Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Trivellin G, Daly AF, Faucz FR, et al. Gigantism and acromegaly due to Xq26 microduplications and *GPR101* mutation. N Engl J Med. DOI: 10.1056/NEJMoa1408028

Supplemental Material

Table of Contents

Methods

Array comparative genomic hybridization (aCGH)

In the gigantism population, all 43 patients underwent aCGH. DNA extracted from peripheral blood leukocytes underwent aCGH analysis using the following commercial arrays: 8x60K (G4827A-031746), and 4x180K (G4890A-029830) (Agilent Technologies, Santa Clara, CA, USA) according to the manufacturer's instructions. The arrays were scanned with a G2565CA microarray scanner (Agilent Technologies, Santa Clara, CA, USA) and the images were extracted and analyzed with CytoGenomics software v2.0 (Agilent Technologies, Santa Clara, CA, USA). An ADM-2 algorithm (cut-off 6.0), followed by a filter to select regions with three or more adjacent probes and a minimum average log2 ratio ±0.25, was used to detect copy number changes. The quality of each experiment was assessed by the measurement of the derivative log ratio spread with CytoGenomics software v2.0. Genomic positions were based on the UCSC February 2009 human reference sequence (hg19) (NCBI build 37 reference sequence assembly). Filtering of copy number changes was carried out using the BENCHlab CNV software (Cartagenia, Leuven, Belgium).

In order to precisely determine the sizes, genomic boundaries and contents of the rearrangements in each individual, we designed an 8x60K format aCGH (Agilent Technologies) with high-density probes tiling the critical region inside Xq26.3 (ChrX: 135001882-136499429, hg19). The probe density averaged five oligonucleotides/Kb for the critical region with copy number changes. It also interrogates the flanking genomic regions of up to 2 Mb in size with probe density of 1-2 per Kb. The experimental procedures of aCGH, including DNA fragmentation, labeling and hybridization, are performed by following the protocols described previously[1] with minor modifications.

Breakpoint junction analysis

Long-range PCR was performed to amplify the breakpoint junctions. The forward and reverse primers were designed by using the sequences from the estimated boundaries of the rearrangements defined by the aCGH. TaKaRa *LA Taq* (Clontech, Mounatin View, CA, US) was used for the PCR amplifications. Sanger sequencing was performed for the PCR products, and the DNA sequences were compared to the reference genome (hg19) in order to map the breakpoint junctions.

Copy Number Variation (CNV) analysis

Individual CNV assays were performed by duplex TaqMan real-time PCR assays in order to confirm the array-CGH results in the giant patients and to extend the analysis in a cohort of 47 patients with sporadic pituitary tumors for which sufficient-quality DNA was available. CNV assays for *CD40LG*, *ARHGEF6*, *RBMX*, and *GPR101*, consisting in a pair of unlabeled primers and a FAM-labeled MGB probe, were supplied from Life Technologies (Assay ID: Hs02425845_cn, Hs01655699_cn, Hs01064297_cn, Hs01730605_cn, respectively). *RNase P* (Life Technologies, #4403328) with a VIC-labeled TAMRA probe was used as reference gene. TaqMan CMV assays were performed according to manufacturer's protocol (Life Technologies, Carlsbad, CA, USA).[2] Briefly, experiments were prepared in 96 microwell plates and consisted of 20 µl reactions containing 20 ng of genomic DNA, 10 µl TaqMan Genotyping Master Mix (Life Technologies, catalog number 4371355) and 1 µl each of one target gene and reference CNV assay mixes. All reactions were run in triplicate on a ViiA 7 Real-Time PCR System (Life Technologies) and thermal cycling conditions were 95°C, 10 min followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. All data were analyzed using the CopyCaller software version 2.0 (Life Technologies). The copy number was determined by the relative relationship between the quantity of the target gene and the reference gene, using a female calibrator sample known to have two copies of each gene as the basis for comparison.

Whole exome analysis (WES)

Thirteen FIPA kindreds with homogeneous acromegaly (2-3 affected patients per kindred) were selected for WES and one affected individual per kindred was studied. Five patients from the gigantism population also underwent WES. Genomic DNA was extracted from 400µl whole blood using the BioRobot M48 (Qiagen). Exome libraries were prepared from high quality DNA following vendors' protocols (TruSeq and Nextera exome capturing kits, Illumina with, respectively, 1µg and 50ng DNA). Exome libraries were multiplexed by 24 and paired-end sequenced on two lanes of an Illumina HiSeq reaching a mean 30x coverage depth. For each sample, reads were mapped and variants were called following the GATK Best Practices (GATK v3.1). Unless stated otherwise, steps were performed with GATK tools (v3.1). Briefly, paired-end reads were mapped to the reference human genome (UCSC hg19) with bwa-mem (0.7.7). Duplicate reads were marked with Picard (1.73) and locally realigned with MarkDuplicates. Base quality scores were recalibrated with BaseRecalibrator. Variants were called on each processed sample with the HaplotypeCaller in gVCF mode followed by a joint GenotypeGVCFs step including 1000Genomes control samples. Variant quality scores were recalibrated with VariantRecalibrator and annotated with Variant Effect Predictor (Ensembl, release 75). Variants falling in a 99% truth sensitivity level were retained.

Fluorescent *in situ* Hybridization (FISH) analysis

FISH analysis was performed for confirmation of the array data. Commercially available probes covering the region of chromosome X provided insufficient coverage and hence new fluorescent probes were designed in collaboration with the commercial partner (Agilent). The red and green probes were designed based on the continuous Xq26.3 duplication intervals of two of the sporadic cases, S1 and S2 (chrx:135620070-136173879, chrx:135617178-136250554) and the two separate duplicated intervals of case S4 (chrx:135624323-135985727; chrx:136 045 379-136 268 105). This provided coverage of the two SROs identified in the Xq26.3 duplicated cases. The nuclei preparation was performed as described previously.[3,4] Analysis was performed with a BX51 epifluorescence microscope (Olympus) equipped with a CV-M4 + CL camera (JAI) and images were captured using the

platform CYTOVISION version 7.3.1 (Leica Microsystems).

DNA preparation and sequencing analysis

All patients with Xq26.3 microduplication, 37 pediatric sporadic patients from the NIH (32 ACTH-secreting and 5 GH-secreting adenomas without a history of FIPA or other inherited pituitary adenoma conditions, described in[5]), and 96 sporadic acromegaly patients (55% males; median age at diagnosis: 38.5 years) from the University of Liège (n=88) and the NIH (n=8) were sequenced for the four protein-coding genes (*CD40LG*, *ARHGEF6*, *RBMX*, and *GPR101*), the small nucleolar RNA, *SNORD61*, and the microRNA, *miR-934*, located in the duplicated Xq26.3 region (Table S2). In addition, the *GPR101* gene was sequenced in a large international cohort of sporadic acromegaly patients and somatic somatotropinoma samples from the NIH, European centers, and a group of samples (reported in[6]) of varied geographic origin. Among these sporadic acromegaly cases, 11 patients had paired genomic and somatic DNA available.

DNA was extracted from peripheral EDTA blood and pituitary tumor samples using the QIAamp DNA Mini Kit (Qiagen, Valencia, CA) according to the manufacturer's protocols. The whole coding region, intron–exon boundaries, and 5'- and 3'-UTRs of all five target genes were PCR amplified and directly sequenced. Primers sequences and PCR conditions are available upon request. DNA sequencing was performed using the BigDye 3.1 Termination Chemistry (Life Technologies) on a Genetic Sequencer ABI3100 apparatus (Life Technologies). Sequences were visualized and aligned to the corresponding wt reference sequence using the SeqMan Pro software (DNAStar, Madison, WI). All variants have been annotated according to Human Genome Variation Society (HGVS) recommendations (www.hgvs.org/mutnomen). The following reference sequences were used: NM_016267.3 for miR-934 (*VGLL1*), NM_000074.2 for *CD40LG*, NM_004840.2 for *ARHGEF6*, NM_002139.3 for *RBMX*, NR_002735.1 for *SNORD61*, and NM_054021.1 for *GPR101*.

In silico analysis

The allelic frequency of every known variant was retrieved from the Population Genetics page available in the Ensembl genome browser for each SNP. Since our patient population is heterogeneous, we compared the allelic frequency of each variant with the corresponding global allelic frequency retrieved from the public databases, including mainly the 1000 Genomes Project[7] and the Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP, http://evs.gs.washington.edu/EVS/) data.

In silico predictions were performed with PONP (Pathogenic-or-Not-Pipeline)[8] and Alamut version 2.3 (Interactive Biosoftware, Rouen, France) softwares. The interpretation of Alamut predictions for potential splice-affecting nucleotide variants was performed following the guidelines reported in *Houdayer et al.*[9] Variants were categorised into five groups according to the classification system reported by *Plon et al.*[10] Inclusion of each variant into one of these groups was done after careful assessment of all the available data[11].
In silico predictions of CpG islands in the four duplicated OMIM genes was done with the freely available online tool CpG Island Searcher (http://cpgislands.usc.edu/) by uploading the 5'UTR and 5 kb of the putative promoter region of each gene.

Immunostaining

The surgical samples were fixed in formalin and embedded in paraffin. Five-µm-thick sections were stained with haematoxylin-eosin (H&E) and reticulin for light microscopy. Immunohistochemistry for pituitary hormones was performed following previously described protocols[12] using the following primary antibodies: anti-Prolactin (PRL) AM031-5M Biogenex®, Fremont-CA, monoclonal mouse, ready to use; anti-Human Growth Hormone (GH) A0570 Dako Cytomation®, Carpinteria-CA, polyclonal rabbit, working dilution 1:2000; anti-Adrenocorticotropin (ACTH) Clone 02A3 Dako Cytomation, Carpinteria-CA, monoclonal mouse, working dilution 1:2000; anti-Follicle Stimulating Hormone (FSH), AM026-5M Biogenex, Fremont-

CA, monoclonal mouse, ready to use; anti-Luteinizing hormone (LH) AM030-5M Biogenex, Fremont-CA, monoclonal mouse, ready to use; anti-Thyroid stimulating hormone (TSH) AM033-5M Biogenex, Fremont-CA, monoclonal mouse, ready to use; anti-alpha-subunit (alpha-hCG) AM037-5M Biogenex, Fremont-CA, monoclonal mouse, ready to use.

GHRH-R, ab150575, Abcam®, Cambridge-MA, primary antibody is a rabbit polyclonal antibody mapping to the C-terminal transmembrane domain of GHRH-R. Immunohistochemistry was performed by Ventana® BenchMarkXT Automated IHC/ISH slide staining system. The histological samples were incubated for 90 min at 37° with this primary antibody diluted at 1:50, after pre-treatment with a pH9 buffer for 60 min. The Ventana® ultraView Universal DAB Detection Kit was used for the detection.

GHRH, GTX81311, Genetex®, Irvine CA, primary antibody is a rabbit polyclonal antibody mapping to the N-terminal region of GHRH. Immunohistochemistry was performed by Ventana® BenchMarkXT Automated IHC/ISH slide staining system. The histological samples were incubated x120' at 37° with this primary antibody diluted at 1:10, without pre-treatment. The Ventana® ultraView Universal DAB Detection Kit was used for detection.

Cytokeratin staining was performed with CAM5.2.

Normal pituitary, pancreatic, and gonadal tissues were used as positive controls.

Negative control reactions were obtained in each test series by omission of the primary antibody. Images from haematoxylin-eosin and reticulin stain for light microscopy and immunohistochemistry were performed with BX41 Olympus Microscope, Axiocam ICc 1 camera and Axiovision 4 software from Zeiss at 100x and 200x magnifications.

For GPR101 immunofluorescence evaluation, a subset of five human pituitary tumor tissues, a human normal pituitary tissue, and five human normal hypothalami, was used. All paraffin embedded slides were submitted to de-parafinization, rehydration

and antigen retrieval for 30 min in citrate buffer solution (pH 6.0). All slides were incubated with 10% normal donkey serum (NDS) for one hour for blocking, and they were incubated with the following primary antibodies: rabbit anti-GPR101 (dilution 1:500; SAB4503289, Sigma-Aldrich, St. Louis, MO, USA), rabbit anti-GPR101 (dilution 1:500; HPA001084, Sigma-Aldrich), goat anti-GH (dilution 1:100, sc-10364; Santa Cruz Biotechnology, Santa Cruz, CA) overnight at 4 °C. Both anti-GPR101 antibodies react with the human and rodent homologues of GPR101. All slides were incubated for 1-2 hours with the following secondary antibodies: donkey anti-rabbit 555 (A-31572, Life Technologies, Foster City, CA) and donkey anti-goat 488 (A-11055, Life Technologies), both at 1:500 dilution. Prolong gold mounting media with DAPI (P36934, Life Technologies) was used to set the slides. As negative control, a section of the same specimen was incubated under identical conditions with no primary antibody. Fluorescence was analyzed with a Leica AF6000 microscope (Leica, Allendale, NJ) at 63x magnification with fixed time of exposure for all samples. Subsequently, images were deconvoluted with the supplied Leica image processing software (Leica). The same linear adjustments for brightness, contrast and color balance have been applied with Adobe Photoshop CS6 to each entire image.

75,000 GH3 cells were seeded onto Lab-Tek II chamber slides (Nunc, Rochester, NY), left overnight and transiently transfected with WT and mutant GPR101 vectors. 24 h after transfection cells were washed 2 x in PBS and fixed in 4 % paraformaldehyde in PBS for 15 min, followed by rehydration and permeabilization in PBS containing 0.01% Tween-20. Slides were blocked in 10% goat serum (Jackson ImmunoResearch, West Grove, PA) for 1 h. Two primary antibodies against GPR101 (SAB4503289, dilution 1:500, and HPA001084, dilution 1:1000, Sigma-Aldrich) in 10% goat serum were applied to the slides at 4 °C overnight alongside negative controls without primary antibody. Slides were washed in PBS-T and secondary Alexa Fluor 488 goat anti-rabbit (dilution 1:1000, Life Technologies) was applied for 1 h. Slides were washed, mounted using Prolong Gold containing DAPI (Life Technologies) and covered. Cells were visualized at 40X magnification with a Zeiss AxioCam MRm microscope camera using the ZEN software.

Five-µm-thick coronal sections of the mouse brain tissues were deparaffinized and the heat induced antigen retrieval was done in a microwave for 5 min with buffer

containing 10 mM Tris, 1 mM EDTA, and 0.05% Tween 20, pH 9. The rabbit polyclonal anti-GPR101 antibody (SAB4503289, Sigma-Aldrich) was applied in 2 μg/ml dilution overnight at 4 ºC. Immunofluorescence staining was performed using a Tyramide signal amplification kit (T20922, Life Technologies) with horseradish peroxidase-goat anti-rabbit IgG and Alexa fluor 488 tyramide, according to the manufacturer's instructions. The sections were mounted in Mowiol and visualized under an inverted confocal microscope (Zeiss LSM 510). Brightness and contrast were adjusted in each entire image in Adobe Photoshop CS4.

RNA isolation and reverse transcription

Total RNA was isolated from PHA stimulated blood cells using the RNeasy Mini Kit (Qiagen, Germantown, MD). Cells were prepared using a modification of a technique first reported by *Nowell et al*.[13] Prior to RNA extraction, in order to reduce nonsense-mediated mRNA decay, cells were incubated with 100 μg/ml cycloheximide (Sigma-Aldrich) for 2 h at 37°C.

Total RNA was isolated from human and mouse pituitary tissues using TRIzol reagent (Life Technologies). 1 μg of RNA was treated with DNase I (Life Technologies) to remove genomic DNA contamination and then reverse transcribed to cDNA using the Superscript III Kit (Life Technologies) according to the manufacturer's protocols.

Total RNA was isolated from GH3 cells transiently transfected with human WT and mutant GPR101 vectors (see below) using the RNeasy Mini Kit (Qiagen) according to the manufacturer's protocols, and reverse transcribed as described above.

Total RNA from mouse hypothalamus and Poly A+ RNA from human hypothalamus were purchased from Clontech (catalog number 636664, 636144, respectively) and reverse transcribed as described above.

mRNA expression analysis

The expression levels of the four duplicated genes and GHRH were measured by RT-qPCR with the TaqMan system using ready-made probe-primer kits supplied from Life Technologies (Assay ID: *CD40LG*, Hs00163934_m1; *ARHGEF6*, Hs00374462_m1; *RBMX*, Hs00953944_g1; *GPR101*, Hs00369662_s1; *GHRH*, Hs00184139_m1). Reactions were performed in triplicate and *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase, Life Technologies, Hs99999905_m1) was used as endogenous control. In GH3-transfected cells beta-actin (*Actb*) was used as endogenous control (Rn00667869_m1, Life Technologies). In mouse tissues *Gpr101* expression was measured with TaqMan assay ID Mm01296083_m1 (Life Technologies) and normalized on beta-actin expression (*Actb*, Mm00607939_s1, Life Technologies). TaqMan assays were performed according to the manufacturer's protocol (Life Technologies). Briefly, experiments were prepared in 96 microwell plates and consisted of 20 µl reactions containing 20 ng of cDNA, 10 µl TaqMan Gene Expression Master Mix (Life Technologies, catalog number 4369016) and 1 µl each of one target gene and endogenous control assay mixes. All reactions were run on a ViiA 7 Real-Time PCR System (Life Technologies) and thermal cycling conditions were 95°C, 10 min followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. Relative gene expression data were analyzed using a relative standard curve using qPCR human reference cDNA, random-primed (50 ng/µl, 639654, Clontech). Data analysis was performed using the ViiA 7 software (Life Technologies).

Protein extraction and Western Blot analysis

Proteins were extracted from GH3 cells transiently transfected with WT and mutant GPR101 vectors. 24 h after transfection cells were washed with PBS, lysed with 100 µl of lysis buffer (50 mM Tris-HCl, 50mM NaCl, 10 mM EGTA, 10 mM EDTA, 80 M sodium molybdate, 5 mM sodium pyrophosphate, 1 mM sodium orthovanadate, 1 mM PMSF, 4 mM pNPP, 1%Triton; Sigma-Aldrich), sonicated, and centrifuged at 20,000 g for 15 min at 4 C. The supernatant was subsequently measured for protein content using the BCA Protein Assay Kit (Pierce Biotechnology, Rockford, IL).

Proteins were also isolated from two post-mortem samples of normal human pituitary using 350 µl of lysis buffer (see above), sonicated, and centrifuged at 20,000 g for 15

min at 4 C. Pools of tissue-specific total proteins isolated from different human brain regions (amygdala, hippocampus, hypothalamus) were purchased from Clontech (catalog number 635317, 635319, 635320, respectively).

Under denaturing conditions, 15–30 μg of protein lysates were loaded onto Bolt 10% Bis-Tris Plus gels (Life Technologies) and electroblotted onto Whatman Protran nitrocellulose membranes (Sigma-Aldrich). Blots were blocked for 30 min in 5% non-fat dry milk in TBST 0.1% and incubated overnight with primary antibodies (GPR101 (SAB4503289, dilution 1:500, Sigma-Aldrich), and GAPDH (sc-137179, dilution 1:400, Santa Cruz)). Subsequently, blots were washed with TBST 0.1% (3 x 15 min) and incubated in goat anti-rabbit IgG horseradish peroxidase-conjugated secondary antibody (dilution 1:2000, Jackson ImmunoResearch) for 1 h at room temperature. Immunoblots were visualized using the Supersignal West Pico Chemiluminescent substrate (Thermo Scientific, Waltham, MA) on a ChemiDoc MP imaging system (Bio-Rad, Hercules, CA), and quantified using the Image Lab software (Bio-Rad). Expression was normalized to GAPDH.

Cell culture

The rat pituitary somatomammotroph GH3 cell line was grown in Dulbecco's modified Eagle's medium (DMEM, high glucose, pyruvate, no glutamine; 10313, Life Technologies) supplemented with 10% fetal bovine serum (100-106, Gemini Bio-Products), and 1% antibiotic-antimycotic (15240-062, Life Technologies) in a humidified atmosphere at 37°C with 5% CO_2.

GH secretion after ARHGEF6, RBMX, and GPR101 over-expression

GH3 cells were seeded in 12-well plates at a density of 2 x 10^5 cells/well. After 24 h, cells were starved with DMEM without serum for 16 h and then transfected with Lipofectamine 2000 (11668030, Life Technologies) according to the manufacturer's protocol, using Opti-MEM I Reduced Serum Medium (31985-070, Life Technologies) and 1 μg of each vector, alone or in combination. The human GPR101 wt (NM_054021.1) and ARHGEF6 wt (NM_004840) coding sequences cloned into

the pCMV-XL5 vector were purchased from Origene (SC120214 and SC100230, respectively), while the human wt RBMX coding sequence (NM_002139) cloned into the pcDNA3.1/V5-His TOPO vector (Life Technologies) was a kind gift of Dr SJ Levine (NIH, Bethesda, USA). The p.E308D and p.A397K variants were introduced into the human GPR101 wt template using the QuikChange Lightning site-directed mutagenesis kit (210518-5, Agilent Technologies), following the manufacturer's protocol. Primer sequences are available on request. The empty pCMV-XL5 and pcDNA3.1/V5-His TOPO vectors were used as negative controls, accordingly. 24 h after transfection supernatants were collected and GH secretion was measured using the Rat/Mouse Growth Hormone ELISA kit (EZRMGH-45K, EMD Millipore) following the manufacturer's protocol. Absorbance was read at 450 nm and 590 nm using a FLUOstar Omega microplate reader (BMG LABTECH). GH secretion values (expressed in ng/ml) were normalized either on protein content or on O.D. values obtained for the same samples from cell proliferation experiments (see section below).

Cell proliferation after ARHGEF6, RBMX, and GPR101 over-expression

GH3 cells were seeded in 96-well plates at a density of 2 x 10^4 cells/well. After 24 h, cells were starved with DMEM without serum for 16 h and then transfected with Lipofectamine 2000 (Life Technologies) according to the manufacturer's protocol, using Opti-MEM I Reduced Serum Medium (Life Technologies) and 125 ng of each vector (human WT ARHGEF6, WT RBMX, WT GPR101, p.E308D GPR101, and p.A397K GPR101), alone or in combination. The empty pCMV-XL5 and pcDNA3.1/V5-His TOPO vectors were used as negative controls. 24h after transfection cell viability was assessed with the Vybrant MTT cell proliferation assay (Life Technologies) following the manufacturer's protocol. Briefly, cells were incubated for 4 h at 37 °C with MTT solution and were then lysed with Dimethyl sulfoxide (DMSO). Absorbance was read at 540 nm using a FLUOstar Omega microplate reader (BMG LABTECH).

Use of a reporter assay to monitor cAMP levels after over-expression of WT and mutant GPR101 constructs

GH3 cells were seeded in 12-well plates at a density of 2 x 10^5 cells/well. After 24 h, cells were starved with DMEM without serum for 16 h and then transfected with Lipofectamine 2000 (11668030, Life Technologies) according to the manufacturer's protocol, using Opti-MEM I Reduced Serum Medium (31985-070, Life Technologies), 1 μg of each GPR101 vector (human WT GPR101, p.E308D GPR101, and p.A397K GPR101), 800 ng of pGL4.29[luc2P/CRE/Hygro] vector containing a cAMP response element (CRE) that drives the transcription of the luciferase reporter gene (Promega), and 40 ng of the Renilla vector (pRL-SV40, Promega). The empty pCMV-XL5 vector was used as negative control. 24h after transfection, a subset of cells was treated with 10 μM forskolin (F6886, Sigma-Aldrich) for 1 h and then lysed. Firefly and Renilla luciferase activities were measured consecutively in the same sample using the Dual-Luciferase Reporter Assay System (E1910, Promega) following the manufacturer's protocol. Ratios of Firefly vs. Renilla luminescence signals, serving as a measure for reporter activity normalized for transfection efficiency, were measured using a FLUOstar Omega microplate reader (BMG LABTECH).

X- chromosome inactivation (XCI) analysis

Inactivation status of the X-chromosomes was determined on lymphocyte genomic DNA using the *androgen receptor* (*AR*), *SLITRK4, PCSK1N,* and *FMR1* methylation assays. Among the patients with the Xq26 duplication six of the eight female sporadic and a familial case (F1A) were tested. DNA was first digested with the methylation-sensitive HpaII restriction endonuclease, purified using QIAquick spin columns (Qiagen), and PCR amplified for *AR, SLITRK4, PCSK1N,* and *FMR1*.[14] The X-chromosome inactivation dosage analysis was calculated as described in.[15] Skewed alleles were considered when more than 80 % of the investigated cells inactivated the same chromosome.

Molecular modeling procedures

Homology models of the human GPR101 receptor were constructed using the structure of the β_2 adrenergic receptor (β_2 AR) crystallized in its activated state in complex with a G_s protein heterotrimer (PDB ID: 3SN6)[16], according to the sequence alignment shown in Figure S13. The template was chosen because, among the crystallographically solved GPCRs, the β_2 AR is the most closely related to GPR101, with 27.5 % amino acid identity within the seven transmembrane domains. Moreover, the presence of a G protein heterotrimer in the 3SN6 structure allowed us to model the putative interactions of GPR101 with the G protein.

Ten homology models of the wild-type GPR101 receptor were constructed with the automodel function and ranked on the basis of the DOPE energy score. For the top scoring models, 100 alternative models were then constructed with the loopmodel function of Modeller and ranked on the basis of the DOPE energy score. The formation of a disulfide bridge between the Cys residues at positions 104 (third transmembrane domain) and 182 (second extracellular loop) was enforced. The modeling parameters were set as follows: deviation = 4; max_var_iterations = 300; md_level = refine.slow.

Two *GPR101* mutations were modelled: one identified in this report (p. E308D) and one previously studied (p. A397K).[17] The p.E08D point mutation was introduced on the model of the wild-type receptor with the "mutate" function of Modeller. Subsequently, the mutated residue (as well as all the residues with at least one atom within a 10 Ang radius from it) was refined with 1 ns of molecular dynamics at a temperature of 1000 K followed by a 20 ps cooling phase that brought the temperature down to 300 K. The molecular dynamics was performed with the CHARMM-22 force field[18] and with a time step of 4 fs.

Statistical analysis

Statistical analysis was performed with StatsDirect software (Addison-Wesley-Longman, Cambridge, UK).

Data are presented as the mean ± standard deviation (SD) of two to five independent experiments, each performed at least in triplicate. Comparisons were calculated using a two-tailed Student's t test for unpaired data and the Kruskal-Wallis test followed by the Conover-Inman test, as appropriate. A Chi-square test was used to compare the allelic frequencies of the different genes in patients and controls. The data were considered to be significant when $P < 0.05$.

CNV databases query

The dbVar (http://www.ncbi.nlm.nih.gov/dbvar/) and DECIPHER v7.0 (http://decipher.sanger.ac.uk/) databases were queried for deletions of the Xq26 region associated with short stature and developmental delay. All the cases retrieved are reported in Table S4.

Literature review

A review of published gigantism cases was performed searching the PubMed database using the keywords "giant", "gigantism", "acromegaly", "pediatric", "infantile", "childhood-onset", "overgrowth", somatotropinoma, "pituitary adenoma". To focus the data collection on definitive cases of GH hypersecretion, only studies in which GH measurements were reported were considered. Other inclusion criteria for analysis were the presence of a defined pituitary abnormality on imaging or surgery, data on the age of onset of disease or age at diagnosis, estimation of the tumor size on imaging or surgery. We excluded cases that could be explained as potentially mutated at the *AIP* gene (as the onset of gigantism in these patients occur during adolescence), and those that had MAS, MEN1 or other syndromes. Data on patient characteristics at birth and during childhood, hormonal secretion at baseline and following dynamic testing, tumor dimensions, surgical findings and pathology findings were extracted, collated and summarized statistically as medians and ranges. A total of 15 cases comprised the literature review dataset. [19-32,43]

Results

CNV analysis

Six patients harboring an Xq26.3 microduplications and 47 patients with sporadic acromegaly and pediatric pituitary tumors were analyzed for CNV by duplex TaqMan real-time PCR assays for *CD40LG*, *ARHGEF6*, *RBMX*, and *GPR101*. CNV assays confirmed the aCGH results in the patients with the duplication (Figure S1), whereas no gain or loss of genetic material was found in any gene in any sporadic acromegaly and pediatric patients. CNV analysis in the tumor DNA of two patients harboring the Xq26.3 duplication confirmed the duplication of all 4 genes in the pituitary tumor tissue.

FISH analysis

The specially designed fluorescent probes covering regions within SRO1 (green) and SRO2 (red) were applied to cultured leukocytes from four of the sporadic cases S1, S2, S4 and S9 (3 females and 1 male). In all four cases a duplication of the two probe signals was seen; a normal single copy of each probe signal was seen in the non-involved X chromosome of the 3 female cases. A representative image of one of the stained nuclei from a female sporadic gigantism case is shown in Figure S2.

High-density array-comparative genomic hybridization (HD-aCGH), breakpoint junction analysis, and potential mechanisms of the genomic rearrangements

The critical region with genomic gains shared amongst the unrelated patients was previously identified by low-resolution whole-genome aCGH. Custom designed HD-aCGH interrogating the critical region delineated 10 different genomic duplications in 12 patients, including four familial and eight sporadic cases (Figure S3). The DNA of patient S3 was not of sufficient quality to be analyzed by HD-aCGH. The remaining 29 patients that did not shown any duplication in the Xq26.3 region by low-resolution

whole-genome aCGH were also investigated in order to exclude an undetectable microduplication. HD-aCGH confirmed that these patients do not harbor any duplication in the studied region.

In the patients S1, S6, S7, S8, and S9, various sized microhomologies (TTCAGCCTTCCA for S1, CT for S6, ATG for S7, AGG for S8, and AGGGXTT for S9) were observed at the breakpoint junctions, suggesting the FoSTeS/MMBIR (Fork Stalling and Template Switching/Microhomology-Mediated Break-Induced Replication) as the potential mechanism for the formation of this tandem duplications.[33,34] Interestingly, one base-pair mismatch (C/G) was observed in the microhomology of S9, reflecting the possibility that a perfect match may not be always required for the microhomology to mediate FoSTeS/MMBIR mechanism.

For case S5, a short insertion was observed (GTGAA). The FoSTeS/MMBIR may act as the mechanism for generating such complexity by switching the template twice.

In the duplication of case S2, a 5 bp microhomology (AAGCA) was observed at the breakpoint of the tandem duplication, while further sequencing revealed a 149 bp insertion that possibly arose from a copy of a template positioned on the reverse strand 122 Mb away (ChrX: 12865862 – 12866010, hg19). Flanking that 149 bp insertion were two 1 bp micro-homologies (G and A) (Figure S3B). A similar level of complexity was apparent for the rearrangement in patient F2A. A 2 bp microhomology was observed at the breakpoint of the tandem duplication, with a 105 bp deletion and then a 1 bp insertion at the same place of the deletion. Non-homologous end-joining (NHEJ) or FoSTeS/MMBIR may act as the deletion/insertion mechanism (Figure S3B).

More complex abnormalities were seen in the duplication of patient S4: HD-aCGH detected proximal and distal duplicated segments separated by a normal segment (this complexity was not seen on the whole genome low density array which simply showed a single duplicated region). Two major breakpoint junction clusters, S4-1 and S4-2, were revealed. A 3 bp insertion was observed at S4-1 and an insertion of 86 bp at the proximal S4-2, which could partially derive from mismatches in the vicinity (e.g. a 7 bp insertion (ACATGAG) and a 9 bp insertion (CATGAGGTA)). This 86 bp

insertion also may provide the 7 bp microhomology (GCCTCCT) for priming the replication to the distal end of S4-2.

PCRs of the breakpoint junctions specific for the patients were performed to confirm the inheritance pattern of the genomic gains. Breakpoint junction PCRs for the families of S2, S4 and S6 suggested the gains to be *de novo* events (Figure S5A). The paternal sample was not available for S5. The breakpoint junction PCRs were negative in the mother and the unaffected brother. Breakpoint junction PCRs for the familial cases F1A, F1B and F1C revealed that the duplications in F1B and F1C were identical and inherited from the mother F1A. PCR was negative in the paternal DNA. For the other familial case F2A, the breakpoint junction PCR was negative for the unaffected sister of the proband (Figure S5B).

In total, 21 potential template-switching events occurred in all the genomic gains, and microhomologies were involved in nine of them. We also observed six small insertions (<10 bp), which may account for up to 12 template switches (2 template switches being required for each small insertion). Although the small insertions (less than 10 bp) were too small to be uniquely located in the genome, it is possible that there were microhomologies flanking the insertions, mimicking the mechanism of the insertion observed in S2. The small insertions could have also been *de novo* synthesized, rather than being template events. Five out of ten genomic gains showed multiple template switches, which introduced small-scale complexities near the large genomic gain breakpoints, reflecting the potential low-processivity replication repair and iterative template switches after the collapse of the replication forks.[35]

Immunostaining

Immunostaining for GPR101 (red) and GH (green) was performed in five patients with the Xq26.3 microduplication and in controls (an age-matched control somatotropinoma without the duplication and a normal pituitary). A higher GPR101 expression was observed in patients harboring the Xq26 duplication compared to both controls (Figure S7).

Immunostaining for GPR101 was also performed in the mouse and human normal brain (Figure S11). GPR101 expression was observed in the mouse hypothalamic area around the third ventricle (3V), including the arcuate nucleus (ARH). In both the mouse and human ARH, GPR101 was detected on neuronal cell bodies and axons.

Haematoxylin-eosin (H&E) and reticulin staining were performed in all cases (Figure 3 and Figure S15). In general, cases were of the mixed-type of GH and prolactin secreting adenoma (WHO 2004 Classification). Tumor cells showed eosinophilia and were organized in solid pattern with some cystic structures and psammomas. The tissue architecture was characterized by expended hyperplastic acini with intact reticulin fiber network compared to normal adenohypophysis. Some areas with enlarged, hyperplastic acini showed an initial breakdown of some reticulin fibers, resulting in confluent acini (transformation areas). Other zones presented clearly a partial or total disruption of reticulin fiber network with pseudo-nodular formations (microadenoma). In patient F1C the normal pituitary tissue was extensively substituted by mammosomatotroph hyperplasia with nodular pattern. There were small foci where the enlarged acini of the hyperplasia become confluent. GHRH-R was expressed in GH-secreting cells in normal adenohypophysis, but at low intensity. In contrast, GHRH-R immunostaining was positive in hyperplastic areas and in adenomas with intensity stronger than normal adenohypophysis (Figure S15). GHRH staining was performed and was low/absent in tumor or hyperplasia from patients with duplication, similar to that of normal pituitary tissue (data not shown).

Sequencing analysis

Several nucleotide changes were identified in the four protein-coding genes located within the common duplicated area in sporadic acromegaly and pediatric patients and are reported in Tables S1 and S2. The duplicated region contains also a small nucleolar RNA, *SNORD61*, and a microRNA, miR-934. However, no pathogenic mutations were observed in our patient population (Table S2).

RT-qPCR analysis

The transcriptional levels of *CD40LG*, *ARHGEF6*, *RBMX*, and *GPR101* were evaluated on peripheral RNA obtained from two patients with Xq26.3 micro duplications and compared with three normal subjects. *CD40LG* levels do not differ between patients and controls, whereas the mRNA levels of *ARHGEF6*, *RBMX*, and *GPR101* are lower in peripheral RNA in the patients; this was particularly evident for *GPR101* (Figure S6).

The transcriptional levels of GHRH were evaluated in two patients with the Xq26.3 duplication and compared with two normal pituitaries and two sporadic GH secreting adenomas. None of the analyzed samples showed expression of GHRH (data not shown).

GH secretion and cell proliferation analysis after over-expression of the duplicated genes

The three human genes that were expressed in the pituitary tumors (*ARHGEF6*, *RBMX*, and *GPR101*) were transiently over-expressed in the GH3 mammosomatotroph cell line. None of the over-expressed human WT proteins significantly affected cell proliferation (data not shown) or GH secretion when singularly transfected compared to cells transfected with the corresponding empty vector (Figure S8A). When human *GPR101* was overexpressed together with either or both human *ARHGEF6* and human *RBMX*, a synergistic effect was shown on cell proliferation (Figure S8C) but not on GH secretion (Figure S8B).

XCI analysis

The X-inactivation pattern in our patients with the Xq26.3 duplication was random in the sporadic cases and skewed in a familial case (F1A), with the duplicated allele being preferentially inactivated in 85% of the cells (Figure S9), as reported in many other disorders.[36]

While this could suggest different methylation triggers in sporadic versus inherited cases[37], the identical clinical features of patients with skewed and random XCI tends to exclude this hypothesis. It may be possible that given the well known monoclonal origin of pituitary tumors[38], the tumor in this patient originated from a cell that has the wild-type allele inactivated. Unfortunately the lack of available tumor did not allow us to determine the level of expression of *GPR101* in the pituitary in that patient.

CpG islands prediction

The possible contribution of each gene to the final phenotype in females could depend on their X-chromosome-related inactivation status. *RBMX* was demonstrated and *GPR101* predicted to be subject to inactivation, whereas both *CD40LG* and *ARHGEF6* were shown to undergo variable escape.[39-42] Accordingly, one CpG island was predicted for both *RBMX* and *GPR101*, whereas no CpG islands were found in *CD40LG* and *ARHGEF6* (Figure S10).

GPR101 expression analysis in normal mouse and human pituitary and hypothalamus

Gpr101 mRNA is expressed at much higher levels in mouse hypothalamus compared to the pituitary, whereas in humans *GPR101* is expressed at similar levels in both tissues (Figure S12A).

GPR101 protein expression in normal human pituitary, amygdala, hippocampus, and hypothalamus was analyzed by Western Blot (Figure S12B). GPR101 expression was detected in all the three brain regions analyzed, whereas only one out of two normal pituitary samples showed low expression levels (the other showed no expression).

GPR101 over-expression analysis in transfected GH3 cells

GH3 cells transiently transfected with the human WT and mutant *GPR101* were analyzed both at the mRNA and protein level for the achieved levels of expression of

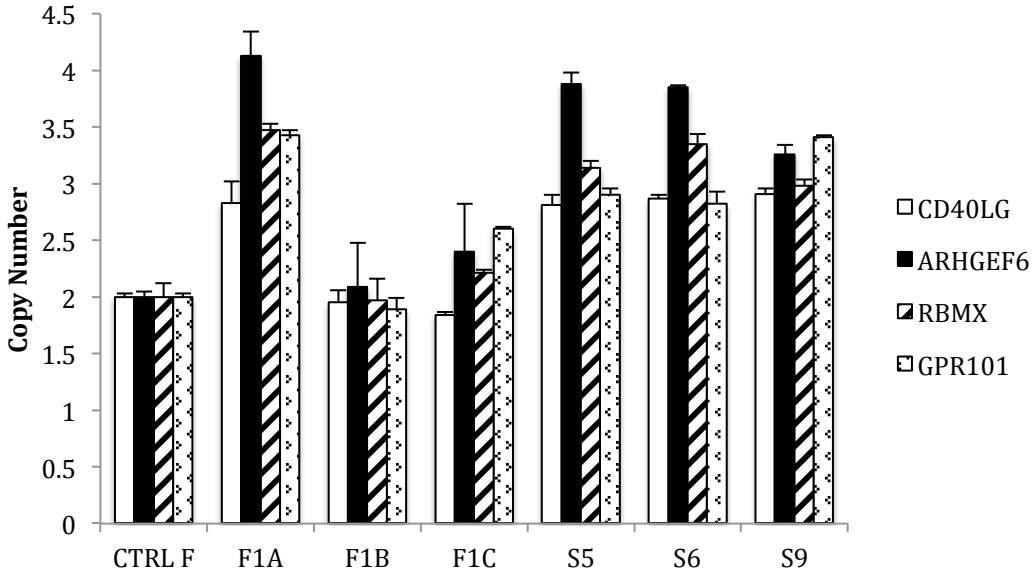
the receptor (Figure S14). Human *GPR101* expression levels are comparable to those observed in the patients with Xq26.3 microduplications.

Literature review of published gigantism cases

Like our patients, these cases were born at full-term and had normal proportions at birth. Excessive growth began generally before the age of two and at diagnosis; both height and weight were markedly increased. Pituitary pathology also showed a spectrum of adenomas and hyperplasia involving mixed GH/prolactin tumors, somatotrophs, lactotrophs, and some well-characterized instances of mammosomatotroph hyperplasia.[24, 25]

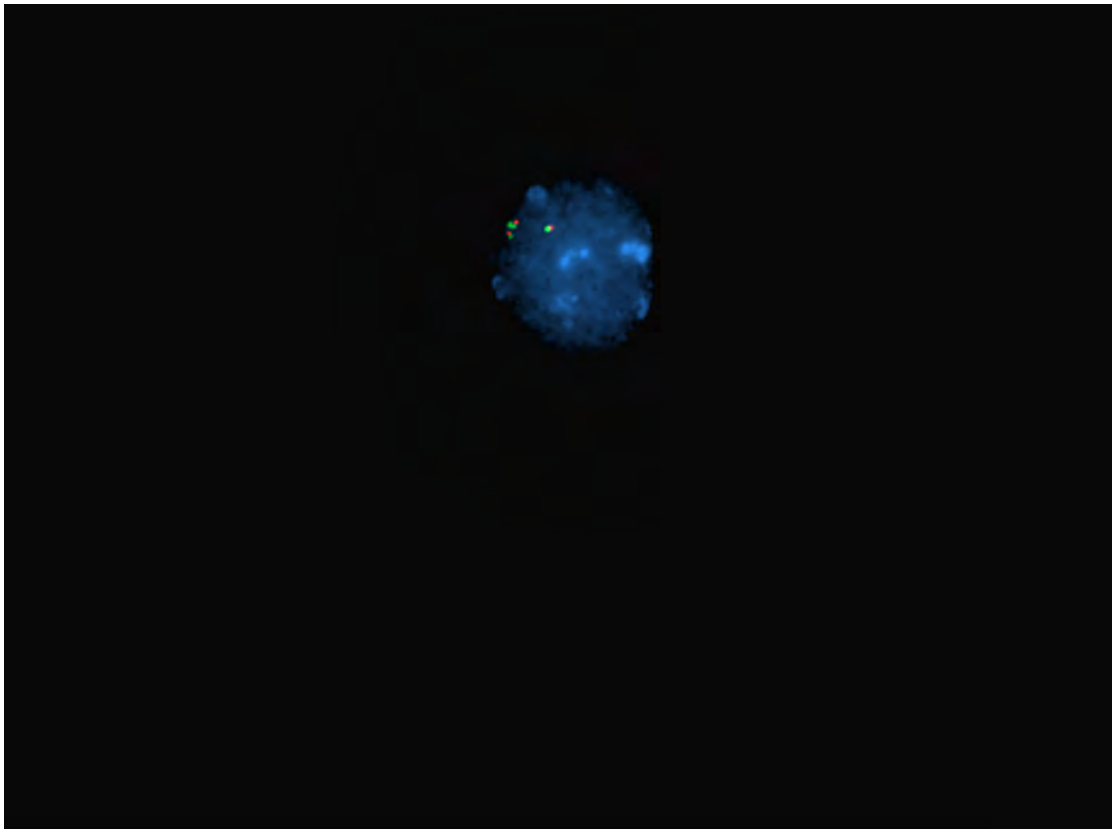
Supplemental Figures

Figure S1: CNV assays for *CD40LG*, *ARHGEF6*, *RBMX*, and *GPR101* to confirm Xq26.3 microduplications.



Six affected patients, three familial and three sporadic, were screened. CTRL F: healthy female control with two copies of the X chromosome. Data are expressed as mean ± SD.

Figure S2: Representative fluorescent image of an interphase nucleus from a female patient with childhood onset gigantism and an Xq26.3 microduplication.

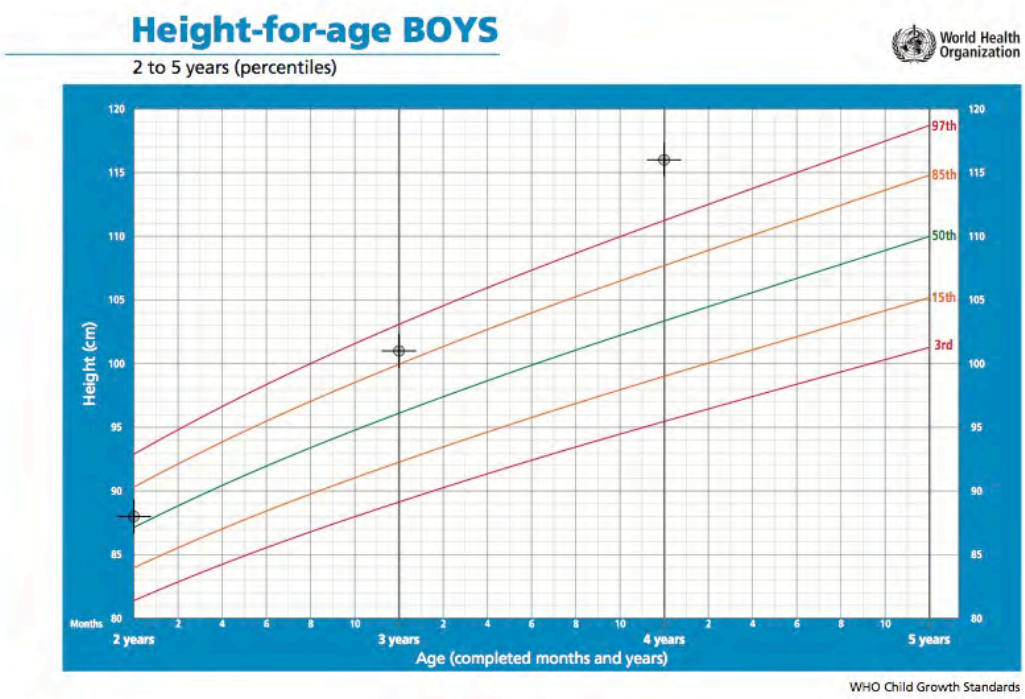


In the image a duplicated signal is seen for both the green probe (*ARHGEF6*) and the red probe (*GPR101*) on one X chromosome, while the other X chromosome gives a normal signal of one copy for both the green and red probes.

Figure S3: Summary of the aCGH data, breakpoint junctions, and potential mechanisms of the genomic rearrangements.

A) aCGH $\log_2$ ratio plots showing the duplications identified in different individuals. The clusters of red dots represent the copy number gains relative to the control. Genders (sporadic cases) and pedigrees (familial cases) are shown on the left of the aCGH plots. The breakpoint junctions are shown underneath each aCGH plot. Microhomologies (purple), insertions (green) and mismatches (grey) are illustrated in the sequences. The genomic coordinates of the proximal and distal junctions are also pointed out. The names of the repetitive elements are in the parentheses after the coordinates if the junctions reside in them. **B)** Schematic diagram showing the potential mechanisms for the genomic arrangements. Duplicated genomic segments (large colored arrows), large insertions (colored triangles), small insertions (colored letters), large deletion (dash box), and microhomologies (purple dots) are illustrated. The occurrences of template switching (FoSTeS #) are also indicated in the diagram. FoSTeS: fork stalling and template switching.

Figure S4: Extended growth chart data for patient F1C shown in Figure 1 panel A in the main text showing the abnormally increased height and weight from 2 to 4 years of age.



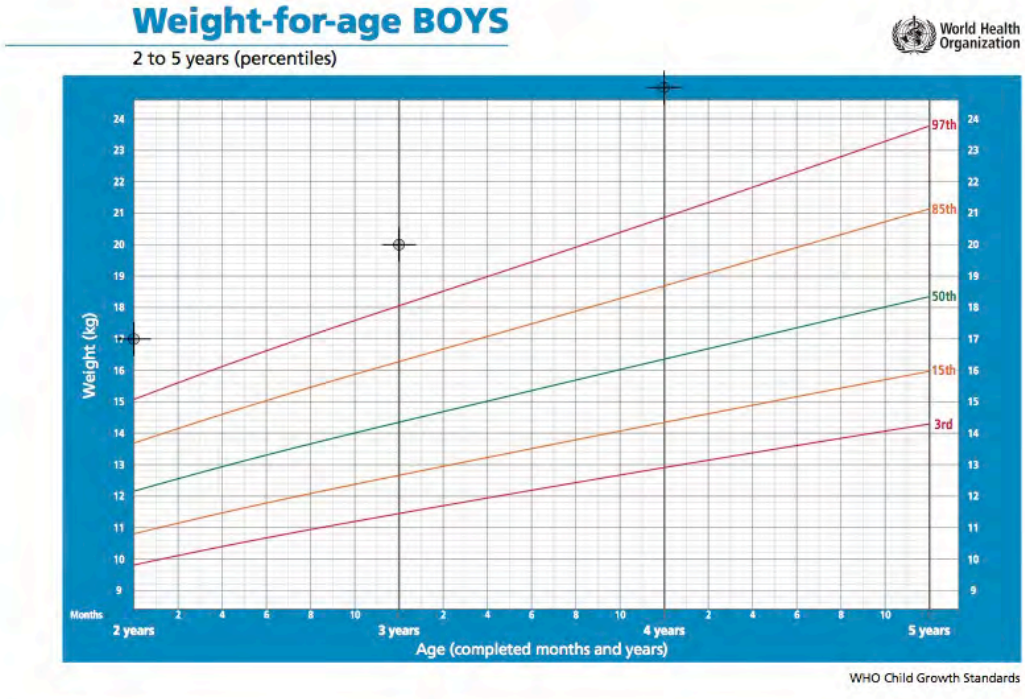
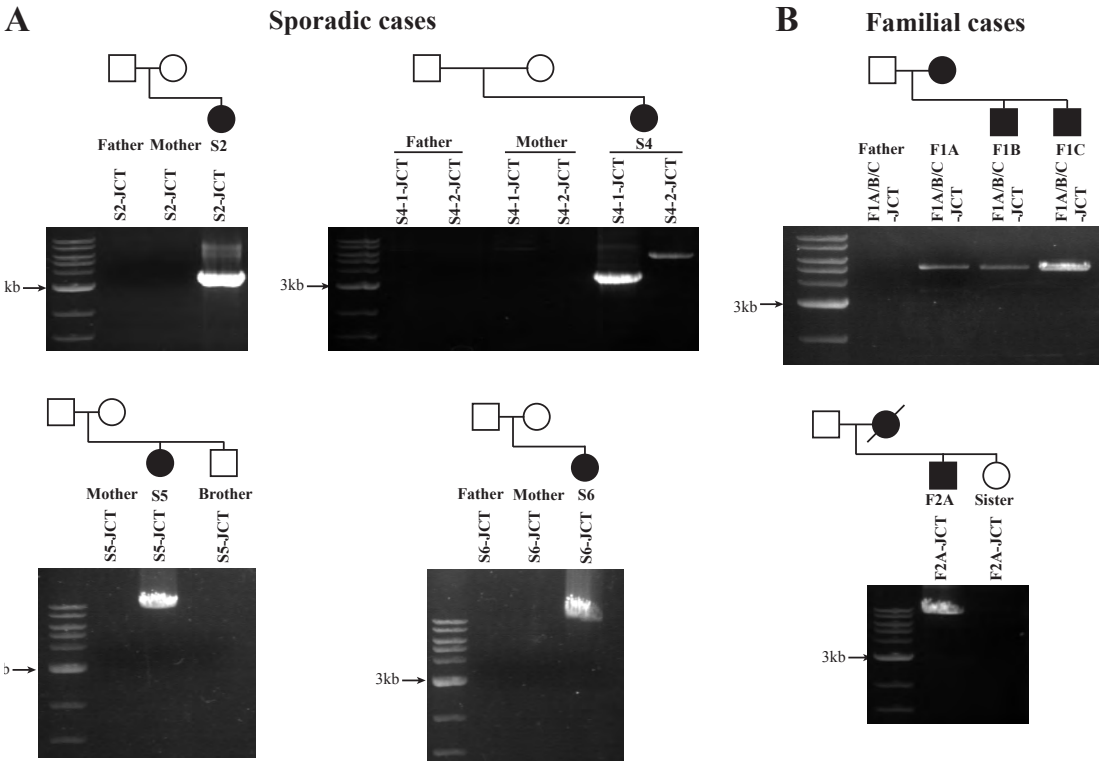
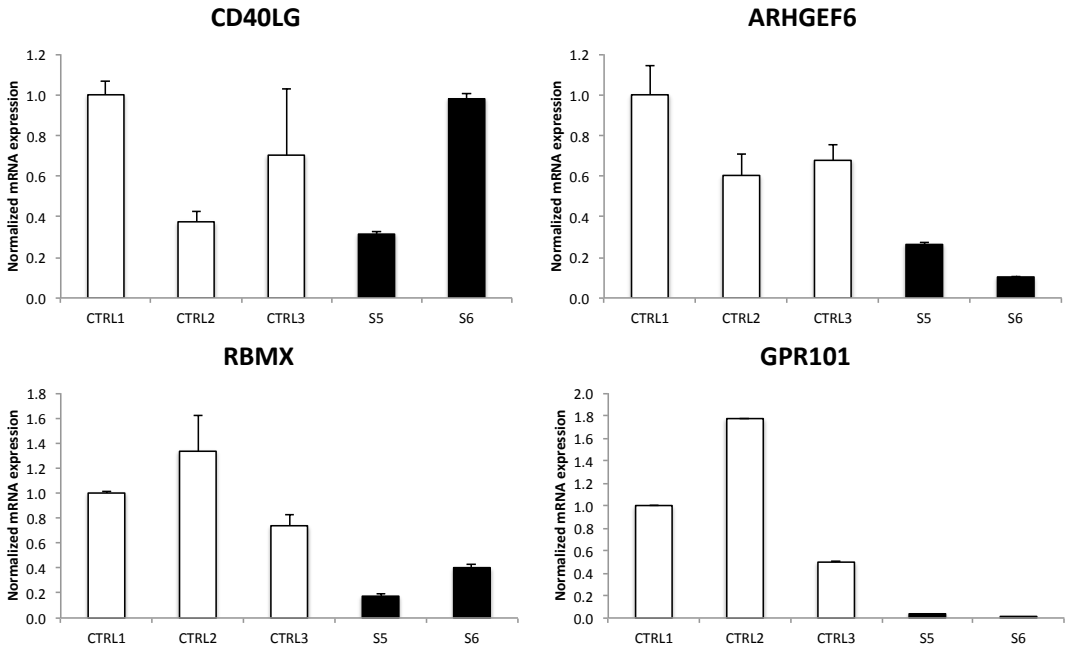


Figure S5: Genotyping of the duplications in the families.



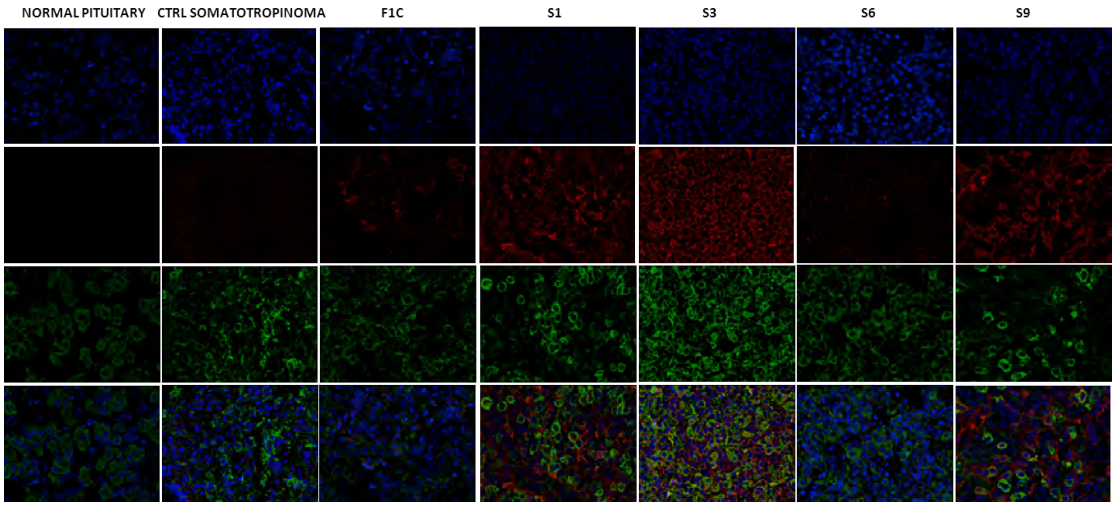
A) The four pedigrees belong to four sporadic cases, S2, S4, S5 and S6, respectively. **B)** The two pedigrees belong to four familial cases, F1A/F1B/F1C and F2A, respectively. Breakpoint junction PCRs are performed for the family members with available DNA. The duplications in S2, S4 and S6 are not observed in the parents. Duplication in S5 is not observed in the mother and the unaffected brother. The duplication in F1B and F1C appears to be inherited from the affected mother, F1A. The duplication in F2A is not identified in the unaffected sister. JCT: breakpoint junction.

Figure S6: Leukocyte mRNA expression profile of the 4 duplicated genes in affected patients.



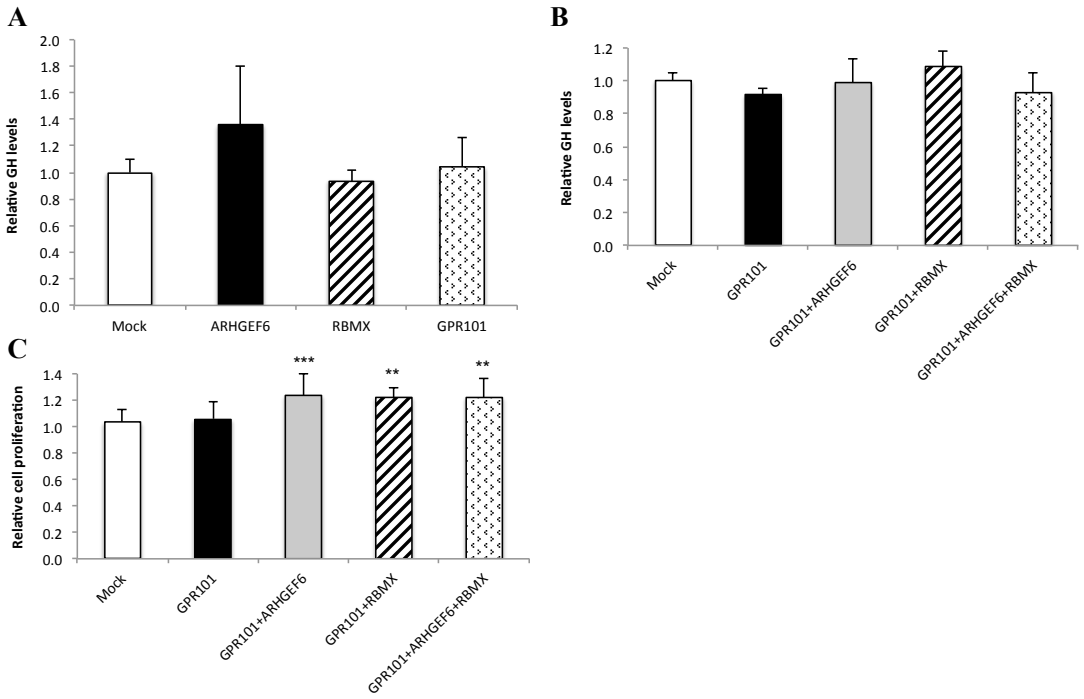
Expression values for each gene were normalized on *GAPDH* expression. Data are expressed as mean ± SD of 3 independent experiments each performed in triplicate.

Figure S7: Immunostaining for GPR101 (red) and GH (green) in patients with Xq26.3 microduplications.



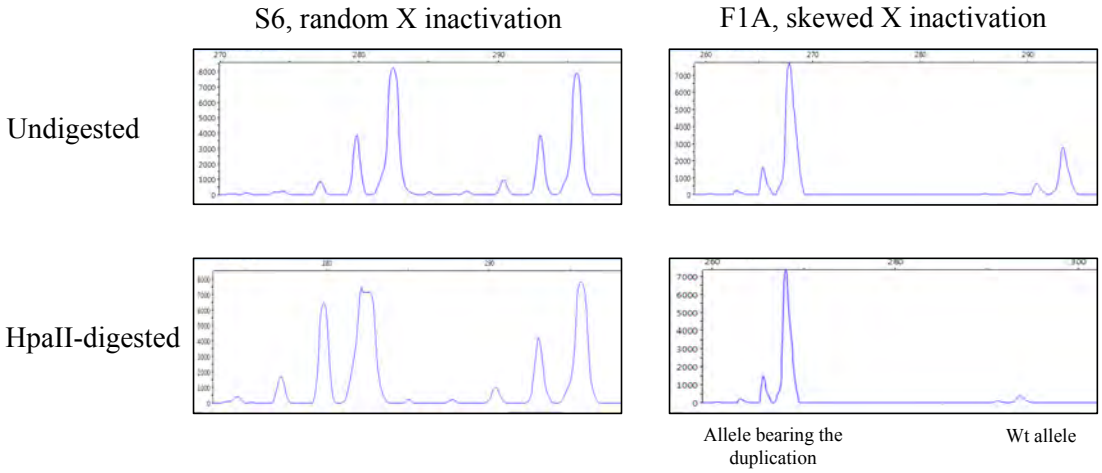
A higher GPR101 expression was observed in five tested patients harboring Xq26.3 microduplications compared to both normal pituitary and an age-matched control somatotropinoma without the duplication. Nuclei (blue) were stained with DAPI. The bottom panel shows the merged staining of GPR101, GH, and the nuclei.

Figure S8: GH secretion and cell proliferation after over-expression in GH3 cells of the duplicated genes expressed in the human pituitary.



A) Overexpression of each single wild-type sequence (wt) gene does not have any significant effect on hormone secretion. When *GPR101* was overexpressed together with either or both *ARHGEF6* and *RBMX*, a synergistic effect was shown on cell proliferation (**C**) but not on GH secretion (**B**). GH and proliferation values of empty vector-transfected cells were set as 1. Data are expressed as mean ± SD of 2-3 independent experiments each performed at least in triplicate. **, $P<0.01$; ***, $P<0.001$.

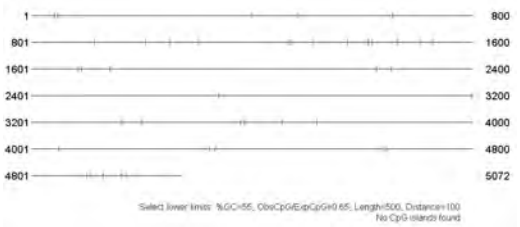
Figure S9: The fragment analysis results for a sporadic patient showing random X inactivation (S6, left panel) and the F1A patient with skewed X inactivation (right panel) are reported.



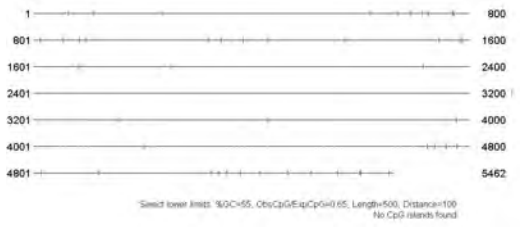
For patient S6 the AR peaks are shown, while in patient F1A the only informative locus was *FMR1*. The identity of the allele bearing an Xq26.3 microduplication was determined by analyzing the DNA of patient F1C.

Figure S10: CpG islands predicted for the four duplicated OMIM genes are showed.

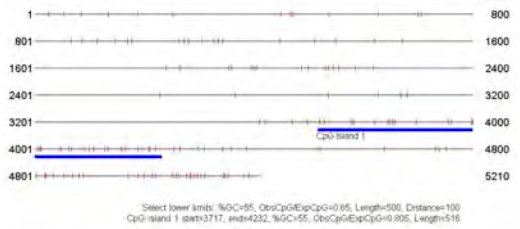
CD40LG, chrX:135,725,336-135,730,407



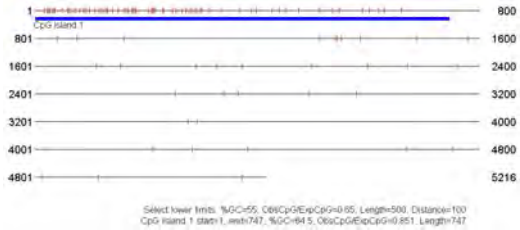
ARHGEF6, chrX:135,863,042-135,868,503



RBMX, chrX:135,961,587-135,967,939

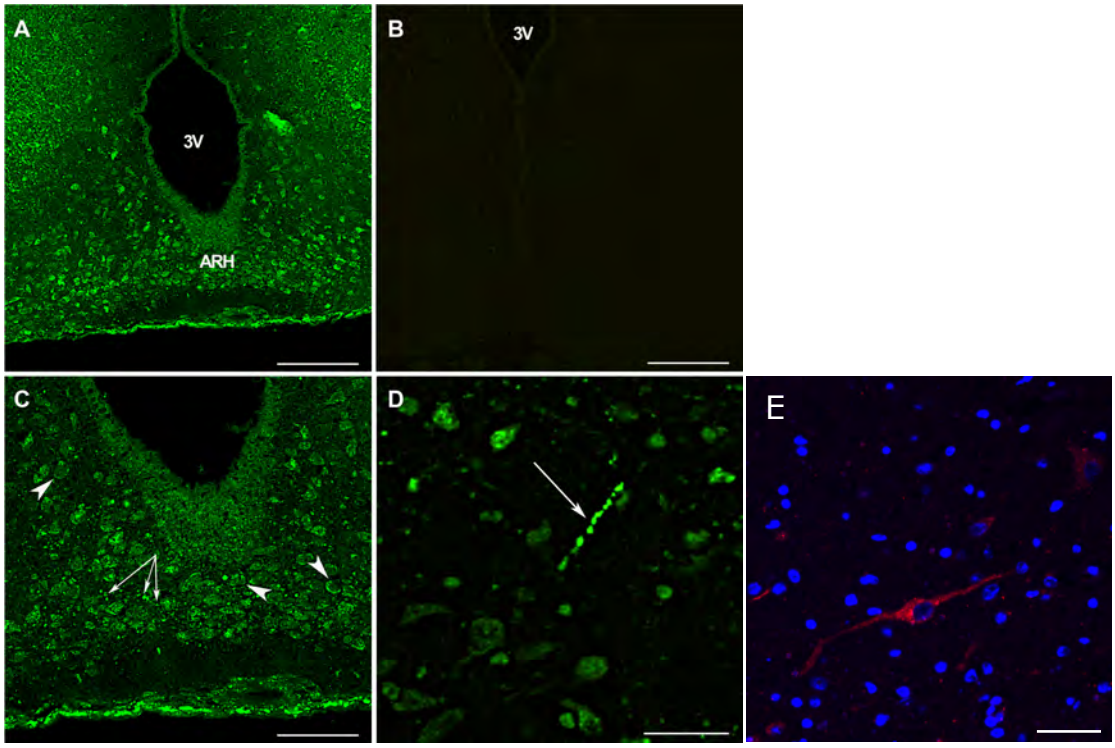


GPR101, chrX:136,115,961-136,121,176



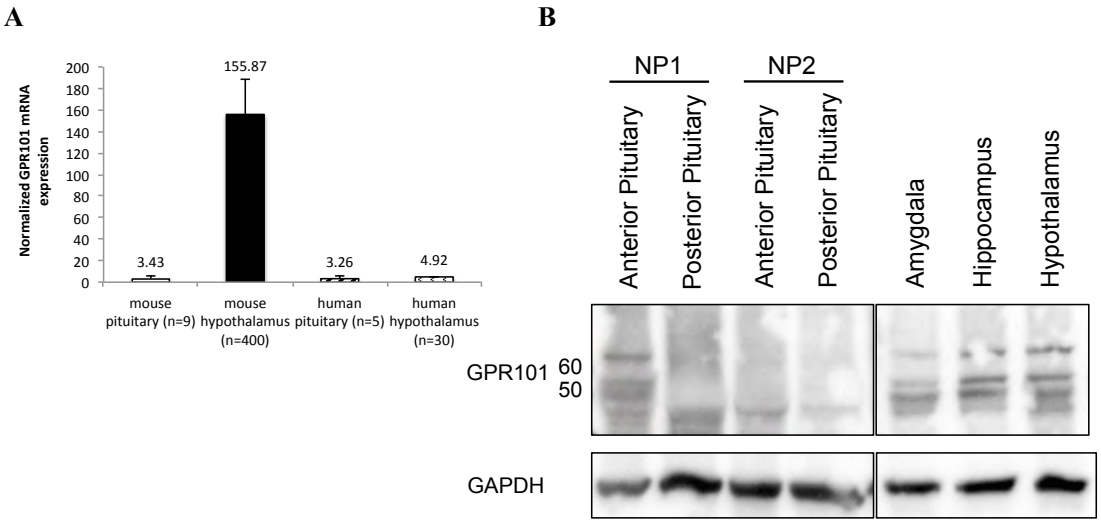
One CpG island was predicted for both *RBMX* and *GPR101*, whereas no CpG islands were predicted for *CD40LG* and *ARHGEF6*.

Figure S11: Immunofluorescence labeling of GPR101 in the mouse and human brain.



A) GPR101 was observed in the mouse hypothalamic area around third ventricle (3V), including the arcuate nucleus (ARH). **B)** Omitting primary antibody resulted in no staining. **C)** In mouse ARH, GPR101 was detected on neuronal cell bodies (arrowheads). Punctuate staining (arrows) suggested the presence of GPR101 receptor on neuronal processes and/or axon terminals. **D)** Thin, intensely labeled, varicose axons could be seen throughout the mouse hypothalamus. **E)** GPR101 was observed in some isolated neurons in the human ARH. Scale bars: 100 µm in A and B, 50 µm in C and E, and 25 µm in D.

Figure S12: RT-qPCR and Western Blot analysis of GPR101 expression in normal mouse and human pituitary and hypothalamus.



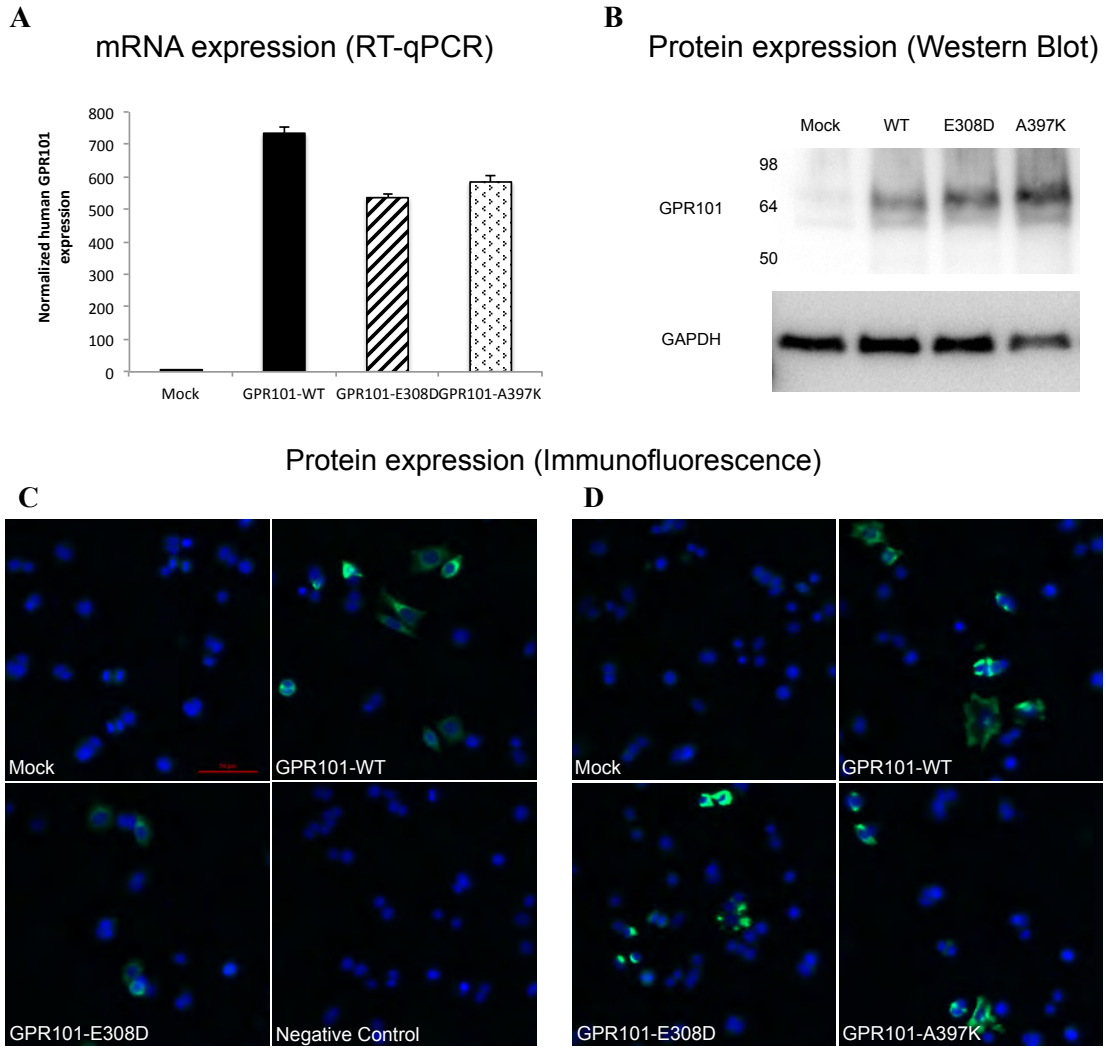
A) *Gpr101* is expressed at much higher levels in mouse hypothalamus compared to the pituitary, whereas in humans *GPR101* is expressed at similar levels in both tissues. Data are expressed as mean ± SD. **B)** Two *post-mortem* human pituitary tissues (NP1 and NP2) separated into the anterior and posterior lobes, and a pool of proteins isolated from amygdala, hippocampus, and hypothalamus from several deceased individuals were analyzed. GPR101 expression was detected in all the three brain regions analyzed (GPR101 predicted molecular weight: 57 kDa), whereas only one out of two pituitary samples (NP1, anterior lobe) showed low expression levels of the receptor.

Figure S13: Alignment of the sequences of human GPR101 (GPR101) and β_2 AR (3SN6).

```
                                                TM1
GPR101   1     MTSTCTNSTR ESNSSHTCMP LSKMPISLAH GIIRSTVLVI FLAASFVGNI
3SN6     30                                EVW VVGMGIVMSL IVLAIVFGNV

                          TM2
GPR101   51    VLALVLQRKP QLLQVTNRFI FNLLVTDLLQ ISLVAPWVVA TSVPLFWPLN
3SN6     53    LVITAIAKFE RLQTVTNYFI TSLACADLVM GLAVVPFGAA HILTKTWTFG

         TM3                                                    TM4
GPR101   101   SHFCTALVSL THLFAFASVN TIVVVSVDRY LSIIHPLSYP SKMTQRRGYL
3SN6     103   NFWCEFWTSI DVLCVTASIE TLCVIAVDRY FAITSPFKYQ SLLTKNKARV

GPR101   151   LLYGTWIVAI LQSTPPLYGW ---------- -----GQAAF DERNALCSMI
3SN6     153   IILMVWIVSG LTSFLPIQMH WYRQEAINCY AEETC----- ------CDFF

             TM5
GPR101   186   WGASPSYTIL SVVSFIVIPL IVMIACYSVV FCAARRQHAL LYNVKRHSLE
3SN6     195   T--NQAYAIA SSIVSFYVPL VIMVFVYSRV FQEAKRQLQK IDKSEGR---

GPR101   236   VRVKDCVENE DEEGAEKKEE FQDESEFRRQ HEGEVKAKEG RMEAKDGSLK
3SN6          ---------- ---------- ---------- ---------- ----------

                                  308
GPR101   286   AKEGSTGTSE SSVEARGSEE VREESSTVASD GSMEGKEGST KVEENSMKAD
3SN6          ---------- ---------- ---------- ---------- ----------

GPR101   336   KGRTEVNQCS IDLGEDDMEF GEDDINFSED DVEAVNIPES LPPSRRNSNS
3SN6          ---------- ---------- ---------- ---------- ----------

                  TM6
GPR101   386   NPPLPRCYQC KAAKVIFIII FSYVLSLGPY CFLAVLAVWV DVETQVPQWV
3SN6     265   -----CLKEH KALKTLGIIM GTFTLCWLPF FIVNIVHVIQ DNL--IRKEV

         TM7
GPR101   436   ITIIIWLFFL QCCIHPYVYG YMHKTIKKEI QDMLKKFFCK EKPPKEDSHP
3SN6     308   YILLNWIGYV NSGFNPLIYC RSPDFRIAFQ ELLC------ ----------

GPR101   486   DLPGTEGGTE GKIVPSYDSA TFP
3SN6          ---------- ---------- ---
```

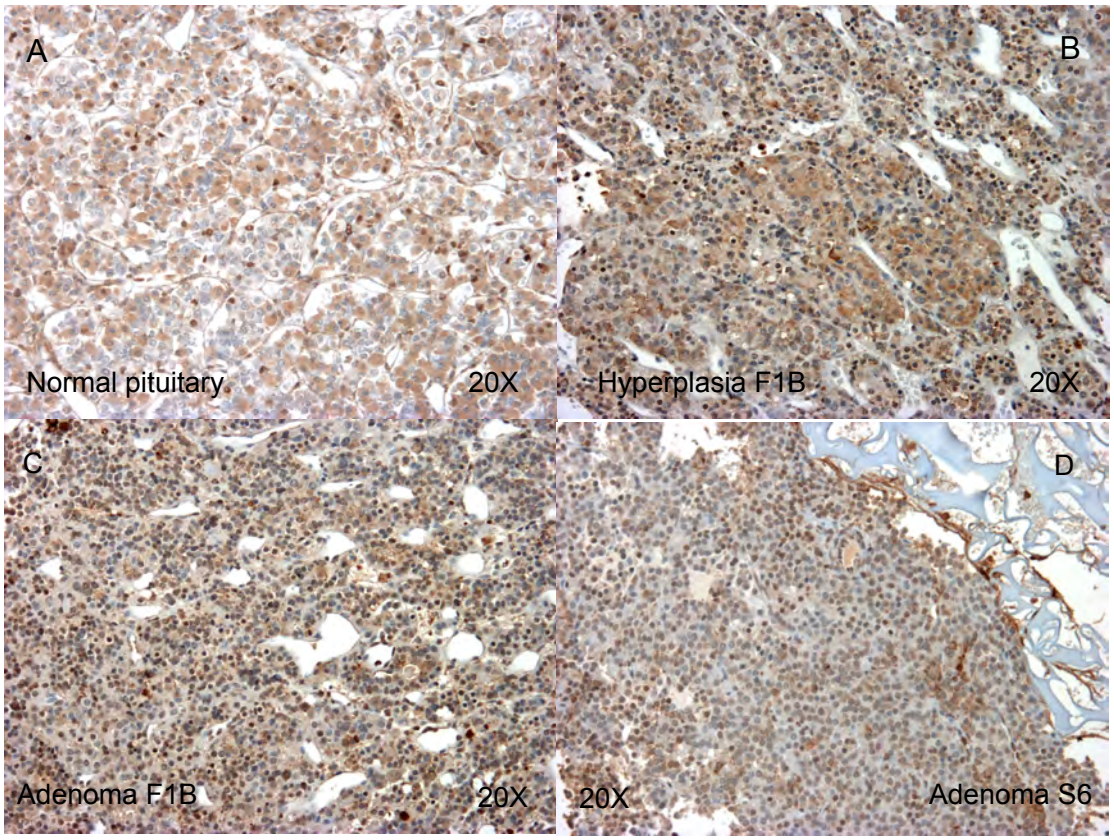
For the β_2 AR, only the residues solved in the 3SN6 structure are shown (30-175; 179-239; 265-341). The seven transmembrane domains (TMs) are highlighted in gray. Residue E308 of GPR101 is highlighted in bold and labeled.

Figure S14: Demonstration of human GPR101 overexpression achieved after *in vitro* transfection of GH3 cells.



WT and mutant human GPR101 expression was analyzed by RT-qPCR (**A**), Western Blot (**B**), and immunostaining (**C** and **D**). **A)** Human GPR101 mRNA levels were normalized on rat beta-actin expression. Data are expressed as mean ± SD. **B)** In Western Blot the anti-GPR101 SAB4503289 (dilution 1:500) antibody was used, whereas two antibodies anti-GPR101 were tested in immunofluorescence: **C)** SAB4503289 (dilution 1:500), and **D)** HPA001084, (dilution 1:1000). Both antibodies react with human and rat GPR101 homologues and were able to detect human GPR101 overexpression in a subset of GH3 cells, reflecting the transfection efficiency of 30-40% normally achieved in those cells. Omitting the primary antibody resulted in no staining (Negative Control in panel **C**). The same scale bar applies to all images: 50 µm.

Figure S15: GHRH-R staining in normal pituitary (Panel A), with increased staining intensity in hyperplasia from case F1B (Panel B) and adenoma samples from patients F1B (Panel C) and S6 (Panel D).



Supplemental Tables

Table S1. Frequency of the *GPR101* p.E308D mutation observed in the peripheral genomic DNA and tumor tissue DNA of patients with sporadic acromegaly.

Country	DNA origin	gender	n.	mutation n.	total sample [#]	sample frequency (%)	allele frequency (%) [$]
UK 1	tumor	n.a.	3	0	11	18.18	14.81
		M	7	2			
		F	1	0			
	blood	n.a.	7	0	18	5.56	4.08
		M	8	1			
		F	3	0			
UK2	tumor	M	5	0	12	0.00	0.00
		F	7	0			
Japan	tumor	n.a.	5	1	33	12.12	7.92
		M	13	2			
		F	15	1			
France	tumor	M	22	1	38	5.26	3.70
		F	16	1			
Belgium	tumor	M	12	1	23	4.35	8.82
		F	11	0			
	blood	M	48	1	88	2.27	1.56
		F	40	1			
US1	tumor	M	20	0	28	0.00	0.00
		F	8	0			
US2	blood	M	4	0	8	0.00	0.00
		F	4	0			
global	tumor	n.a.	8	1	145	6.21	4.35
		M	79	6			
		F	58	2			
	blood	n.a.	7	0	114	2.63	1.82
		M	60	2			
		F	47	1			

n - number; n.a. - not available; M - male; F - Female; [#] The total number of patient is 248, 11 patients have both DNA origins; [$] for allele frequency analysis, patients with unknown gender were counted as 1.5

Table S2. Nucleotide changes identified in the genes located within the common duplicated area in patients with sporadic acromegaly and pediatric pituitary adenomas

Gene	DNA change	Protein change	SNP id	Patient allelic frequency (%)	Control allelic frequency (%)	P-value	*In silico* pathogenicity
miR-934	c.688+65G>A	n.a.	rs73558572	1.439	0.146	0.0105	unlikely
CD40LG	c.148T>C	n.a.	rs1126535	15.108	19.108	n.s.	unlikely
	c.410-13T>C	n.a.	rs3092923	17.986	26.160	0.0431	unlikely
ARHGEF6	c.169T>C	p.C57R	rs147131853	0.719	0.301	n.s.	unknown
	c.250-25C>T	n.a.	rs5975789	1.460	1.145	n.s.	unlikely
	c.362G>A	p.R121H	rs35106300	0.730	0.314	n.s.	unknown
	c.891G>T	p.Q297H	rs5974620	0.730	1.206	n.s.	unlikely
RBMX	c.-1C>A	n.a.	rs2011584	70.896	33.180	<0.0001	unlikely
	c.541+11C>G	n.a.	rs145225005	2.143	1.688	n.s.	unknown
	c.656+30A>T	n.a.	rs2306222	4.317	4.159	n.s.	unlikely
	c.902C>A	p.P301Q	rs78646793	0.719	n.a.	n.a.	unknown
GPR101	c.91G>A	p.G31S	rs138068185	0.714	0.045	n.s.	unlikely
	c.370G>T	p.V124L	rs1190736	42.169	35.227	n.s.	unknown
	c.878C>T	p.T293I	rs73566014	2.857	5.003	n.s.	unlikely
	c.924G>C	p.E308D	rs73637412	2.86	0	<0.0001	unknown
	c.1127T>C	p.L376P	rs5931046	10.169	14.300	n.s.	unknown

The allelic frequencies identified in patients and in controls (according to public databases, including mainly the 1000 Genomes Project[7] and the Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP, http://evs.gs.washington.edu/EVS/) data) are shown. The nomenclature for the variant found in miR-934 refers to *VGLL1*, since the miRNA resides within an intron of this gene. No variants were identified in *SNORD61*. n.s.: not significant; n.a.: not applicable.

Table S3. Literature cases of early childhood onset gigantism due to proven GH hypersecretion from pituitary adenoma/hyperplasia.[19-32,43]

Sex	9F/6M (all sporadic)
Gestation (weeks)	All cases >34 weeks
Birth weight normal	11/14 reported cases (2 cases >95% for weight; 1 case 2100 g)
Birth length normal	6/7 reported (1 case >95% for length)
Median paternal height (cm; range)	176.5 (154.0-183.0)
Median maternal height (cm; range)	162 (146.0-170.0)*
Sibling growth normal	7/7 reported cases with siblings; all normal
Median age at onset of rapid growth (months; range)	18 (1-60)
Median age at diagnosis (months; range)	54.0 (10-132)
Median height age at diagnosis (months; range)	81 (33-132)
Median height SDS score at diagnosis (range)	+4.0 (+2.5 - +4.4)
Median weight SDS score at diagnosis (range)	+4.7 (+1.2 - +6.6)
Galactorrhea	1 episode in 1 case (others none)
Pubertal development begun at diagnosis	4 cases
Elevated GH/prolactin levels at diagnosis	100% of cases
GH unsuppressed by OGTT	0/11 of cases reported**
GH response on TRH test	Increased in 7/8 patients tested
Serum GHRH levels	Elevated in 3/5 cases tested***
Adenoma/hyperplasia confirmed pathologically**	11/4 (3 of the cases had both adenoma and hyperplasia)
Median maximum tumor diameter (mm; range)	16 (10-39)

* One set of parents were both >95% for height

** Four cases had a paradoxical increase in GH following OGTT

*** Only one case had levels increased above the threshold for a possible ectopic source of GHRH, although no ectopic source was found

**** 12 cases had reported pituitary pathological data

ULN: upper limit of the normal range

Table S4. Cases with deletions of the Xq26 region associated with short stature and developmental delay listed in the dbVar and DECIPHER databases as on June 2014.

Database	Code	Sex	ChrX region affected	Size (bp)	Phenotype
dbVar	nsv530253	n.a.	135866325-154903852	19,037,527	**Failure to thrive**; Global developmental delay; Microcephaly
dbVar	nsv533390	female	127434894-154886088	27,451,194	Developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing
dbVar	nsv530250	female	125499525-154879290	29,379,765	Developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing
dbVar	nsv529705	female	114558717-154905065	40,346,348	Global developmental delay
dbVar	nsv498095	female	113673848-154905065	41,231,217	Developmental delay and additional significant developmental and morphological phenotypes referred for genetic testing
dbVar	nsv530234	female	99779559-154899615	55,120,056	Global developmental delay, Short stature
dbVar	nsv531697	n.a.	76102-155226096	155,149,994	Intellectual functioning disability, Global developmental delay, Muscular hypotonia, **Short stature**, Hypothyroidism, Abnormality of the skeletal system, Seizure
DECIPHER	Patient 2552	n.a.	133634030-154754171	21,120,141	Abnormality of the endocrine system
DECIPHER	Patient 283237	female	118921340-155134088	36,212,748	Global developmental delay
DECIPHER	Patient 286978	female	110828029-155159533	44,331,504	Global developmental delay
DECIPHER	Patient 280504	female	108685763-154929420	144,060,844	Global developmental delay
DECIPHER	Patient 277955	female	61091-155232907	155,171,816	Abnormality of the thumb, Specific learning disability, Hip dysplasia, Leukonychia, Hip dislocation, Abnormality of the wrist, **Short stature**

n.a.: not available.

REFERENCE LIST

1. Shinawi M, Liu P, Kang SH, et al. Recurrent reciprocal 16p11.2 rearrangements associated with global developmental delay, behavioural problems, dysmorphism, epilepsy, and abnormal head size. Journal of medical genetics 2010;47:332-41.
2. Mayo P, Hartshorne T, Li K, McMunn-Gibson C, Spencer K, Schnetz-Boutaud N. CNV analysis using TaqMan copy number assays. Current protocols in human genetics 2010;Chapter 2:Unit2.13.
3. Manolakos E, Vetro A, Kefalas K, et al. The use of array-CGH in a cohort of Greek children with developmental delay. Molecular cytogenetics 2010;3:22.
4. Taymans SE, Pack S, Pak E, Torpy DJ, Zhuang Z, Stratakis CA. Human CYP11B2 (aldosterone synthase) maps to chromosome 8q24.3. The Journal of clinical endocrinology and metabolism 1998;83:1033-6.
5. Stratakis Ca, Tichomirowa Ma, Boikos S, et al. The role of germline AIP, MEN1, PRKAR1A, CDKN1B and CDKN2C mutations in causing pituitary adenomas in a large cohort of children, adolescents, and patients with genetic syndromes. Clinical genetics 2010;78:457-63.
6. Sandrini F, Kirschner LS, Bei T, et al. PRKAR1A, one of the Carney complex genes, and its locus (17q22-24) are rarely altered in pituitary tumours outside the Carney complex. Journal of medical genetics 2002;39:e78.
7. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491:56-65.
8. Olatubosun A, Väliaho J, Härkönen J, Thusberg J, Vihinen M. PON-P: integrated predictor for pathogenicity of missense variants. Human mutation 2012;33:1166-74.
9. Houdayer C. In Silico Prediction of Splice-Affecting Nucleotide Variants. In: Yu B, Hinchcliffe M, eds. In Silico Tools for Gene Discovery. Totowa, NJ: Humana Press; 2011:269-81.
10. Plon SE, Eccles DM, Easton D, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Human mutation 2008;29:1282-91.

11. Bell J, Bodmer D, Sistermans E, Ramsden S. Practice guidelines for the interpretation and reporting of unclassified variants (UVs) in clinical molecular genetics. Birmingham, United Kingdom: CMGS 2007.

12. Villa C, Magri F, Morbini P, et al. Silent familial isolated pituitary adenomas: histopathological and clinical case report. Endocrine pathology 2008;19:40-6.

13. Nowell PC, Hungerford DA. Minute chromosome in human chronic granulocytic leukemia. Science 1960;132:1497-.

14. Bertelsen B, Tümer Z, Ravn K. Three new loci for determining x chromosome inactivation patterns. The Journal of molecular diagnostics 2011;13:537-40.

15. Thouin MM, Giron JM, Hoffman EP. Detection of nonrandom X chromosome inactivation. Current protocols in human genetics 2003;Chapter 9:Unit9.7.

16. Rasmussen SG, DeVree BT, Zou Y, et al. Crystal structure of the beta2 adrenergic receptor-Gs protein complex. Nature 2011;477:549-55.

17. Lowell BB, Dhillon H, inventors; Novel gpr101 transgenic mice and methods of use thereof (http://www.google.com/patents/US20110173706). 2011.

18. MacKerell AD, Bashford D, Bellott M, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586-616.

19. Guyda H, Robert F, Colle E, Hardy J. Histologic, ultrastructural and hormonal characterization of a pituitary tumor secreting both HGH and prolactin. The Journal of clinical endocrinology and metabolism 1973;36:531-47.

20. Blumberg DL, Sklar CA, David R, Rothenberg S, Bell J. Acromegaly in an infant. Pediatrics 1989;83:998-1002.

21. Felix IA, Horvath E, Kovacs K, Smyth HS, Killinger DW, Vale J. Mammosomatotroph adenoma of the pituitary associated with gigantism and hyperprolactinemia. A morphological study including immunoelectron microscopy. Acta neuropathologica 1986;71:76-82.

22. Ritzen EM, Wettrell G, Davies G, Grant DB. Management of pituitary gigantism. The role of bromocriptine and radiotherapy. Acta paediatrica Scandinavica 1985;74:807-14.

23. Rusconi R, Ballabio G, Piceni Sereni L. [Gigantism caused by a double-secreting (GH-PRL) acidophile adenoma in a 6-year-old girl]. Minerva pediatrica 1985;37:815-9.

24. Moran A, Asa SL, Kovacs K, et al. Gigantism due to pituitary mammosomatotroph hyperplasia. The New England journal of medicine 1990;323:322-7.
25. Dubuis JM, Deal CL, Drews RT, et al. Mammosomatotroph adenoma causing gigantism in an 8-year old boy: a possible pathogenetic mechanism. Clin Endocrinol (Oxf) 1995;42:539-49.
26. Alikasifoglu A, Kandemir N, Akalan N, Yordam N. Pituitary adenoma associated with gigantism and hyperprolactinemia. Pediatric neurosurgery 2001;35:325-8.
27. Fazekas I, Pasztor E, Slowik F, et al. Pathological and experimental investigations in a case of gigantism. Acta neuropathologica 1993;85:167-74.
28. Goldenberg N, Racine MS, Thomas P, Degnan B, Chandler W, Barkan A. Treatment of pituitary gigantism with the growth hormone receptor antagonist pegvisomant. The Journal of clinical endocrinology and metabolism 2008;93:2953-6.
29. Nanto-Salonen K, Koskinen P, Sonninen P, Toppari J. Suppression of GH secretion in pituitary gigantism by continuous subcutaneous octreotide infusion in a pubertal boy. Acta Paediatr 1999;88:29-33.
30. Zimmerman D, Young WF, Jr., Ebersold MJ, et al. Congenital gigantism due to growth hormone-releasing hormone excess and pituitary hyperplasia with adenomatous transformation. J Clin Endocrinol Metab 1993;76:216-22.
31. Gelber SJ, Heffez DS, Donohoue PA. Pituitary gigantism caused by growth hormone excess from infancy. The Journal of pediatrics 1992;120:931-4.
32. Narayanaswamy V, Rettig KR, Bhowmick SK. Excessive growth. Clinical pediatrics 2008;47:705-8.
33. Zhang F, Khajavi M, Connolly AM, Towne CF, Batish SD, Lupski JR. The DNA replication FoSTeS/MMBIR mechanism can generate genomic, genic and exonic complex rearrangements in humans. Nature genetics 2009;41:849-53.
34. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. Cell 2007;131:1235-47.
35. Hastings PJ, Lupski JR, Rosenberg SM, Ira G. Mechanisms of change in gene copy number. Nature reviews Genetics 2009;10:551-64.

36. Plenge RM, Stevenson RA, Lubs HA, Schwartz CE, Willard HF. Skewed X-chromosome inactivation is a common feature of X-linked mental retardation disorders. American journal of human genetics 2002;71:168-73.
37. Grasshoff U, Bonin M, Goehring I, et al. De novo MECP2 duplication in two females with random X-inactivation and moderate mental retardation. European journal of human genetics 2011;19:507-12.
38. Herman V, Fagin J, Gonsky R, Kovacs K, Melmed S. Clonal origin of pituitary adenomas. J Clin Endocrinol Metab 1990;71:1427-33.
39. Carrel L, Willard HF. X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 2005;434:400-4.
40. Cotton AM, Lam L, Affleck JG, et al. Chromosome-wide DNA methylation analysis predicts human tissue-specific X inactivation. Human genetics 2011;130:187-201.
41. Yasukochi Y, Maruyama O, Mahajan MC, et al. X chromosome-wide analyses of genomic DNA methylation states and gene expression in male and female neutrophils. Proceedings of the National Academy of Sciences of the United States of America 2010;107:3704-9.
42. Zhang Y, Castillo-Morales A, Jiang M, et al. Genes that escape X-inactivation in humans have high intraspecific variability in expression, are associated with mental impairment but are not slow evolving. Molecular biology and evolution 2013;30:2588-601.
43. Hindmarsh PC, Pringle PJ, Brook CG. The 24-hour growth hormone secretion in a boy with giantism. Acta endocrinologica 1988;117:403-8.