

Syndromic cases of hemophilia B and morbid obesity due to contiguous gene deletions on Xq26.3-q27.2: unsuspected phenotype-genotype associations by bioinformatics and extensive clinical data mining

Pamela Radic¹, Miguel Abelleiro¹, Betiana Ziegler¹, Vanina Marchione¹, Julian Nevado², Pablo Lapunzina², Gabriela Sciuccati³, Daniela Neme⁴, Liliana Rossetti¹, Mariana Bonduel⁵, and Carlos De Brasi¹

¹Instituto de Medicina Experimental (IMEX-CONICET/ANM)

²Instituto de Genética Médica y Molecular (INGEMM)-IdiPAZ, Hospital Universitario La Paz

³Hospital de Pediatría Prof Dr Juan P Garrahan Servicio de Hematología y Oncología

⁴Fundación de la Hemofilia Alfredo Pavlovsky.

⁵Hospital de Pediatría Prof Dr Juan P Garrahan Servicio de Hematología y Oncología

August 20, 2020

Abstract

Hemophilia B (HB) associates with pathogenic F9-variants. The literature showed that hemizygous deletions encompassing F9 and vicinal genes may express extra-phenotypes suggesting new causal relationships. Aim: Analyze the molecular basis of syndromic cases of HB, obesity (OB), severe global developmental delay (SGDD) and generalized hypotonia (GH). Whole F9-deletions were detected in 3 hemizygous probands with HB. Dense SNP-array and case-specific STS walking strategies allowed amplification and characterization of the deletion breakpoints. Bioinformatic/statistical analyses included data mining in HPO (Human Phenotype Ontology), OMIM, STRING (protein-protein interaction networks) databases and estimation of null-hypothesis-based Expected-values. Patients (cases#3-case#1) showed complete F9 deletions involving 0.16-4.34-Mb and 1-17 additional genes on Xq26.3-q27.2. Bioinformatic/statistical approaches revealed highly significant STRING-associations ($P[?][0.00115]$) between case#1/#2 common deleted genes (SOX3, FGF13, CXorf66) and those HPO associated with OB (20/343), GH (36/923) and SGDD (10/119). Only case#2 showed two extra-phenotypic abnormalities, anal atresia and pituitary hypothyroidism, and one extra-deleted gene, MAGEC2. Our bioinformatic/statistical approaches confirmed a previous involvement of SOX3 in OB suggesting additional roles in GH and SGDD, similar to FGF13 and CXorf66 (experimentally linked to POMGNT1, HPO-connected with all 3 phenotypes). Our findings highlight the importance to characterize X-chromosome deletion syndromes to unveil functional associations of the involved genomic regions.

Syndromic cases of hemophilia B and morbid obesity due to contiguous gene deletions on Xq26.3-q27.2: unsuspected phenotype-genotype associations by bioinformatics and extensive clinical data mining

Claudia Pamela Radic^{*1}, Miguel Martín Abelleiro^{*1}, Betiana Ziegler¹, Vanina Daniela Marchione¹, Julian Nevado^{2,3}, Pablo Lapunzina^{2,3}, Gabriela Sciuccati⁴, Daniela Neme⁵, Liliana Carmen Rossetti¹, Mariana Bonduel^{**4}, Carlos Daniel De Brasi^{**1,6}

Affiliations: ¹Laboratorio de Genética Molecular de la Hemofilia (GMH), Instituto de Medicina Experi-

mental (IMEX, CONICET-Academia Nacional de Medicina (ANM)), Buenos Aires, Argentina; ²Instituto de Genética Médica y Molecular (INGEMM)-IdiPAZ, Hospital Universitario La Paz, Universidad Autónoma, Madrid, Spain; ³CIBERER, Centro de Investigación Biomédica en Red de Enfermedades Raras, ISCIII, Madrid, Spain; ⁴Servicio de Hematología y Oncología, Hospital de Pediatría Prof. Dr. Juan P. Garrahan (HPJG); ⁵Fundación de la Hemofilia Alfredo Pavlovsky (FHAP); ⁶Instituto de Investigaciones Hematológicas Mariano R. Castex (IIHEMA, ANM), Buenos Aires, Argentina.

*CPR and MMA contributed equally as co-first authors.

** MB and CDB contributed equally as co-senior authors.

Address correspondence to: Claudia Pamela Radic, Molecular Genetics of Haemophilia Laboratory, Academia Nacional de Medicina. Pacheco de Melo 3081, Buenos Aires (1425), Argentina. yocpamelita@yahoo.com.ar, 5411 48058803, Ext.: 261.

Claudia Pamela Radic ORCID: 0000-002-3901-3904

Summary

Hemophilia B (HB) associates with pathogenic *F9*- variants. The literature showed that hemizygous deletions encompassing *F9* and vicinal genes may express extra-phenotypes suggesting new causal relationships.

Aim: Analyze the molecular basis of syndromic cases of HB, obesity (OB), severe global developmental delay (SGDD) and generalized hypotonia (GH).

Whole *F9* -deletions were detected in 3 hemizygous probands with HB. Dense SNP-array and case-specific STS walking strategies allowed amplification and characterization of the deletion breakpoints. Bioinformatic/statistical analyses included data mining in HPO (Human Phenotype Ontology), OMIM, STRING (protein-protein interaction networks) databases and estimation of null-hypothesis-based Expected-values.

Patients (cases#3-case#1) showed complete *F9* deletions involving 0.16-4.34-Mb and 1-17 additional genes on Xq26.3-q27.2.

Bioinformatic/statistical approaches revealed highly significant STRING-associations ($P[?]0.00115$) between case#1/#2 common deleted genes (*SOX3*, *FGF13*, *CXorf66*) and those HPO associated with OB (20/343), GH (36/923) and SGDD (10/119). Only case#2 showed two extra-phenotypic abnormalities, anal atresia and pituitary hypothyroidism, and one extra-deleted gene, *MAGEC2*.

Our bioinformatic/statistical approaches confirmed a previous involvement of *SOX3* in OB suggesting additional roles in GH and SGDD, similar to *FGF13* and *CXorf66* (experimentally linked to *POMGNT1*, HPO-connected with all 3 phenotypes). Our findings highlight the importance to characterize X-chromosome deletion syndromes to unveil functional associations of the involved genomic regions.

Keywords: Hemophilia B, obesity, severe global developmental delay, syndromic Xq27 deletions, bioinformatics.

INTRODUCTION

According to Human Phenotype Ontology (HPO), clinical characteristics as hemarthroses (HP:0005261), spontaneous joint hemorrhage (HP:0005261) associated with spontaneous or prolonged hemorrhages, are frequently found in hemophilia B (HB, OMIM:306900), which associates with defects in the factor IX gene (*F9*, OMIM 300746). However, clinical characteristics as developmental delay (HP:0001263), generalized hypotonia (HP:0001290), central hypothyroidism (HP:0011787), anal atresia (HP:0002023) and obesity (HP:0001513) are each associated with hundreds of gene defects genome-wide as occurred with most human diseases according to HPO.

The plasmatic FIX activity levels (FIX:C) permits classification of HB in severe (<1%), moderate (1–5%) and mild (5–40%), in which the clinical phenotypic severity, and its accompanying FIX:C, is closely related to the deleterious nature of the causative genetic variant (Tjarnlund-Wolf et al. 2019). However, dissimilar

clinical HB phenotypes have been reported in patients with identical genetic variants in the *F9*, e.g., discordant bleeding phenotypes were described in probands of the same family (Santagostino et al. 2010). Nevertheless, all cases with entire *F9* deletions showed FIX:C<1%. Development of inhibitory antibodies against exogenous FIX (inhibitors) is an important complication for patient's treatment and while HB cases show inhibitor frequencies ranging 1.4-11.7%, patients with complete *F9* deletions may show >20% (Radic et al. 2013; Santoro et al. 2018; Thorland et al. 1999). But in *F9*, as in most human genes, both partial and complete gene deletions are uncommon (Ketterling et al 1994).

Due to technical reasons most large deletions are only detected in hemizygous patients but they are rarely characterized at the sequence level. However, some reported patients with deletions of the entire *F9* and vicinal genes on Xq26-q27 presented additional phenotypic abnormalities different from those associated with HB, such as intellectual disability (ID), hypopituitarism, seizures, scoliosis, short stature, hypotonia, overweight, bilateral cryptorchidism (Nakamura et al. 2016; Hewitt et al. 2014; Jourdy et al. 2016; Stoof et al. 2018; Vencesla et al. 2007; Alatzoglou et al. 2014) while others patients only showed bleeding symptoms associated with HB (Nakamura et al. 2016; Koeberl et al. 1990).

The aim of this work was to characterize the pathogenic event in three patients with severe HB showing large deletions encompassing all *F9* exons and vicinal genes configuring complex syndromic cases of hemophilia and morbid obesity. The precise molecular definition of these events allowed a comprehensive correlation of their genetic extents with their specific clinical manifestations using available genotype-phenotype databases, gene-gene associations and reviewing virtually all cases of entire *F9* deletions reported in the literature.

MATERIAL AND METHODS

Studied populations and samples

Patients with severe HB included in this study showed a lack of PCR amplification of all eight exons of the *F9*. Genomic-DNA was obtained from peripheral blood leukocytes using a standard salting out method (Lahiri et al. 1991).

Our local institutional Ethical Committee (CEIANM) approved this study (Res.215/96-April/2007) and a written informed consent was obtained from all cases.

Case Report: clinical features and outcomes

Case#1. A 7-month-old boy was transferred to the Intensive Care Unit of the HPJG for severe intracranial hypertension secondary to meningitis due to *Streptococcus pneumoniae*. Neuroimaging detected arterial ischemic stroke of the left middle cerebral artery with an extensive hemorrhagic transformation, severe cerebral edema and cerebral herniation. An emergency decompressive craniotomy was performed without bleeding complications. Postoperatively, after a new laboratory, he was diagnosed with severe HB without inhibitor, and plasma-derived Factor IX concentrates were given. No history of allergic reactions or inhibitor development was observed (less than 50 exposures days). The patient presented a mild bleeding phenotype. He was a term neonate with normal birth weight born by vaginal delivery after an uncomplicated pregnancy. He was the son of unrelated parents with no family history of hemophilia. He developed severe obesity (OB) since the first months of life and physical evaluation also revealed severe global developmental delay (SGDD) and generalized hypotonia (GH). He had normal male karyotype. Severe sequelae related to meningitis, arterial ischemic and cerebral hemorrhage were observed during his follow-up. The boy died at 3 years-old due to aspiration pneumonia. The proband's mother shows a FIX activity of 103 IU/dL.

Case#2. A 2 year-old boy was transferred to HPJG for colostomy closure. He was a 36 weeks' gestation neonate with normal birth weight born by vaginal delivery after an uncomplicated pregnancy. He was the only son of unrelated parents with no family history of hemophilia. The neonate underwent urgent abdominal surgery in his local town 12 hours after his birth because of anal atresia (AA). The surgery was performed without knowledge of a diagnosis of hemophilia and no bleeding complications were observed. At 8 months of age in a preoperative evaluation he was subsequently diagnosed with severe HB. He developed OB since the first year of life and physical evaluation also revealed SGDD and GH. A trail of thyroid hormones

confirmed central pituitary hypothyroidism (PH). He had normal male Karyotype. The patient presented a mild bleeding phenotype. After more than 50 exposure day to plasma-derived Factor IX concentrates he presented an allergic reaction and he developed a low response (<2 UB/mL) transient inhibitor. The colostomy closure was performed when he was 2 1/2 years old. The boy died one month later due to severe sepsis related to postoperative complications. The proband's mother shows a FIX:C of 95 IU/dL.

Case#3 . Patient with severe HB from FHAP who developed FIX inhibitor at the age of 30 years (titer: 22 BU/mL) after 50-100 exposures days to therapeutic plasma-derived FIX concentrates. At 7 months of age, he had central nervous system hemorrhage but did not receive the recommended management and presented a right hemiparesis with severe walking disability. The patient has two brothers with severe hemophilia B, one of them developed FIX inhibitor and the other was not studied, and one sister showed FIX:C levels of 10 IU/dL.

Clinical characteristic of all cases are summarized in Table 1.

Molecular genotype characterization

All relevant *F9* sequences were analyzed by PCR amplification, according to Radic et al (Radic et al. 2013). The presence/absence and molecular size identity of each product was analyzed to detect large *F9* deletions in multiplex amplification using an uninvolved autosomal region within the β -globin gene as a control reference *locus* .

SNP-array analysis was performed, using the platform CytoSNP850K (Illumina, San Diego, CA). A genome-wide scan of 850,00 tag SNPs was conducted according the manufacture's specifications (Illumina, San Diego, CA). Resolution is 1SNP/2.5Kb and generally (95%), specific enrichment coverage for chromosomal cytobands/telomeres and centromeres and large regions defined from ISCA (International Standards for Cytogenomic Arrays) and CCMC (Cancer Cytogenomics Microarray Consortium). Image data were analyzed using the Chromosome Viewer tool contained in Genome Studio (Illumina, San Diego, CA). The metric used was the log R ratio which is the log (base 2) ratio of the observed normalized R value for a SNP divided by the expected normalized R value (under manufacturer's specifications). In addition, an allele frequency analysis was applied for all SNPs. All genomic positions were based upon Genome Reference Consortium Human Build 38 patch release 13 (GRCh38.13).

Case-specific sequence tagged sites (STS) were designed and applied to refine a delimitation of each deletion breakpoints, taking into account the results obtained from the SNP-array experiments. Sets of 3 or 4 primer pairs were used from each positive-negative SNP interval of SNP-array (Supplementary Figure SF1; Supplementary Table ST1). Gap-PCR was performed bridging (encompassing) each deletion breakpoint junction.

Long range-PCR (LR-PCR) : LR-PCR amplification was performed in a volume of 25 μ l with 500 ng of gDNA input using 0.75 U of KAPA Long Range DNA Polymerase (KAPA Biosystem, USA) with 0.6 μ M primers located by the inferred 5' and 3' deletion breakpoints (Supplementary Table ST1) in a Thermal-Cycler BIOER (China) under conditions recommended by the manufacturers.

Standard size gap-PCR : Amplifications ranging 700-2,100 bp were performed in a volume of 25 μ l with 300 ng of gDNA input using 1.5 U of Go Taq DNA Polymerase (Promega, Argentina) and primers designed to characterize the breakpoint junctions by Sanger sequencing (Supplementary Table ST1) in a Thermal-Cycler Biometra (Germany) applying standard reagents and thermal-cycling conditions.

Bioinformatics, genotype-phenotype data mining and statistical analysis

Most DNA sequence analysis and primer annotations were performed using SeqBuilder 7.1 (DNA Star Laser-gene). Each characterized 5' and 3' deletion breakpoint was mapped to the Reference Genome (GenBank GRCh38.p13, ChrX: NC_000023.11). Genetic variant nomenclature follows the guidelines of the Human Genome Variation Society (HGVS) (den Dunnen 2017). HPO was accessed to precisely classify phenotypic abnormalities and to annotate the listed panel of associated genes (<http://hpo.jax.org/app/>).

To study the relationships between gene-products, the STRING protein-protein interaction network was accessed online (<http://string-db.org/>) (Szklarczyk et al. 2019). In order to determine whether or not a proposed connection between a gene function lost in the hemizygous deletion is significantly associated with a particular phenotypic abnormality (part of a syndromic case) a set of null-hypothesis (Ho) associated expected-value (E-value) was obtained applying Monte-Carlo-like simulations based on picking 100 genes at or near random genomic locations (control gene set) by choosing GRCh38.p13 coordinates using aleatory numbers (<http://www.random.org>).

RESULTS

Genotype characterization of the structural variants

Patients failed to amplify all 12 *F9* specific amplimers and SNP-array studies revealed that *F9* deletions involved additional regions nearby. SNP-array results narrowed case#1 deletion to about 4.33 Mb on Xq27.1-27.2, case#2, to 3.9 Mb on Xq27.1-q27.2 and case#3 to 0.165 Mb on Xq27.1 allowing the setting up of a STS walking strategy to delimitate further the deletion breakpoints to PCR amplifiable intervals (Supplementary Fig. SF1).

Gap-PCR amplifications were achieved bridging each breakpoint junctions in all cases (case#1-case#3) using standard, long-range and/or nested PCR amplification approaches (Supplementary Table ST1, Supplementary Figure SF1).

Sanger sequencing permitted full characterization of all three deletions involving the entire *F9*. *Case#1* showed a deletion of 4,339,584 bp, NC_000023.11: g.137,786,488_142,126,072del, involving chromosome Xq26.3-q27.2 interval; *case#2*, a 3,904,082 bp deletion, g.138,932,331_142,836,413del, involving chrXq27.1-q27.2; and *case#3*, a 160,027 bp deletion, g.139,517,752_139,677,779del, involving a region within chrXq27.1 (Table 2).

Functional relationship between genotype-phenotype

Case#1, Case#2 and Case#3 expressed severe-HB

The deletion of the entire *F9* is the primary cause for the *reduced factor IX activity* (HP:0011858), which in term triggers a number of phenotypic abnormalities associated with severe HB in all cases (Figure 1, Table 1). In apparent contrast, cases #1 and #2, which showed larger deletions than case#3 (4.3 and 3.9 vs 0.16), both expressed mild bleeding phenotypes and did not develop FIX inhibitors whilst case#3 did. Case#3 also presented a partial deletion of the *MCF2* gene associated with an oncogenic potential, by encoding a guanine nucleotide exchange factor (GEF) that exerts control over some members of the Rho family of small GTPases (Figure 1, Supplementary Table ST2).

Supplementary Table ST2 presents the characteristics and functions associated with genes (n=18) involved in all cases and genomic elements collected from NCBI (National Center for Biotechnology Information) databases (Table 2).

Case#1 and Case#2 expressed SGDD, GH and OB

In order to unveil the relationship between the genotypes of case#1 and #2 characterized by a common region of hemizygous deletions on Xq26.3-q27.2 both sharing 13 genes (*FGF13*, *ATP11C*, *CXorf66*, *SOX3*, *CDR1*, *SPANXB1*, *LDOC1*, *SPANXC*, *SPANXA1*, *SPANXA2*, *SPANXD*, *MAGEC3*, *MAGEC1*) (Figure 1) excluding *F9* and *MCF2* (that are deleted in the case#3 showing only a hemophilia related phenotype) and their common extra-hemophilia phenotype (i.e., SGDD, GH, and OB) (Table 1), an unbiased blind-connecting bioinformatic approach was devised and applied. This approach studies the degree of protein-protein interaction by use of the STRING network by weighting the eventual connection between this set of 13 case#1/#2 common genes with the gene sets HPO associated with each of the phenotypic abnormalities, i.e., SGDD (n=119), GH (n=923) and OB (n=343). Given that STRING analysis involves evidence from broadly different sources (Genome Neighborhood, Gene Fusion, Co-occurrence, Co-expression,

Experiments, Databases, Text-mining and Homology) the first approach only considered the *number* of connections classifying 0 (no interaction), 1, 2 and [?] 3 connections (Figure 2). This approach considering all evidence indicated the connections between SGDD (HP:0011344) with the case#1/#2 common deleted genes involved *SOX3* (2/100, only 2 in 100 randomly chosen genes were found to show equal or more connections with SGDD than *SOX3*), *FGF13* (5/100) and *CXorf66* (40/100). The Ho associated with obtaining this degree of connection genotype-phenotype, calculated by computer simulations as indicated in M&M, estimated a combined probability of SGDD, P (SGDD)= 0.0004. The same analysis with GH (HP:0001290) connected with *FGF13* (4/100), *SOX3* (18/100), *ATP11C* (31/100), *CXorf66* (64/100), *MAGEC3* (66/100) and *SPANXC* (8/100) configuring a combined probability based on Ho of P (GH)= 0.000754. The analysis of OB (HP:0001513) connected with *FGF13* (6/100), *SOX3* (11/100), *ATP11C* (51/100), *MAGEC3* (51/100) and *CXorf66* (67/100) estimating a combined Ho probability based on randomness of P (GH)= 0.00115. These highly significant P figures ($P < 0.0012$) indicated that obtaining such degree of STRING interactions between case#1/#2 common deleted gene-products and those associated with their shared phenotype features (i.e., SGDD, GH and OB) are exceedingly unlikely to be obtained by chance.

In order to refine this unbiased bioinformatic approach, a similar investigation of STRING connections but focusing only on the *experimental evidence* (Figure 2) was applied to analyze the same set of 13 deletion genes with the same sets of genes related to SGDD (n=119), GH (n=923) and OB (n=343). This approach showed connections between *CXorf66* with a SGDD associated gene *POMGNT1* with a score 0.797 (Ho: 0/100 simulations with *score* [?]0.797) estimating a P (exp.SGDD)< 0.01. The same analysis in GH associated *CXorf66* with *POMGNT1* (*score* 0.797) ($P = 0.06$), *SOX3* with *ATXN1* (*score* 0.750) ($P = 0.06$) and *ATP11C* with *ATP6VOA2* (*score* 0.546) ($P = 0.09$) estimating a global P (exp.GH)= 0.000324. In OB, *CXorf66* showed a 0.797 scored association with *POMGNT1* estimating a P (exp.OB)= 0.03. It is interesting to note that *POMGNT1*, experimentally linked to *CXorf66* with a high evidence *score*, is a HPO-associated gene with all three clinical features. Again, the Ho associated P estimations (P exp.< 0.05) indicated that the obtained magnitude of STRING experimentally-demonstrated interactions between case#1 and #2 shared deleted genes and those associated with their three common phenotype features are not expected to be obtained by chance.

The all-evidence STRING showed a comprehensive map of connections between the common deleted genes of cases#1 and #2 (6 out of 13) with the genes associated with SGDD (7/119), GH (6/923) and OB (9/343) (Supplementary Figure SF2). Interestingly, despite of their connection to the genes that define cases' genotypes, three genes, *NTRK2*, *OTX2* and *POMGNT1*, are related to all three phenotypes by HPO (SGDD-GH-OB); another three, *FGFR3*, *HESX1* and *MECP2*, are related to SGDD and OB; two, *CTNNB1* and *CNKSRR2*, are related to OB and GH; and *POU1F1*, to SGDD and GH, while *SOX3*, a gene associated with OB, is directly deleted both clinical cases #1 and #2 (Supplementary Figure SF2).

Case#2 expressed PH and AA

Although case#1 encompassed a bigger deletion than case#2 (4.3 vs 3.9 Mb), the latter involved an additional gene respect to case#1, *MAGEC2* (Figure 1), and expressed two additional phenotypic abnormalities (i.e., PH and AA) (Table 1). HPO database presents 115 genes linked to PH and 26 genes to AA. A STRING analysis focused only on the case#2 extra-gene *MAGEC2* indicated only a modest single direct association with an AA associated gene, *ZIC3* and no association with any PH gene (Figure 3). Understanding that the complex syndrome expressed by case#2 including severe HB, OB, SGDD, GH, AA and PH may be consequence of the entire set of deleted genes (Figure 1), a comprehensive STRING analysis similar to those blind-connecting bioinformatic approach performed above with equally devised Ho/E-values was applied to link with AA and PH (Figure 3).

This approach only showed connections with PH and AA when considering all type of evidence but it did not when only experimental evidence was considered. The STRING analysis showed weighted connections between PH (HP:0008245) genes with *FGF13* (found in a single control out of 100 randomly picked genes), *ATP11C* (22/100), *SOX3* (0<1/100) and *CDR1* (22/100) among the set of 15 genes encompassed in case#2 deletion estimating a global Ho-based P (PH)= 0.000005 (Figure 3). The same analysis with AA

(HP:0002023) connected the genes *FGF13* (4/100), *SOX3* (12/100) and *MAGEC2* (42/100) assessing a P (AA) = 0.002 (Figure 3).

DISCUSSION

The present study showed a full molecular characterization of three patients with severe HB caused by entire *F9* deletions and contiguous genes.

The widely available applied practical approach, including SNP-array analysis, case-specific STS walking and gap-PCR amplification-based characterization of the deletion breakpoints in hemizygous probands, has proved to be capable to cover all desirable genetic objectives.

The sequence characterization of the deletion events allowed the exact determination of genetic loss permitting reliable genotype/phenotype correlations based on the extensive use of bioinformatic-statistical approaches and publicly accessible protein networks and databases.

Inconsistent with the FIX:C <1%, normally associated with severe HB (e.g., our typical case#3), cases #1 and #2 expressed a mild bleeding phenotype defined as two or fewer bleeding episodes per year, a median lifelong concentrate consumption lower than 500 IU/kg and less than one hemarthroses (Beltran-Miranda et al. 2005). Despite the comprehensive search efforts from different sources, the literature has not reported cases of HB caused by deletions involving the *F9* and some adjacent megabases, expressing a mild bleeding phenotype.

Although HPO classification of phenotype abnormalities in the patients was based on precise clinical diagnoses, syndromic cases express as an intricate composite plexus of signs and symptoms hindering accurate dissections between defined function losses and phenotypic characteristics. Related to this later statement is the tight HPO relationship among the three main common phenotypes expressed by our cases #1 and #2 (i.e., SGDD, GH and indeed OB) and a number of genes HPO associated with each of these phenotypic entities and case#1/#2 common deleted genes (Supplementary Figure SF2).

Bioinformatic/statistical analysis performed in this work weighting the number of interactions among case#1/#2 common deleted genes and the common expressed phenotypes focus the main attention on two genes over the others: *FGF13* and *SOX3* (Figure 2, Supplementary Figure SF2). Moreover, our approach allowed estimation of the probabilities of the combined deletion of *SOX3* and *FGF13* genes indicating highly significant associations with SGDD ($P = 0.001$), GH ($P = 0.0072$) and OB ($P = 0.0066$). Focusing on single gene deletion and the Ho-based on randomness, *SOX3* deletion associates our findings with the phenotype combination SGDD-GH-OB estimating a highly significant association $P < 0.0004$, whilst *FGF13* deletion estimated even a more significant $P = 0.00012$.

Notably, a comprehensive and detailed analysis of the literature evidence compiling all entire *F9* deletions expanding from Xq26.3 to Xq27.2 reaches similar conclusions about highlighting *FGF13* and *SOX3* as playing key roles impacting the observed phenotypes (Supplementary Figure SF3) (Nakamura et al. 2016; Hewitt et al. 2014; Jourdy et al. 2016; Wu et al. 2014).

For example, Jourdy et al (Jourdy et al. 2016) reported five deletions including the entire *F9* one of them involving from *FGF13* to *SPANXN4* including *SOX3* and *MAGEC2* among other genes and expressed overweight, hypotonia, bilateral cryptorchidism, scoliosis, mild intellectual disability (ID), special school needs and developmental delay (DD) (late walking); and three cases encompassing several genes on *F9* downstream region reaching at least up to *SPANXC* expressing mostly severe to mild ID, special school needs, short stature, hypotonia, DD (late walking) and delayed speech acquisition, but none of them obesity (Supplementary Figure SF3). Nakamura et al (Nakamura et al. 2016) and Hewitt et al (Hewitt et al. 2014) reported four and one large deletions targeting the entire *F9*, respectively; two of them, who showed deletions spanning all genes in the studied region, expressed seizures and ID adding scoliosis and partial hypopituitarism but there was not evidence of suffering obesity or overweight (Supplementary Figure SF3).

Although it may reflect a lack of a detailed and systematic clinical diagnosis of syndromic patients bearing

large deletions in this region expressing severe HB and other severe conditions, the evidence from Nakamura and Hewitt cases do not support the causative linkage between OB and *FGF13*. Supporting the association OB/*FGF13*, experimental murine models demonstrated that *FGF13* loss associates with weight gain (Sinden et al. 2019). There have been persistent suggestions that *FGF* gene family has broader roles in craniofacial development, neural differentiation, presynaptic neurotransmitter vesicle number, and hair growth (Muñoz-Sanjuan et al. 2001; Nishimoto et al. 2007; Xiao et al. 2007; DeStefano et al. 2013).

According to the logical association, smaller deletions involving the *F9* and minor extensions nearby are only associated with HB, our case#3, with full *F9* and partial *MCF2* deletion expressed only HB related clinical features and no additional phenotype in concordance with the literature (e.g., Wu and Nakamura reported cases, Supplementary Figure SF3) (Nakamura et al. 2016; Jourdy et al. 2016; Wu et al. 2014; Anson et al. 1988).

Several *loci* responsible for mental disorders were described, from those that have been mapped to the Xq26-27 region of the human X chromosome. For example, the Borjeson-Forssman-Lehmann syndrome (BFLS) that includes the *SOX3* deletion showed clinical symptoms of mild-moderate intellectual handicap (Stevanovic et al. 1993; Gedeon et al. 1996). The expression pattern of *SOX3* was associated with normal physiological processes during neural and pituitary development whilst variants in *SOX3* were associated with intellectual delay in some cases (Alatzoglou et al. 2014; Laumonnier et al. 2002).

According to evidence obtained from STRING, the deleted gene *CXorf66* is experimentally associated with *POMGNT1*, which is HPO-related to SGDD, OB and GH (Figure 2, Supplementary Figure SF2). *POMGNT1* (protein O-linked mannose N-acetyl-glucosaminyl-transferase 1) has been related to inherited disorders that affect development of the muscles, brain, and eyes, and some cases of muscular dystrophies associated with pathogenic mutations in *POMGNT1* often presented with mental retardation (Mercuri et al. 2009).

Case#2 expressed the entire set of five extra-HB clinical features (i.e., AA and PH, added to OB, GH and SGDD) and only one additional deleted gene, *MAGEC2*. Although HPO database did not reveal a direct association between *MAGEC2* with PH or AA, some suggestive observations should be taken into account. All MAGE gene family members, including the deleted *MAGEC2*, *MAGEC1* and *MAGEC3* genes, were shown to enhance ubiquitin ligase activity of RING-type zinc finger-containing E3 ubiquitin-protein ligases (Doyle et al. 2010). Ubiquitination Cascade Pathway is known to be involved in the regulation of a majority, if not all, cellular processes, signaling pathways, and disease states. Moreover, STRING analysis indicated a co-expression in testis of *MAGEC2* and an AA-related gene, *ZIC3* (a member of the Zinc finger of the cerebellum (ZIC) protein family) (Figure 3) both mapping to vicinal regions on Xq26.3-q27.3. *ZIC3* was described as a transcriptional factor required in the earliest stages in both axial midline development and left-right asymmetry specification, development of the foregut, urogenital and hindgut malformations, central nervous system anomalies, heterotaxy, and complex congenital heart diseases (Chung et al. 2011). Taking into account all clinical features and all deleted genes, *SOX3* was directly associated with PH-related genes in the HPO database, and showed interactions with AA-related genes (Figure 3).

Our findings revealed that our practical approach combining the information from public databases with the statistical evaluation of null hypothesis likelihoods may be used as a shortcut to achieve meaningful associations between gene function losses and phenotypic abnormalities. However, an intrinsic drawback of the approach, about the difficulty to get univocal associations between a single gene loss and a single well-defined phenotypic feature, highlights the necessity to combine all the mined information considering all genes (and genetic elements) and clinical characteristics of syndromic cases.

In this broader sense, our practical approach has been able to show the probable involvement of *FGF13*, *SOX3*, *CXorf66* and MAGE family genes in obesity, developmental delay, hypotonia, pituitary hypothyroidism and anal atresia.

Beyond these considerations, the detailed clinical report of syndromic cases of fully characterized X-linked deletions contributes with indispensable information to enrich public databases, ultimately improving prediction and eventually prevention of the full expression of genetic disorders from the mere definition of a

properly characterized genotype.

Acknowledgements: The authors thank Miguel de Tezanos Pinto and Mirta Hepner for their kind help in different phases of this work.

Funding/Support: This study was supported by grants from the René Barón Foundation, the Alberto J. Roemmers Foundation, the Florencio Fiorini Foundation, the National Research Council (CONICET), the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT) and the World Federation of Hemophilia.

Conflict of Interest Disclosures: The other authors have no example conflicts of interest to disclose.

Ethics approval: Our local institutional Ethical Committee (CEIANM) approved this study (Res.215/96-April/2007).

Consent to participate: A written informed consent was obtained from all cases.

Contributors' Statement:

Drs Radic and Abelleyro conceptualized and designed the study, drafted the initial manuscript, and reviewed and revised the manuscript.

MSc Ziegler, Drs Marchione, Nevado, Lapunzina, Sciuccati, Neme and Rossetti designed the data collection instruments, collected data, carried out the initial analyses, and reviewed and revised the manuscript.

Drs Bonduel and De Brasi conceptualized and designed the study, coordinated and supervised data, and critically reviewed the manuscript for important intellectual content.

All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

REFERENCES

- Alatzoglou KS, Azriyanti A, Rogers N, Ryan F, Curry N, Noakes C, Bignell P, Hall GW, Litooij AS, Saunders D, Thomas P, Stewart H, Dattani MT (2014) SOX3 deletion in mouse and human is associated with persistence of the craniopharyngeal canal. *J Clin Endocrinol Metab* 99:E2702–8. doi: 10.1210/jc.2014-1160.
- Anson DS, Blake DJ, Winship PR, Birnbaum D, Brownlee GG (1988) Nullisomic deletion of the mcf.2 transforming gene in two haemophilia B patients. *EMBO J* 7(9): 2795-9
- Beltran-Miranda CP, Khan A, Jaloma-Cruz AR, Laffan MA (2005) Thrombin generation and phenotypic correlation in haemophilia A. *Haemophilia* 11:326–34. doi: 10.1111/j.1365-2516.2005.01107.x.
- Chung B, Shaffer LG, Keating S, Johnson J, Casey B, Chitayat D (2011) From VACTERL-H to heterotaxy: variable expressivity of ZIC3-related disorders. *Am J Med Genet* 2011;155A:1123–8. doi: 10.1002/ajmg.a.33859.
- den Dunnen JT (2017) Describing Sequence Variants Using HGVS Nomenclature. *Methods Mol Biol* 1492:243-51. doi: 10.1007/978-1-4939-6442-0_17.
- DeStefano GM, Fantauzzo KA, Petukhova L, Kurban M, Tadin-Strapps M, Levy B, Warburton D, Cirulli ET, Han Y, Sun X, Shen Y, Shirazi M, Jobanputra V, Cepeda-Valdes R, Salas-Alanis JC, Christiano AM (2013) Position Effect on FGF13 Associated With X-linked Congenital Generalized Hypertrichosis. *Proc Natl Acad Sci USA* 7;110(19):7790-5. doi: 10.1073/pnas.1216412110.
- Doyle JM, Gao J, Wang J, Yang M, Potts PR (2010) MAGE-RING protein complexes comprise a family of E3 ubiquitin ligases. *Mol Cell* 39(6):963-74. doi: 10.1016/j.molcel.2010.08.029.
- Gedeon AK, Kozman HM, Robinson H, Pilia G, Schlessinger D, Turner G, Mulley JC (1996) Refinement of the Background Genetic Map of Xq26-q27 and Gene Localisation for Borjeson-Forssman-Lehmann Syndrome.

Am J Med Genet 12;64(1):63-8. doi: 10.1002/(SICI)1096-8628(19960712)64:1<63::AID-AJMG9>3.0.CO;2-S.

Hewitt J, Chou EM, Brown LA, Smith VC, Yong SL, Wadsworth LD, Wu JK, Macgillivray RT (2014) Molecular characterization of a 4,409,480 bp deletion of the human X chromosome in a patient with haemophilia B. *Haemophilia* 20(3):e230-4. doi: 10.1111/hae.12395.

Jourdy Y, Chatron N, Carage ML, Fretigny M, Meunier S, Zawadzki C, Gay V, Negrier C, Sanlaville D, Vinciguerra C (2016) Study of six patients with complete F9 deletion characterized by cytogenetic microarray: role of the SOX3 gene in intellectual disability. *J Thromb Haemost* 14(10): 1988-93. doi: 10.1111/jth.13430.

Ketterling RP, Vielhaber EL, Lind TJ, Thorland EC, Sommer SS (1994) The rates and patterns of deletions in the human factor IX gene. *Am J Hum Genet* 54(2):201-13.

Koeberl DD, Bottema CD, Ketterling RP, Bridge PJ, Lillicrap DP, Sommer SS (1990) Mutations causing hemophilia B: direct estimate of the underlying rates of spontaneous germ-line transitions, transversions, and deletions in a human gene. *Am J Hum Genet* 47(2):202-17.

Lahiri DK, Nuremberg JI (1991) A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. *Nucleic Acids Res* 19:5444.

Laumonnier F, Ronce N, Hamel BCJ, Thomas P, Lespinasse J, Raynaud M, Paringaux C, Van Bokhoven H, Kalscheuer V, Fryns JP, Chelly J, Moraine C, Briault S (2002) Transcription factor SOX3 is involved in X-linked mental retardation with growth hormone deficiency. *Am J Hum Genet* 71:1450-5. doi: 10.1086/344661.

Mercuri E, Messina S, Bruno C, Mora M, Pegoraro E, Comi GP, D'Amico A, Aiello C, Biancheri R, Berrardinelli A, Boffi P, Cassandrini D (2009) Congenital muscular dystrophies with defective glycosylation of dystroglycan: a population study. *Neurology* 72:1802-9. doi: 10.1212/01.wnl.0000346518.68110.60.

Munoz-Sanjuan I, H-Brivanlou A (2001) Early posterior/ventral fate specification in the vertebrate embryo. *Dev Biol* 1;237(1):1-17. doi: 10.1006/dbio.2001.0350.

Nakamura Y, Ando Y, Takagi Y, Murata M, Kozuka T, Nakata Y, Hasebe R, Takagi A, Matsushita T, Shima M, Kojima T (2016) Distinct X chromosomal rearrangements in four haemophilia B patients with entire F9 deletion. *Haemophilia* 22(3):433-9. doi: 10.1111/hae.12849.

Nishimoto S, Nishida E (2007) Fibroblast Growth Factor 13 is essential for neural differentiation in *Xenopus* early embryonic development. *J Biol Chem* 17;282(33):24255-61. doi: 10.1074/jbc.M704277200.

Radic CP, Rossetti LC, Abelleiro MM, Candela M, Perez Bianco R, de Tezanos Pinto M, Larripa IB, Goodeve A, De Brasi CD (2013) Assessment of the F9 genotype-specific FIX inhibitor risks and characterisation of 10 novel severe F9 defects in the first molecular series of Argentinian patients with haemophilia B. *Thromb. Haemost* 109:24-33. doi: 10.1160/TH12-05-0302.

Santagostino E, Mancuso ME, Tripodi A, Chantarangkul V, Clerici M, Garagiola I, Mannucci PM (2010) Severe hemophilia with mild bleeding phenotype: molecular characterization and global coagulation profile. *J. Thromb. Haemost* 8:737-43. doi: 10.1111/j.1538-7836.2010.03767.x.

Santoro C, Quintavalle G, Castaman G, Baldacci E, Ferretti A, Riccardi F, Tagliaferri A (2018) Inhibitors in Hemophilia B. *Semin Thromb Hemost* 44(6):578-89. doi: 10.1055/s-0038-1660817.

Sinden DS, Holman CD, Bare CJ, Sun X, Gade AR, Cohen DE, Pitt GS (2019) Knockout of the X-linked *Fgf13* in the Hypothalamic Paraventricular Nucleus Impairs Sympathetic Output to Brown Fat and Causes Obesity. *FASEB J* 33(10):11579-94. doi: 10.1096/fj.201901178R.

Stevanović M, Lovell-Badge R, Collignon J, Goodfellow PN (1993) SOX3 Is an X-linked Gene Related to SRY. *Hum Mol Genet* 2(12): 2013-8.

Stoof SCM, Kersseboom R, de Vries FAT, Kruip MJHA, Kievit AJA, Leebeek FWG (2018) Hemophilia B in a female with intellectual disability caused by a deletion of Xq26.3q28 encompassing the F9. *Mol Genet Genomic Med* 6(6):1220-4. doi: 10.1002/mgg3.425.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ, von Mering C (2019) STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 47: D607-13. doi: 10.1093/nar/gky1131.

Thorland EC, Drost JB, Lusher JM, Warriar I, Shapiro A, Koerper MA, Dimichele D, Westman J, Key NS, Sommer SS (1995) Anaphylactic response to factor IX replacement therapy in haemophilia B patients: complete gene deletions confer the highest risk. *Haemophilia* 5:101-5.

Tjárnrlund-Wolf A, Lassila R (2019) Phenotypic characterization of haemophilia B - Understanding the underlying biology of coagulation factor IX. *Haemophilia* 25(4):567-574. doi: 10.1111/hae.13804.

Vencesla A, Barcelo MJ, Baena M, Quintana M, Baiget M, Tizzano EF (2007) Marker and real-time quantitative analyses to confirm hemophilia B carrier diagnosis of a complete deletion of the *F9* gene. *Haematologica* 92:1583-4. doi: 10.3324/haematol.10693.

Wu X, Lu Y, Ding Q, You G, Dai J, Xi X, Wang H, Wang X (2104) Characterisation of large F9 deletions in seven unrelated patients with severe haemophilia B. *Thromb Haemost* 2;112(3):459-65. doi: 10.1160/TH13-12-1060.

Xiao M, Xu L, Laezza F, Yamada K, Feng S, Ornitz DM (2007) Impaired Hippocampal Synaptic Transmission and Plasticity in Mice Lacking Fibroblast Growth Factor 14. *Mol Cell Neurosci* 34(3):366-77. doi: 10.1016/j.mcn.2006.11.020.

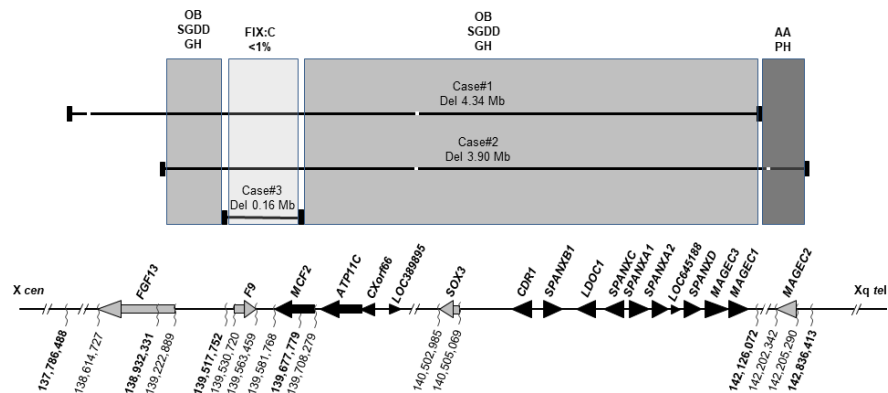
LEGENDS

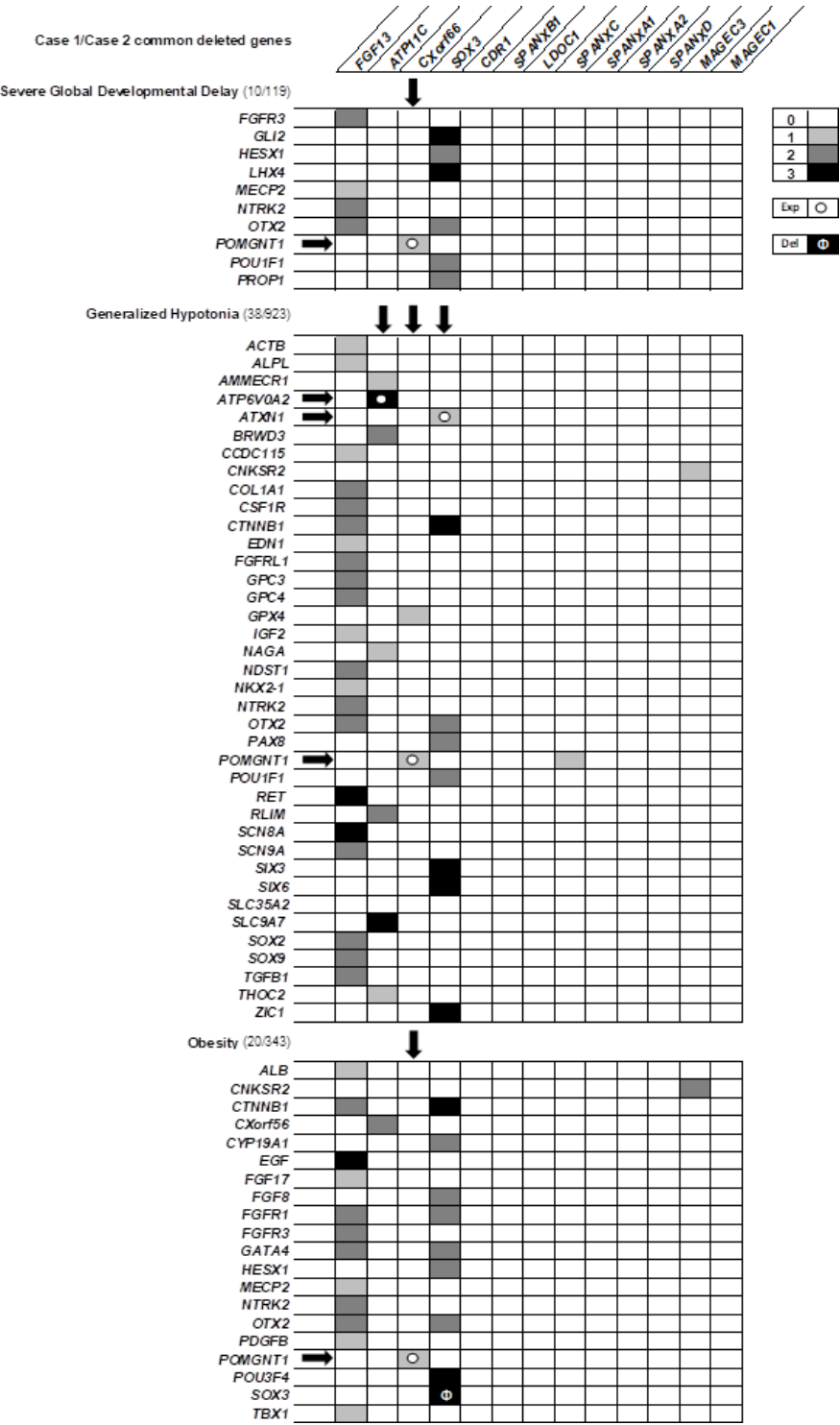
Figure 1. Scheme showing the extent of characterized large deletions from our series on Xq26-q27 reference region including all intervenient genes . Upper panel: extent (Mb, mega base pairs) of the three presented cases (case#1-#3). Lower panel: graphical representation of genes (official names in *italics*) closed arrows. Relevant X-chromosome gene coordinates (NC_000023.11, GRCh38.p13) are shown below. Deletion breakpoint's coordinates are highlighted in boldface. Most relevant genes (i.e., *FGF13* , *F9* , *MCF2* , *SOX3* and *MAGEC2*) coordinates are indicated (grey). Genes involved in the deleted regions (GRCh38.p13) are shown. Shared clinical features between cases are depicted as grey rectangle areas on the background; SGDD: Severe Global Developmental Delay; GH: Generalized Hypotonia; OB: Obesity; FIX:C <1%: Reduced factor IX activity, severe Hemophilia B; AA: Anal Atresia; PH: Pituitary Hypothyroidism.

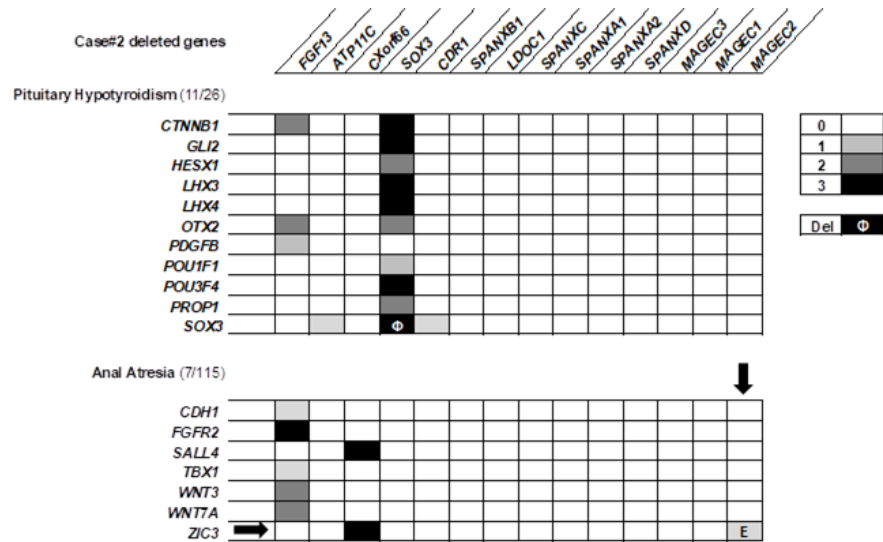
Figure 2. Heat map showing STRING based interactions between case#1/#2 common deleted genes and the genes HPO associated with their common phenotypic abnormalities. From white to black, the gray gradient correlates with the number of connections classifying: 0 (no interaction, open box), 1 (1 connection, pale gray), 2 (dark gray) and [?]3 (closed box). Case#1 and #2 (characterized by a common region of hemizygous deletions) share regions encompassing 13 deleted genes, upper row; common phenotype abnormalities classified by HPO, i.e., Severe Global Developmental Delay (HP:0011344), Generalized Hypotonia (HP:0001290) and Obesity (HP:0001513). Exp (open circle) indicates *experimental* only evidence item in the STRING database (Protein-Protein Interaction Networks and Functional Enrichment Analysis); black arrows highlight evidence connected genes; Del (Greek capital Phi, Φ) indicates gene deletion (*SOX3* in Obesity). Numbers in brackets, e.g., Obesity (20/343), indicate number of genes showing [?]1 connection divided by all genes associated with the given phenotype.

Figure 3. Heat map showing the same STRING based analysis of Figure 2 but associating case#2 deleted genes (n=14) with Human Phenotype Ontology genes of Pituitary Hypothyroidism (HP:0008245) and Anal Atresia (HP:0002023). The same conventions as Figure 2 in number of STRING connections are used. E indicates that *MAGEC2* is the only extra deleted gene in case#2. Arrows

underline this direct connection between MAGEC2 and ZIC3 (an Anal Atresia linked gene). No experimental evidence-based STRING connections were found.







Hosted file

Table 1.docx available at <https://authorea.com/users/334229/articles/476235-syndromic-cases-of-hemophilia-b-and-morbid-obesity-due-to-contiguous-gene-deletions-on-xq26-3-q27-2-unsuspected-phenotype-genotype-associations-by-bioinformatics-and-extensive-clinical-data-mining>

Hosted file

Table 2.docx available at <https://authorea.com/users/334229/articles/476235-syndromic-cases-of-hemophilia-b-and-morbid-obesity-due-to-contiguous-gene-deletions-on-xq26-3-q27-2-unsuspected-phenotype-genotype-associations-by-bioinformatics-and-extensive-clinical-data-mining>