 However, very little is known regarding GPR101 function, expression in different tissues and species, and

384 transcript structure. We undertook the present study in order to better understand GPR101 physiology and

385 how its de-regulation might cause X-LAG, and possibly other diseases (Tariq 2013). Four GPR101

386 transcripts were identified. These transcripts are characterized by alternative splicing at the 5'UTR due to

387 different TSSs, while they share a common 6.1 kb-long 3'UTR.

388

389 Isoform 1 seems to be the predominant isoform in the human pituitary. The higher expression levels

390 observed in the pituitary tumors harboring a duplication of *GPR101* compared to normal pituitary tissue

391 or tumors without the duplication suggests that this is the isoform that leads to the previously observed

392 GPR101 over-expression (Trivellin et al. 2014). The other three isoforms are expressed at much lower

393 levels and were readily detectable by quantitative PCR only in *GPR101*-duplicated pituitary tumors.

394 These isoforms might therefore represent minor transcripts but we cannot exclude that they are

395 specifically expressed in other tissues, where they might have a biological function. Isoform 3 was

396 identified by both techniques only in mRNA isolated from pituitary tumors harboring an Xq26.3

397 microduplication. While it is possible that this is a pituitary-specific transcript, we could not exclude that

398 this transcript might be generated due to the rearranged surrounding genomic region. Interplay between

399 the tridimensional organization of the genome and the regulation of alternative splicing has indeed been

400 previously shown (Curado, et al. 2015).

401

402 Different approaches support the real nature of the identified transcripts. First, the *in vitro* data matched

403 well with the *in silico* predictions. The predicted 5'UTR matches exactly with exon 1 of isoform 2.

18

404   Moreover, we previously showed that *GPR101* is predicted to harbor a 747 bp CpG island (chrX:

405   136,115,961–136,116,708) (Trivellin et al. 2014). Since CpG islands normally lie in the promoter region

406   and can extend into the 5'UTR (Medvedeva, et al. 2010), this result is in line with the location of the

407   identified 5'UTR. In addition, the canonical splicing sites identified for all alternative intronic regions

408   support their *bona fide* nature. While we did not observe a splicing event within the 3'UTR as inferred by

409   the *in silico* alignment, the end of the predicted 3'UTR supports the *in vitro* observation. Second,

410   concordant results were observed between the two techniques used for the *in vitro* characterization,

411   RACE-PCR and RNAseq. It would be interesting for future studies to investigate the function of the very

412   long 3'UTR, which could harbor several binding sites for regulatory molecules, such as miRNAs or

413   lncRNAs.

414

415   Our *in silico* analyses showed that the putative *GPR101* promoter region likely extends at least until 1 kb

416   upstream of isoform 2 TSS (Figure 1D). Several features suggest that the *GPR101* promoter is broad and

417   TATA-less, including the identification of a putative core promoter element associated with TATA-less

418   genes (XCPE1, Supplemental Figure 1C), the high GC content, and the multiple TSSs (Carninci, et al.

419   2006; Sandelin, et al. 2007; Tokusumi, et al. 2007). A couple of putative TATA boxes were identified but

420   their location far away from the TSS (Supplemental Figure 1C) argues against a role in transcription

421   activation. A conserved TATA box is indeed usually located around 30 bp upstream of the TSS (Sandelin

422   et al. 2007). Future *in vitro* studies are necessary to decipher the function of the promoter region as well

423   as of possible enhancers, which could lead to abnormally high *GPR101* expression levels when

424   rearranged.

425

426   Our analysis of GPR101 expression in adult human tissues showed that GPR101 is generally expressed at

427   very low levels, with the exception of the brain, where relatively high expression was seen in several

428   regions. This brain-specific expression correlates well with the *in silico* identification of a CpG-enriched

429   promoter since this type of promoter could be advantageous in this organ due to its property of fine-

430   tuning gene transcription (Gustincich, et al. 2006). We also previously showed that when GPR101 is

431   over-expressed, even in the absence of a ligand, it exhibits high constitutive activity by strongly activating

432   the cAMP pathway (Trivellin et al. 2014). It is therefore conceivable that GPR101 expression is normally

433   turned down or completely switched-off in most tissues, in line with our observations.

434

435   Within the brain, the highest GPR101 expression levels were observed in the nucleus accumbens, a region

436   that plays an important role as the reward center and where other GPCRs characterized by constitutive

437   activity play vital roles (Meye, et al. 2014). This finding suggests that GPR101 might be involved in the

438   regulation of different behaviors, such as food seeking: indeed, about a quarter of X-LAG cases presented

439   with increased eating/food seeking that accompanied their markedly increased statural overgrowth

440   (Beckers et al. 2015). Further studies are thus needed to determine if a functional link between GPR101,

441   behavior, and physical growth exists.

442

443   The pituitary gland, by responding to hormonal signals from the hypothalamus, constitutes an anatomical

444   and physiological link between the nervous and the endocrine systems to control several key body

445   functions, such as somatic growth. In order to gain insight into GPR101's putative role in regulating

446   growth, we decided to investigate its expression in the brain and pituitary of different species both during

447   and after development. GPR101 expression in the adult anterior pituitary gland showed remarkable

448   differences among species. While GPR101 is expressed at low levels in adult human pituitary, GPR101-

449   positive cells were seen in both monkey and rat pituitary. However, the cell types expressing the receptor

450   differ in those two species. These results therefore suggest that GPR101 might have different functions in

451   the pituitary of different species during development and in adult life. In particular, in normal adult

452   human pituitary, the main physiological role of GPR101 might already be finished. This is supported by

453   the fact that high GPR101 expression was observed in the human pituitary gland during fetal

454   development, which suggests that the receptor might play an important physiological role in this organ as

455   it proliferates and matures. In particular, the observation that GPR101 is mainly expressed in the lateral

456   wings of the anterior lobe, where the vast majority of GH- and PRL-secreting cells reside (Al-Gahtany, et

457   al. 2003; Melmed, et al. 2011), suggests that the receptor might be involved in regulating, directly or

458   indirectly, the embryonic differentiation of mammosomatotroph cells. Even if no causative association

459   can be demonstrated by these data, it is interesting to note that in humans this is the stage were the

460   maximum rate of growth in length and weight is achieved (Bogin 1999). Furthermore, duplication of

461   GPR101 in the setting of X-LAG syndrome is associated with a specific pituitary pathology related to

462   somatotroph, mammosomatotroph and lactotroph cells (Trivellin et al. 2014). It will be of interest in

463   future studies to further define the main sites of hormonal dysregulation due to GPR101 duplication in X-

464   LAG syndrome, as we recently showed likely hypothalamic GHRH hypersecretion can accompany the

465   pituitary pathology in this condition (Daly, et al. 2016a).

466

467   We also investigated GPR101 expression in the pituitary during postnatal development. We observed

468   very few GPR101-positive cells in normal pituitary tissue from a two year old-subject. This result seems

469   to be in line with the observation that postnatal growth rates start to rapidly decelerate during

470   infancy/early childhood (Bogin 1999; Stevens, et al. 2013). We hypothesized that GPR101 might also

471   play a role during the so-called "adolescent growth spurt", a rapid and intense post-pubertal increase in

472   the rate of growth in height and weight that, among all primate species, is observed only in humans

473   (Bogin 1999; Lui, et al. 2015). This growth stage can account for up to 15-20 % of adult final height

474   (Cousminer, et al. 2013) and, even if there is considerable variation between individuals and populations,

475   begins on average at 10 years for girls and 12 for boys, and ends at about 18-19 years for girls and 20-22

476   years for boys (Bogin 1999; Lui et al. 2015). To test this hypothesis we assessed GPR101 expression in

477   pituitary tissues from adolescent individuals, to see if it might increase. We observed many GPR101-

478   expressing cells in two male subjects aged 10 and 17 year old, whereas practically absent expression was

479   seen in a 21 year-old male. These results indicate that GPR101 expression seems indeed to increase

480   during the adolescent growth spurt and then wanes once this growth phase terminates, thus supporting our

481   hypothesis. However, we cannot exclude that GPR101 might also act at a superior (e.g. hypothalamic)

482   level as suggested recently (Daly et al. 2016a). Altogether these results suggest that the expression of

483   GPR101 in the human pituitary is tightly timed during pre- and post-natal development and lead us to

484   speculate that the continuously high expression levels observed in the tumor tissues of patients with X-

485   LAG compromise this fine regulation leading to uncontrolled growth.

486

487   GPR101 expression during pre- and post-natal development was also assessed in rat and zebrafish brain

488   and pituitary. GPR101 was found to be expressed in several regions of the fetal rat brain, including the

489   hypothalamus, but excluding the pituitary. These results are in corroboration with what was previously

490   observed in the adult rat brain (Bates et al. 2006). The absence of GPR101 expression in fetal rat pituitary

491   could be explained by the fact that growth in rodents shows a very different course than in humans, with

492   maximum growth rate occurring in the early postnatal period (Stevens et al. 2013). In agreement with this

493   hypothesis, GPR101 expression in the pituitary was detected only after birth, even if its expression levels

494   were generally quite low. We observed sexual dimorphism with the highest relative GPR101 levels

495   observed in 2-week old females. Interestingly, at this point of female development were also observed the

496   highest mRNA levels for *Gnrhr*, *Lhb*, *Fshb* and gonadotroph/thyrotroph-specific α subunit (*Cga*)

497   (Bjelobaba et al. 2015). These results lead us to speculate that in rodents, in contrast to humans,

498   GPR101's putative role in growth is largely absent after birth. Moreover, GPR101 related effects seem to

499   be exerted both in the pituitary and in superior brain regions (Bates et al. 2006; Nilaweera et al. 2007;

500   Nilaweera et al. 2008; Regard et al. 2008; Trivellin et al. 2014). In the zebrafish embryo, strong *gpr101*

501   expression was detected in the brain area. It will be important for future studies to generate specific

502   knockout animals in order to investigate the role(s) played by GPR101 during development.

503

504   In conclusion, this study shows that different human GPR101 transcripts exist and that the brain is the

505   major site of GPR101 expression across different species, suggesting a conservation of function for this

506   GPCR. On the other hand, we also observed species-specific expression patterns, especially concerning

507   the pituitary, with GPR101 expression being regulated in a cell type- and developmental stage-specific

508   manner in different species. Altogether these findings indicate that a) *GPR101* plays an important role in

509   both brain and pituitary development and b) the differences we observed might reflect the very different

510   growth, development and maturation patterns among species.

511

531   **Declaration of interest:**

532   There is no conflict of interest that could be perceived as prejudicing the impartiality of the research

533   reported.

534 **References**

535

536 Al-Gahtany M, Horvath E & Kovacs K 2003 Pituitary hyperplasia. *Hormones (Athens, Greece)* **2** 149-

537 158.

538 Barresi MJ, Hutson LD, Chien CB & Karlstrom RO 2005 Hedgehog regulated Slit expression determines

539 commissure and glial cell position in the zebrafish forebrain. *Development* **132** 3643-3656.

540 Bates B, Zhang L, Nawoschik S, Kodangattil S, Tseng E, Kopsco D, Kramer A, Shan Q, Taylor N,

541 Johnson J, et al. 2006 Characterization of Gpr101 expression and G-protein coupling selectivity. *Brain*

542 *research* **1087** 1-14.

543 Beckers A, Lodish MB, Trivellin G, Rostomyan L, Lee M, Faucz FR, Yuan B, Choong CS, Caberg JH,

544 Verrua E, et al. 2015 X-linked acrogigantism syndrome: clinical profile and therapeutic responses.

545 *Endocr Relat Cancer* **22** 353-367.

546 Bjelobaba I, Janjic MM, Kucka M & Stojilkovic SS 2015 Cell Type-Specific Sexual Dimorphism in Rat

547 Pituitary Gene Expression During Maturation. *Biol Reprod* **93** 21.

548 Bogin B 1999 Evolutionary perspective on human growth. *Annu Rev Anthropol* **28** 109-153.

549 Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS,

550 Engstrom PG, Frith MC, et al. 2006 Genome-wide analysis of mammalian promoter architecture and

551 evolution. *Nat Genet* **38** 626-635.

552 Cousminer DL, Berry DJ, Timpson NJ, Ang W, Thiering E, Byrne EM, Taal HR, Huikari V, Bradfield

553 JP, Kerkhof M, et al. 2013 Genome-wide association and longitudinal analyses reveal genetic loci linking

554 pubertal height growth, pubertal timing and childhood adiposity. *Hum Mol Genet* **22** 2735-2747.

555 Curado J, Iannone C, Tilgner H, Valcarcel J & Guigo R 2015 Promoter-like epigenetic signatures in

556 exons displaying cell type-specific splicing. *Genome Biol* **16** 236.

557 Daly AF, Lysy PA, Desfilles C, Rostomyan L, Mohamed A, Caberg JH, Raverot V, Castermans E,

558 Marbaix E, Maiter D, et al. 2016a GHRH excess and blockade in X-LAG syndrome. *Endocr Relat*

559 *Cancer* **23** 161-170.

560 Daly AF, Yuan B, Fina F, Caberg JH, Trivellin G, Rostomyan L, de Herder WW, Naves LA, Metzger D,

561 Cuny T, et al. 2016b Somatic mosaicism underlies X-linked acrogigantism syndrome in sporadic male

562 subjects. *Endocr Relat Cancer* **23** 221-233.

563 Gustincich S, Sandelin A, Plessy C, Katayama S, Simone R, Lazarevic D, Hayashizaki Y & Carninci P

564 2006 The complexity of the mammalian transcriptome. *J Physiol* **575** 321-332.

565 Kakarala KK & Jamil K 2014 Sequence-structure based phylogeny of GPCR Class A Rhodopsin

566 receptors. *Mol Phylogenet Evol* **74** 66-96.

567 Kimmel CB, Ballard WW, Kimmel SR, Ullmann B & Schilling TF 1995 Stages of embryonic

568 development of the zebrafish. *Dev Dyn* **203** 253-310.

569 Lee DK, Nguyen T, Lynch KR, Cheng R, Vanti WB, Arkhitko O, Lewis T, Evans JF, George SR &

570 O'Dowd BF 2001 Discovery and mapping of ten novel G protein-coupled receptor genes. *Gene* **275** 83-

571 91.

572 Li S, Bjelobaba I, Yan Z, Kucka M, Tomic M & Stojilkovic SS 2011 Expression and roles of pannexins

573 in ATP release in the pituitary gland. *Endocrinology* **152** 2342-2352.

574 Lui JC, Garrison P & Baron J 2015 Regulation of body growth. *Curr Opin Pediatr* **26** 502-510.

575 McCurley AT & Callard GV 2008 Characterization of housekeeping genes in zebrafish: male-female

576 differences and effects of tissue type, developmental stage and chemical treatment. *BMC molecular*

577 *biology* **9** 102.

578 Medvedeva Ya, Fridman MV, Oparina NJ, Malko DB, Ermakova EO, Kulakovskiy IV, Heinzel A &

579 Makeev VJ 2010 Intergenic, gene terminal, and intragenic CpG islands in the human genome. *BMC*

580 *genomics* **11** 48.

581 Melmed S, Kleinberg D & Ho KEN 2011 Pituitary Physiology and Diagnostic Evaluation. In *Williams*

582 *Textbook of Endocrinology*, edn Twelfth Ed, pp 175-228. Boston, Massachusetts: Elsevier Inc.

583 Meye FJ, Ramakers GMJ & Adan RaH 2014 The vital role of constitutive GPCR activity in the

584 mesolimbic dopamine system. *Translational psychiatry* **4** e361.

585   Mukhopadhyay S, Wen X, Ratti N, Loktev A, Rangell L, Scales SJ & Jackson PK 2013 The ciliary G-

586   protein-coupled receptor Gpr161 negatively regulates the Sonic hedgehog pathway via cAMP signaling.

587   *Cell* **152** 210-223.

588   Nilaweera KN, Ozanne D, Wilson D, Mercer JG, Morgan PJ & Barrett P 2007 G protein-coupled receptor

589   101 mRNA expression in the mouse brain: altered expression in the posterior hypothalamus and

590   amygdala by energetic challenges. *Journal of neuroendocrinology* **19** 34-45.

591   Nilaweera KN, Wilson D, Bell L, Mercer JG, Morgan PJ & Barrett P 2008 G protein-coupled receptor

592   101 mRNA expression in supraoptic and paraventricular nuclei in rat hypothalamus is altered by

593   pregnancy and lactation. *Brain research* **1193** 76-83.

594   Regard JB, Sato IT & Coughlin SR 2008 Anatomical profiling of G protein-coupled receptor expression.

595   *Cell* **135** 561-571.

596   Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y & Hume DA 2007 Mammalian RNA

597   polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet* **8** 424-436.

598   Stevens A, Hanson D, Whatmore A, Destenaves B, Chatelain P & Clayton P 2013 Human growth is

599   associated with distinct patterns of gene expression in evolutionarily conserved networks. *BMC genomics*

600   **14** 547.

601   Tariq MC, A. E.; Belmont, J. W.; Ware, S. M. 2013 Identification of a Novel Cause of X-linked

602   Heterotaxy (Abstract). In *ASHG 2013*. Boston, USA.

603   Thisse C & Thisse B 2008 High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat*

604   *Protoc* **3** 59-69.

605   Tokusumi Y, Ma Y, Song X, Jacobson RH & Takada S 2007 The new core promoter element XCPE1 (X

606   Core Promoter Element 1) directs activator-, mediator-, and TATA-binding protein-dependent but TFIID-

607   independent RNA polymerase II transcription from TATA-less promoters. *Mol Cell Biol* **27** 1844-1858.

608   Toth ZE & Mezey E 2007 Simultaneous visualization of multiple antigens with tyramide signal

609   amplification using antibodies from the same species. *J Histochem Cytochem* **55** 545-554.

610   Trivellin G, Daly AF, Faucz FR, Yuan B, Rostomyan L, Larco DO, Schernthaner-Reiter MH, Szarek E,

611   Leal LF, Caberg JH, et al. 2014 Gigantism and acromegaly due to Xq26 microduplications and GPR101

612   mutation. *N Engl J Med* **371** 2363-2374.

613

614

Figure legends

Figure 1: *In silico* **analyses to predict human** *GPR101* **promoter and untranslated regions.** A) This panel shows the human *GPR101* aligned with mouse and rat *Gpr101*. The track displaying non-human RefSeq genes follows the display conventions for gene prediction tracks. The color shading indicates the level of review the RefSeq record has undergone: predicted (light), provisional (medium), reviewed (dark). The RNAs were aligned against the human genome using blat. A track with the predicted CpG islands is also shown underneath. The number on the left of each island represents the CpG count. B) The mouse *Gpr101* mRNA was BLASTed to the human genome to determine the predicted UTRs of *GPR101* The alignment predicted a 215 bp 5'UTR located about 2 kb upstream of the start codon and a splicing event in the 3'UTR. C and D) Analysis of the putative human *GPR101* promoter region. 5 kb upstream of the predicted *GPR101* TSS were aligned to the mouse genome (C). The alignment returned four regions which were further analyzed by MPromDb (orange bar), the UCSC genome browser track annotating TF binding sites assayed by ChIP-seq (grey bars), and Genomatix (yellow bars) (D). The identity of the predicted TFs is shown in Supplementary Figure 1B and 1C. The location of the previously predicted CpG island by CpG Island Searcher (Trivellin et al. 2014) is also shown in panel D (green bar).

Figure 2: Structure and expression of *GPR101* **transcripts.** A) Four isoforms were observed in the analyzed samples by both 5'-RACE and RNAseq. All isoforms shared a common 3'UTR, while they differ for the 5'UTR. The coding sequence (CDS) is represented as a dark grey box, the UTRs as a smaller light grey box, and the intron as a black line. B) Relative expression of the four isoforms assessed by RT-qPCR in a pituitary tumor harboring an Xq26.3 microduplication. Isoform 1 is expressed at much higher levels than the other isoforms. C) Isoform 1 expression in different tissues. The relative expression of isoform 1 was determined by RT-qPCR in the following samples: pituitary tumor and lymphocytes of an X-LAG patient, lymphocytes of her unaffected parents, normal pituitary, GH-secreting tumor without Xq26.3 microduplication, HEK293 cells. Isoform 1 expression

was detected only in pituitary samples, showing higher expression in the tumor harboring a *GPR101* duplication. All values are relative and normalized on *ACTB* expression.

Figure 3: *GPR101* mRNA expression in normal adult human tissues. A) *GPR101* expression in a panel of 48 different human tissues. B) *GPR101* expression in a panel of 24 human brain regions. The highest *GPR101* expression levels were observed in the nucleus accumbens. All values are relative and were normalized on *ACTB* expression. The dotted lines represent the average expression in all tissues analyzed in each cDNA panel.

Figure 4: Labeling of GPR101 in the anterior pituitary gland of female rhesus monkey (A-F) and rat (G-I). A) Preincubation of GPR101 antibody with blocking peptide, resulted in no staining. B) Punctuate GPR101 (green) labeling could not be attributed to somatotrophs (red). C) GPR101 (green) is expressed by LHβ positive cells (red). D-F) High magnification images showing membranous and cytoplasmic distribution of GPR101 in monkey gonadotrophs, excluding nucleus. G-I) GPR101 (green) is expressed by subpopulation of rat somatotrophs (red), as indicated by arrows. Scale bars: A-D: 20 µm; E: 10 µm.

Figure 5: Fetal pituitary gland staining characteristics for GPR101. Panel A shows an H&E stain transverse section of an entire pituitary gland from a 14-week-old fetus at low magnification (50X). The medial (M) and lateral (L) portions of the gland are highlighted. Panel B shows GPR101 staining (brown immunopositivity highlighted) in the lateral portion of a 14-week-old fetal pituitary at higher magnification (400X). Immunopositivity is localized to the cytoplasm and is negative in the nuclei. The different staining intensities for GPR101 at 24 weeks in the lateral (Panel C) and medial (Panel D) portions of the fetal pituitary can be seen (400X). Some positivity for GPR101 in capillaries can also be seen. The increasing GPR101 staining intensity in the lateral portion of the fetal pituitary is illustrated in Panel E at 34 weeks, where very high density is seen (400X). Panels F-I show GPR101 (red) and GH (green) immunofluorescence staining in the anterior pituitary during post-natal life. Almost absent GPR101 expression was seen in a two-year-old female child (F) and in a 21-year-old

male (I). Several GPR101 positive cells were present in two males aged 10 (G) and 17 (H) year old. Very little co-localization of GPR101 with GH was seen in Panels G and H. Nuclei (blue) were counterstained with DAPI.

Figure 6: Immunolocalization of GPR101 in coronal sections of rat embryo forebrain. A) GPR101 immunoreactive cells and processes were pronounced in accumbens subventricular zone (SVZ) and rhinencephalic differential field (Rh). B) Strong GPR101 immunoreactivity was observed in striatal neuroepithelium and striatal subventricular zone (SVZ) around lateral ventricle (LV) and striatal differentiating field. GPR101 positive cells and cell processes were also observed in the septal area. C) In more posterior sections, GPR101 immunoreactivity was observed throughout the hippocampus and lateral to third ventricle (3V), in the anterior hypothalamus. GPR101 positive cells and processes were also noticed in the amygdaloid differentiating field (A). Scale bar applies to all images: 0.5 mm.

Figure 7: *In situ* staining for *gpr101* during WT zebrafish development (lateral view). A) 1-cell, B) 8-cell, C) 30% epiboly, D) bud, E) 3 somite, F) 20 somite, G) 24hpf, H) 2dpf, I) 5dpf. *gpr101* mRNA (blue) is visible from 1-cell to 8-cell stage (A, B), probably representing maternal transcripts, and then reappears during the bud stage (D), approximately 10hpf. Starting at 2dpf (H) a strong and brain-specific staining is seen. Scale bar: 100 μm; hpf: hours post-fertilization; dpf: days post-fertilization.

Figure 8: *gpr101* expression in the brain of WT Zebrafish embryo. A) and B) WISH staining for *gpr101* and *gh1* (purple) in 2dpf distinct Zebrafish embryos (ventral view). *gpr101* is strongly expressed in the whole brain region (A), while a pituitary-specific staining is visible with the *gh1* probe (B). C) and D) Double WISH/immunofluorescence staining for *gpr101* (purple) and a pituitary (PRL, green, panel C) and hypothalamic marker [Tyrosine hydroxylase (TH) specific for diencephalic dopaminergic neurons, green, panel D] in a 2dpf Zebrafish embryo (ventral view). *gpr101* staining overlaps with the location of pituitary and hypothalamic neurons. The qualitatively different aspect of

gpr101 staining in panels C and D is due to the different focal planes selected for PRL and TH. Scale

bar: 50 μm applies to all panels.

Supplemental Figure legends

Supplemental Figure 1: A) The genomic sequence of the predicted *GPR101* 3'UTR is well conserved among vertebrates (red line). Each peak corresponds to a single nucleotide and its height is proportional to the degree of conservation among the 100 vertebrate species considered. In the track underneath, a multiple alignment of human *GPR101* 3'UTR with the four species investigated in this study (rhesus monkey, mouse, rat, and zebrafish) is also shown. B) UCSC track showing regions of TF binding motifs identified by ChIP-seq experiments performed by the ENCODE project in different cell lines within 5 kb upstream of *GPR101* TSS. TF binding motifs are represented by boxes. The four TFs shown are EZH2 (http://genome.ucsc.edu/cgi-bin/hgEncodeVocab?ra=encode%2Fcv.ra&term=%22EZH2%22), SUZ12 (http://v1.factorbook.org/mediawiki/index.php/SUZ12), CTCF (http://v1.factorbook.org/mediawiki/index.php/CTCF), and RAD21 (http://v1.factorbook.org/mediawiki/index.php/RAD21). The darkness of each box is proportional to the maximum signal strength observed in any cell line. The name of the TF is reported on the left of each box, while the code on the right identifies the cell lines where a signal is detected. The genomic coordinates and Cluster Score for each TF binding site is also reported below the figure. C) TF binding sites predicted by Genomatix within 5 kb upstream of GPR101 TSS. TF binding sites were filtered by strand (we considered the reverse strand only for TATA-box predictions), location (only sites located within the four human-mouse conserved regions were considered), and vertebrate elements.

Supplemental Figure 2: Sashimi plot generated with the IGV software showing the transcripts structure of *GPR101*. Total RNA extracted from the pituitary tumors of two patients with X-LAG underwent RNAseq analysis. In both patients high *GPR101* expression is seen, as previously reported (Trivellin et al. 2014). The 6.1 kb-long 3'UTR (black line) can clearly be seen. The splicing event observed in isoform 2 by 5'-RACE can also be appreciated in both patients, with "19' and "38"

representing the number of reads supporting this event. The height of each bar is proportional to the number of reads. *GPR101* is shown in the reverse orientation.

Supplemental Figure 3: Immunostaining for GPR101 (red) and GH (green) in normal adult human anterior pituitary gland. A-D) No GPR101 expression was seen in four different samples. Nuclei (blue) were counterstained with DAPI. Images were taken at 40X magnification.
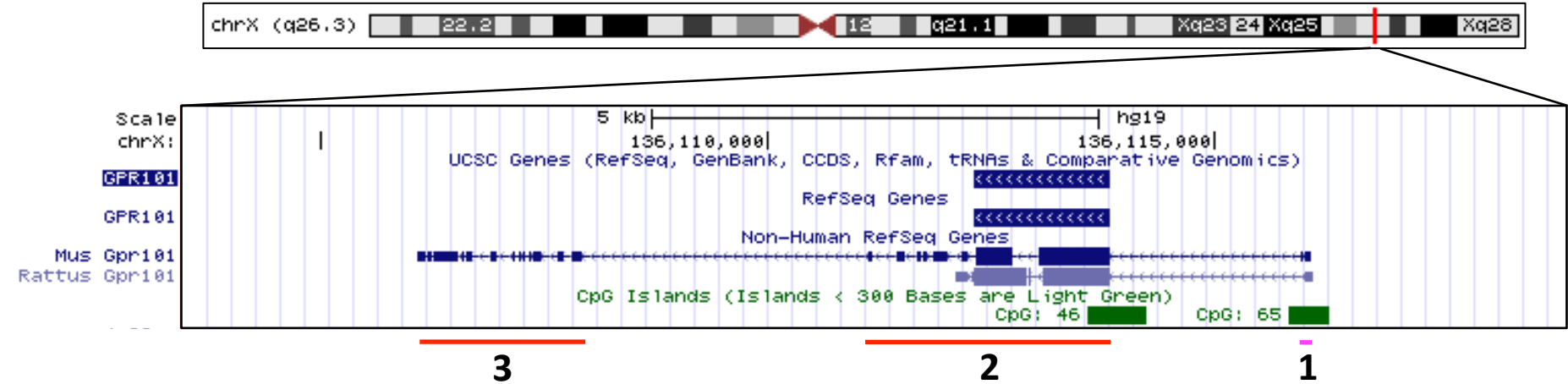
Supplemental Figure 4: Double immunofluorescence labeling for GPR101 (green) and pituitary hormones (red) showed that rat pituitary gonadotrophs (A), corticotrophs (B), thyrotrophs (C) and lactotrophs (D) do not express GPR101. Scale bar: 20 µm for both upper and lower panels.

Supplemental Figure 5: *Gpr101* mRNA expression in rat anterior pituitary during development. The highest *Gpr101* expression was observed in two week-old female pituitary tissue, then the messenger levels gradually decreased, reaching levels similar to those measured in males. In male pituitary *Gpr101* expression was low but relatively stable throughout development. Data points shown are mean ± SEM values from six to 37 animals per group, relative to *Gapdh* expression.

Supplemental Figure 6: RT-PCR for *gpr101* during WT zebrafish development. The results are consistent with what observed in WISH (Figure 7 in the main text), showing expression of presumably maternal *gpr101* mRNA during the first cell divisions and then embryo-specific transcript levels gradually appearing from the epiboly stage. β-actin expression was assessed to ensure equal loading. CTRL+: *gpr101* plasmid; CTRL-: water; h: hours; d: days.
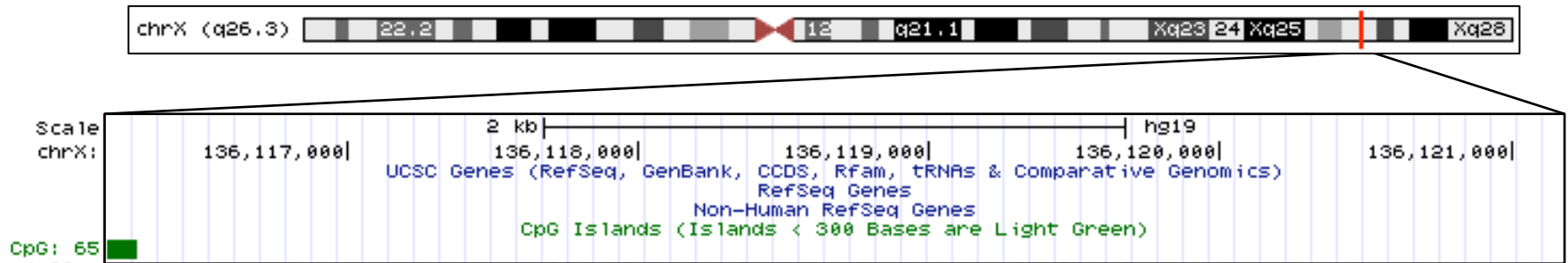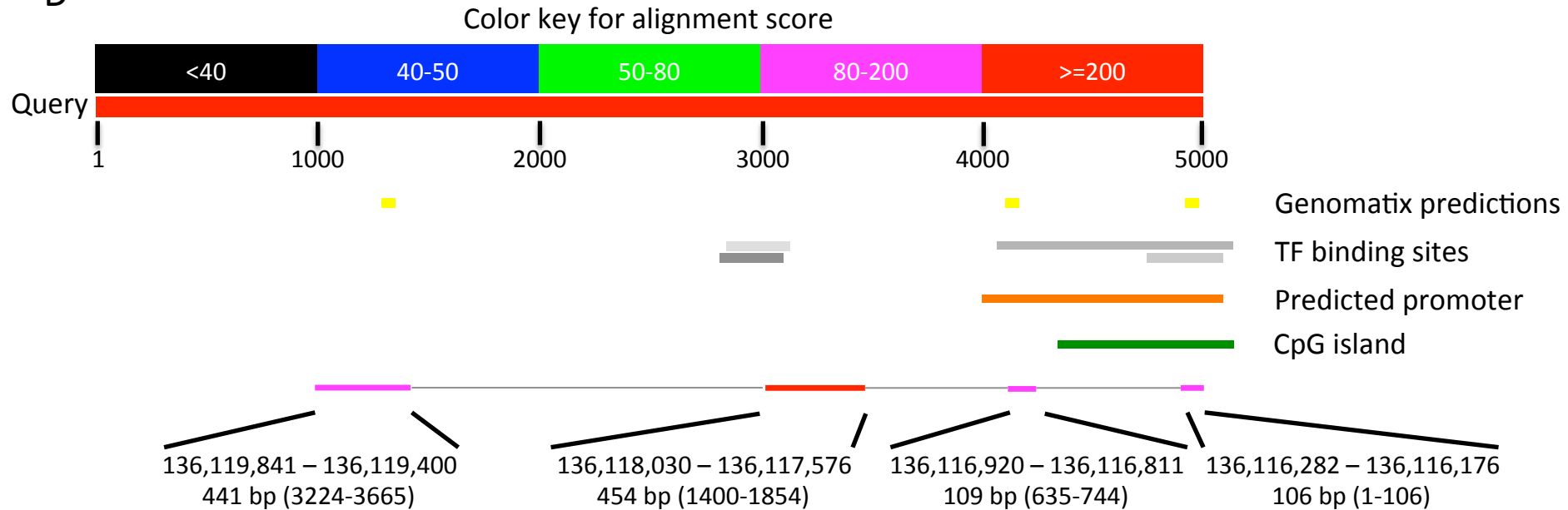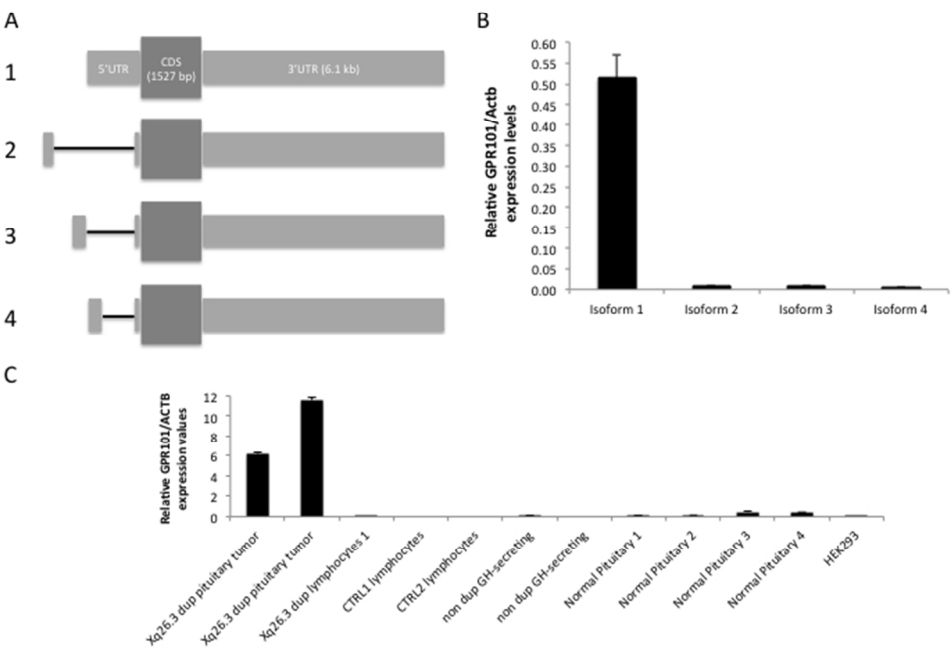
Figure 1

A

B

Figure 1

C

D

Color key for alignment score

| <40 | 40-50 | 50-80 | 80-200 | >=200 |

Query

1   1000   2000   3000   4000   5000

Genomatix predictions

TF binding sites

Predicted promoter

CpG island

136,119,841 – 136,119,400
441 bp (3224-3665)

136,118,030 – 136,117,576
454 bp (1400-1854)

136,116,920 – 136,116,811
109 bp (635-744)

136,116,282 – 136,116,176
106 bp (1-106)

Structure and expression of GPR101 transcripts. A) Four isoforms were observed in the analyzed samples by both 5'-RACE and RNAseq. All isoforms shared a common 3'UTR, while they differ for the 5'UTR. The coding sequence (CDS) is represented as a dark grey box, the UTRs as a smaller light grey box, and the intron as a black line. B) Relative expression of the four isoforms assessed by RT-qPCR in a pituitary tumor harboring an Xq26.3 microduplication. Isoform 1 is expressed at much higher levels than the other isoforms. C) Isoform 1 expression in different tissues. The relative expression of isoform 1 was determined by RT-qPCR in the following samples: pituitary tumor and lymphocytes of an X-LAG patient, lymphocytes of her unaffected parents, normal pituitary, GH-secreting tumor without Xq26.3 microduplication, HEK293 cells. Isoform 1 expression was detected only in pituitary samples, showing higher expression in the tumor harboring a GPR101 duplication. All values are relative and normalized on ACTB expression.
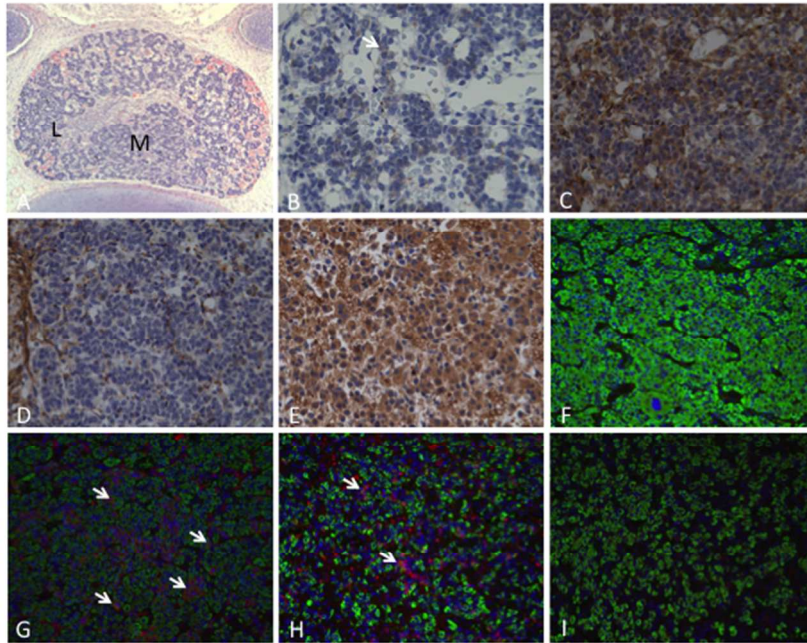254x190mm (72 x 72 DPI)
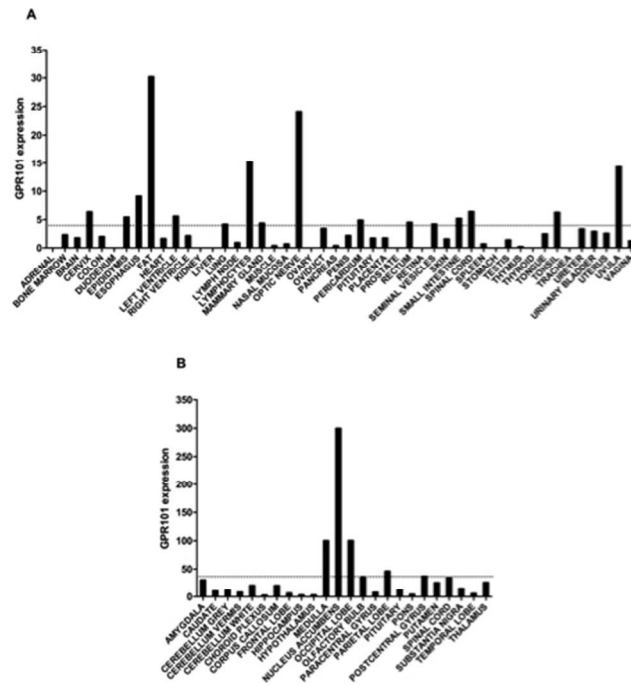
GPR101 mRNA expression in normal adult human tissues. A) GPR101 expression in a panel of 48 different human tissues. B) GPR101 expression in a panel of 24 human brain regions. The highest GPR101 expression levels were observed in the nucleus accumbens. All values are relative and were normalized on ACTB expression. The dotted lines represent the average expression in all tissues analyzed in each cDNA panel.
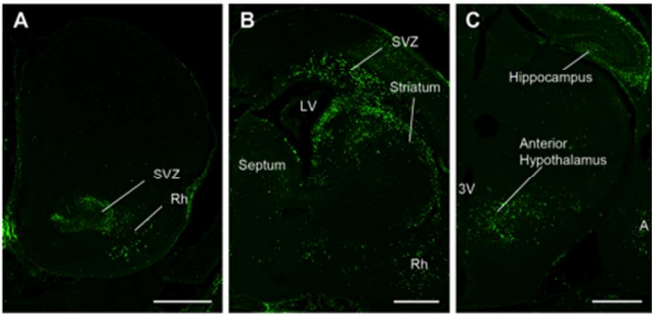254x190mm (72 x 72 DPI)
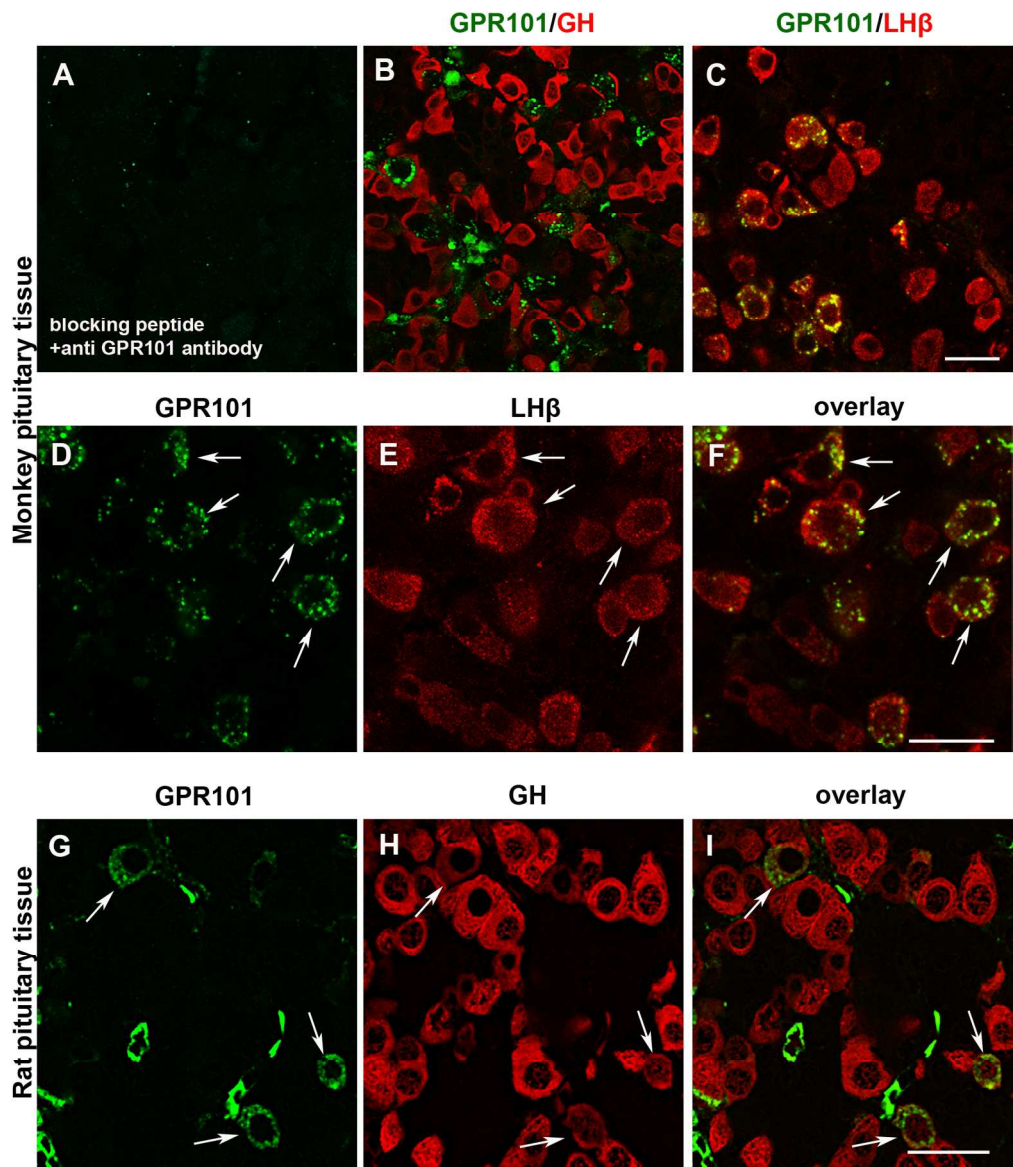
Labeling of GPR101 in the anterior pituitary gland of female rhesus monkey (A-F) and rat (G-I). A) Preincubation of GPR101 antibody with blocking peptide, resulted in no staining. B) Punctuate GPR101 (green) labeling could not be attributed to somatotrophs (red). C) GPR101 (green) is expressed by LHβ positive cells (red). D-F) High magnification images showing membranous and cytoplasmic distribution of GPR101 in monkey gonadotrophs, excluding nucleus. G-I) GPR101 (green) is expressed by subpopulation of rat somatotrophs (red), as indicated by arrows. Scale bars: A-D: 20 μm; E: 10 μm.
162x189mm (300 x 300 DPI)

Fetal pituitary gland staining characteristics for GPR101. Panel A shows an H&E stain transverse section of an entire pituitary gland from a 14-week-old fetus at low magnification (50X). The medial (M) and lateral (L) portions of the gland are highlighted. Panel B shows GPR101 staining (brown immunopositivity highlighted) in the lateral portion of a 14-week-old fetal pituitary at higher magnification (400X). Immunopositivity is localized to the cytoplasm and is negative in the nuclei. The different staining intensities for GPR101 at 24 weeks in the lateral (Panel C) and medial (Panel D) portions of the fetal pituitary can be seen (400X). Some positivity for GPR101 in capillaries can also be seen. The increasing GPR101 staining intensity in the lateral portion of the fetal pituitary is illustrated in Panel E at 34 weeks, where very high density is seen (400X). Panels F-I show GPR101 (red) and GH (green) immunofluorescence staining in the anterior pituitary during post-natal life. Almost absent GPR101 expression was seen in a two-year-old female child (F) and in a 21-year-old male (I). Several GPR101 positive cells were present in two males aged 10 (G) and 17 (H) year old. Very little co-localization of GPR101 with GH was seen in Panels G and H. Nuclei (blue) were counterstained with DAPI.
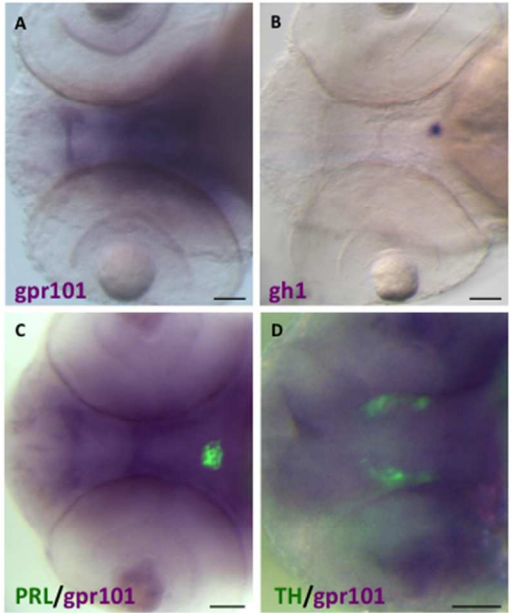254x190mm (72 x 72 DPI)

Immunolocalization of GPR101 in coronal sections of rat embryo forebrain. A) GPR101 immunoreactive cells and processes were pronounced in accumbens subventricular zone (SVZ) and rhinencephalic differential field (Rh). B) Strong GPR101 immunoreactivity was observed in striatal neuroepithelium and striatal subventricular zone (SVZ) around lateral ventricle (LV) and striatal differentiating field. GPR101 positive cells and cell processes were also observed in the septal area. C) In more posterior sections, GPR101 immunoreactivity was observed throughout the hippocampus and lateral to third ventricle (3V), in the anterior hypothalamus. GPR101 positive cells and processes were also noticed in the amygdaloid differentiating field (A). Scale bar applies to all images: 0.5 mm.
254x190mm (72 x 72 DPI)
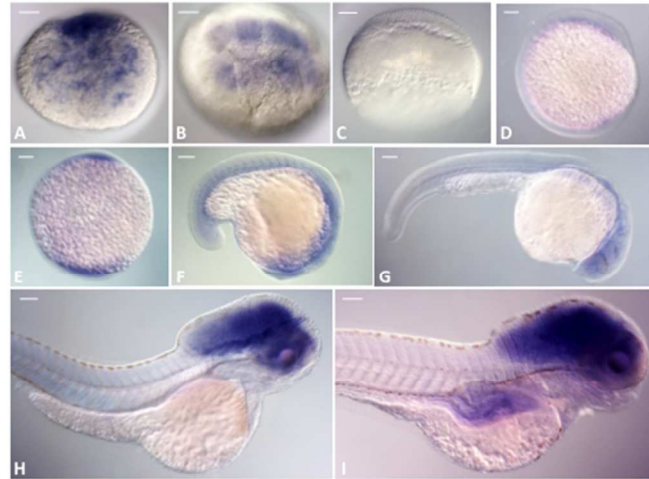
In situ staining for gpr101 during WT zebrafish development (lateral view). A) 1-cell, B) 8-cell, C) 30% epiboly, D) bud, E) 3 somite, F) 20 somite, G) 24hpf, H) 2dpf, I) 5dpf. gpr101 mRNA (blue) is visible from 1-cell to 8-cell stage (A, B), probably representing maternal transcripts, and then reappears during the bud stage (D), approximately 10hpf. Starting at 2dpf (H) a strong and brain-specific staining is seen. Scale bar: 100 μm; hpf: hours post-fertilization; dpf: days post-fertilization.

254x190mm (72 x 72 DPI)

gpr101 expression in the brain of WT Zebrafish embryo. A) and B) WISH staining for gpr101 and gh1 (purple) in 2dpf distinct Zebrafish embryos (ventral view). gpr101 is strongly expressed in the whole brain region (A), while a pituitary-specific staining is visible with the gh1 probe (B). C) and D) Double WISH/immunofluorescence staining for gpr101 (purple) and a pituitary (PRL, green, panel C) and hypothalamic marker [Tyrosine hydroxylase (TH) specific for diencephalic dopaminergic neurons, green, panel D] in a 2dpf Zebrafish embryo (ventral view). gpr101 staining overlaps with the location of pituitary and hypothalamic neurons. The qualitatively different aspect of gpr101 staining in panels C and D is due to the different focal planes selected for PRL and TH. Scale bar: 50 μm applies to all panels.
254x190mm (72 x 72 DPI)